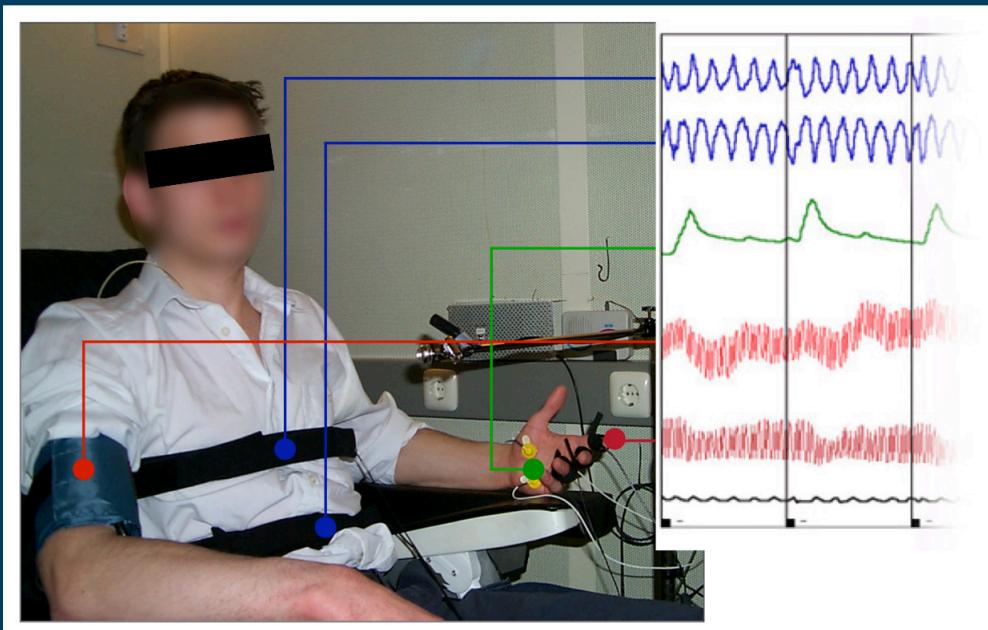
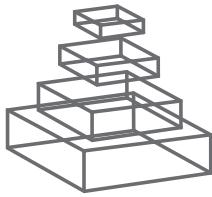


frontiers RESEARCH TOPICS



BASIC AND APPLIED RESEARCH ON
DECEPTION AND ITS DETECTION

Topic Editors
Wolfgang Ambach and Matthias Gamer



FRONTIERS COPYRIGHT STATEMENT

© Copyright 2007-2014
Frontiers Media SA.
All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88919-254-0

DOI 10.3389/978-2-88919-254-0

ABOUT FRONTIERS

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

FRONTIERS JOURNAL SERIES

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing.

All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

DEDICATION TO QUALITY

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

WHAT ARE FRONTIERS RESEARCH TOPICS?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area!

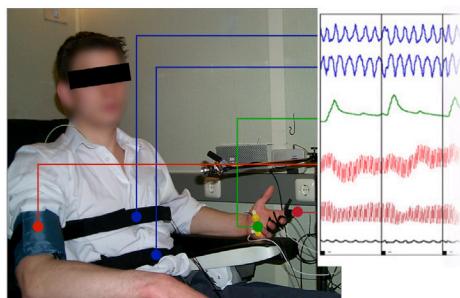
Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

BASIC AND APPLIED RESEARCH ON DECEPTION AND ITS DETECTION

Topic Editors:

Wolfgang Ambach, Institute for Frontier Areas of Psychology and Mental Health (IGPP), Germany

Matthias Gamer, University Medical Center Hamburg-Eppendorf, Germany



Detecting deception (picture by Matthias Gamer).

Deception is a ubiquitous phenomenon in social interactions and has attracted a significant amount of research during the last decades. The majority of studies in this field focused on how deception modulates behavioral, autonomic, and brain responses and whether these changes can be used to validly identify lies. Especially the latter question, which historically gave rise to the development of psychophysiological “lie detection” techniques, has been driving research on deception and its detection

until today. The detection of deception and concealed information in forensic examinations currently constitutes one of the most frequent applications of psychophysiological methods in the field.

With the increasing use of such methods, the techniques for detecting deception have been controversially discussed in the scientific community. It has been proposed to shift from the original idea of detecting deception *per se* to a more indirect approach that allows for determining whether a suspect has specific knowledge of crime-related details. This so-called Concealed Information Test is strongly linked to basic psychological concepts concerning memory, attention, orienting, and response monitoring.

Although research in this field has intensified with the advancement of neuroimaging techniques such as PET and fMRI in the last decade, basic questions on the psychological mechanisms underlying modulatory effects of deception and information concealment on behavioral, autonomic, and brain responses are still poorly understood.

This Research Topic brings together contributions from researchers in experimental psychology, psychophysiology, and neuroscience focusing on the understanding of the broad concept of deception including the detection of concealed information, with respect to

basic research questions as well as applied issues. This Research Topic is mainly composed of original research articles but reviews and papers elaborating on novel methodological approaches have also been included. Experimental methods include, but are not limited to, behavioral, autonomic, electroencephalographic or brain imaging techniques that allow for revealing relevant facets of deception on a multimodal level. While this Research Topic primarily includes laboratory work, relevant issues for the field use of such methods are also discussed.

Table of Contents

- 06 Deception Research Today**
Matthias Gamer and Wolfgang Ambach
- 09 Current Research and Potential Applications of the Concealed Information Test: An Overview**
Gershon Ben-Shakhar
- 20 Effects of the Combination of P3-Based GKT and Reality Monitoring on Deceptive Classification**
Ki-Won Jang, Deok-Yong Kim, Sungkun Cho and Jang-Han Lee
- 27 Concealed Semantic and Episodic Autobiographical Memory Electrified**
Giorgio Ganis and Haline E. Schendan
- 48 P300 Amplitudes in the Concealed Information Test are Less Affected by Depth of Processing than Electrodermal Responses**
Matthias Gamer and Stefan Berti
- 58 "Have You Ever Seen This Face?" – Individual Differences and Event-Related Potentials During Deception**
Anja Leue, Sebastian Lange and André Beauducel
- 71 Face and Voice as Social Stimuli Enhance Differential Physiological Responding in a Concealed Information Test**
Wolfgang Ambach, Birthe Assmann, Bennet Krieg and Dieter Vaitl
- 83 A Functional Analysis of Deception Detection of a Mock Crime Using Infrared Thermal Imaging and the Concealed Information Test**
Kevin K. Park, Hye Won Suk, Heungsun Hwang and Jang-Han Lee
- 100 Combining Blink, Pupil, and Response Time Measures in a Concealed Knowledge Test**
Travis L. Seymour, Christopher A. Baker and Joshua T. Gaunt
- 115 Detecting False Intent Using Eye Blink Measures**
Frank M Marchak
- 124 The Current and Future Status of the Concealed Information Test for Field Use**
Izumi Matsuda, Hiroshi Nittono and John J. B. Allen
- 135 Advancing Lie Detection by Inducing Cognitive Load on Liars: A Review of Relevant Theories and Techniques Guided by Lessons from Polygraph-Based Approaches**
Jeffrey J. Walczyk, Frank P. Igou, Alexa P. Dixon and Talar Tcholakian
- 148 When Interference Helps: Increasing Executive Load to Facilitate Deception Detection in the Concealed Information Test**
George Visu-Petra, Mihai Varga, Mircea Miclea and Laura Visu-Petra

- 159 *Detecting Concealed Information from Groups Using a Dynamic Questioning Approach: Simultaneous Skin Conductance Measurement and Immediate Feedback***
Ewout H. Meijer, Gary Bente, Gershon Ben-Shakhar and Andreas Schumacher
- 165 *The Autobiographical IAT: A Review***
Sara Agosta and Giuseppe Sartori
- 177 *Learning to Lie: Effects of Practice on the Cognitive Cost of Lying***
B. Van Bockstaele, B. Verschueren, T. Moens, Kristina Suchotzki, Evelyne Debey and Adriaan Spruyt
- 185 *A Repeated Lie Becomes a Truth? The Effect of Intentional Control and Training on Deception***
Xiaoqing Hu, Hao Chen and Genyue Fu
- 192 *When Pinocchio's Nose Does Not Grow: Belief Regarding Lie-Detectability Modulates Production of Deception***
Kamila E. Sip, David Carmel, Jennifer L. Marchant, Jian Li, Predrag Petrovic, Andreas Roepstorff, William B. McGregor and Christopher D. Frith
- 203 *Does the Inferior Frontal Sulcus Play a Functional Role in Deception? A Neuronavigated Theta-Burst Transcranial Magnetic Stimulation Study***
Bruno Verschueren, Teresa Schuhmann and Alexander T. Sack
- 210 *Markers of Deception in Italian Speech***
Katelyn Spence, Gina Villar and Joanne Arciuli
- 219 *Exploring the Movement Dynamics of Deception***
Nicholas D. Duran, Rick Dale, Christopher T. Kello, Chris N. H. Street and Daniel C. Richardson
- 235 *How Do Incentives Lead to Deception in Advisor-Client Interactions? Explicit and Implicit Strategies of Self-Interested Deception***
Barbara Mackinger and Eva Jonas



Deception research today

Matthias Gamer^{1*} and Wolfgang Ambach^{2*}

¹ Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

² Institute for Frontier Areas of Psychology and Mental Health, Freiburg, Germany

*Correspondence: m.gamer@uke.uni-hamburg.de; ambach@igpp.de

Edited and reviewed by:

Eddy J. Davelaar, Birkbeck College, United Kingdom

Keywords: deception, Concealed Information Test, differentiation of deception paradigm, application, theory

INTRODUCTION

Deception is a complex social behavior which involves a set of higher cognitive functions. Studying this common phenomenon in humans has in all epochs been driven not merely by the wish to understand the underlying framework of cognitive functioning but rather by the ambition to detect deceptive behavior in criminal suspects. Thus, identifying valid indicators of deceptive behavior has always been in the focus of deception research. Such indicators can be defined in terms of specific behavior, physiological correlates, or content of verbal reports. The question of how validly each indicator allows for differentiating truthful and deceptive accounts is inherent in the majority of research efforts in this domain.

Another important aspect concerns the development of deception theory. According to current opinion, deception is not characterized by a single cognitive process but rather involves the combination of a variety of basic cognitive processes such as working memory, response monitoring and inhibition. Identifying these processes, modeling their interplay and their modulation by personality and situational factors is still one major challenge in deception research. Furthermore, deception is no unitary phenomenon. Correspondingly, researchers need to examine and describe different variants of this phenomenon occurring in distinct contexts, which entails a variety of experimental and theoretical approaches that largely differ in scope and methods.

CURRENT INTERESTS

One major field in deception research concerns the use of psychophysiological methods to detect deceptive behavior. Over time, the traditional physiological measures (electrodermal, cardiovascular, and respiratory responses) have been supplemented by electroencephalographic, functional imaging, and other innovative procedures. Finding measures that validly discriminate between truth and lie, and the wish to optimize their use, have received new impetus from recent technological development. Neuroimaging techniques, for example, yield new promises and deserve a deep evaluation. Thermal imaging and eyetracking are other innovative methods which might provide additional information about the mental processes involved in generating deceptive responses. However, even "classic" behavioral measures such as response times are still frequently used in this domain for theoretical as well as applied purposes.

Different techniques for detecting deception with the help of physiological measures have been controversially discussed in the scientific community. Among the most influential experimental

paradigms, the so-called Concealed Information Test (CIT, Lykken, 1959) has received broad scientific attention. The CIT does not target at identifying deception *per se* but rather aims at detecting whether a suspect has concealed knowledge of specific (e.g., crime-related) details. A different approach is the so-called differentiation-of-deception paradigm (Furedy et al., 1988) which follows the aim to identify specific patterns in behavioral or physiological variables that differ systematically between truthful and deceptive behavior. Particularly this latter approach has been newly fueled by brain imaging techniques which claim to mirror mental processes accompanying deceptive responses more directly.

The current Research Topic brings together contributions from experimental psychology, psychophysiology, and neuroscience focusing on the understanding of the broad concept of deception including the detection of concealed information, with respect to basic research questions as well as applied issues. Due to the interdisciplinary focus of this approach, articles were published in Frontiers in Psychology or Frontiers in Human Neuroscience, respectively.

CURRENT RESEARCH

Most articles of this Research Topic focus on the detection of concealed information using variants of the CIT. A large body of previous research has documented that perpetrators show larger electrodermal responses, respiratory suppression, as well as heart rate deceleration to crime-related probes (e.g., a murder weapon) as compared to neutral items (e.g., other unrelated weapons). More recently, comparable differences were reported for behavioral responses, specific components of event-related brain potentials, and neurovascular changes in specific brain regions measured by neuroimaging techniques (for a comprehensive review see Verschuere et al., 2011). Under the premise that innocents do not possess crime-related knowledge, the CIT can be used to validly differentiate perpetrators from innocents. Although the CIT has been first described in the middle of the last century (Lykken, 1959), there are still a number of open questions concerning the theoretical background, the validity of new measures, or special applications for specific circumstances of crimes. Some of these questions were addressed by current articles included in this Research Topic.

Several studies focused on event-related brain potentials and demonstrated that ERP components were susceptible to contextual factors such as the personal involvement in a misdeed during encoding (Jang et al., 2013) or the nature of memory tested in the CIT (episodic vs. semantic, Ganis and Schendan,

2013). Furthermore, it was found that depth of encoding modulated electrodermal responses to crime-related details but did not affect P300 responses in the CIT (Gamer and Berti, 2012). A large study with more than 100 participants reported a modulation of ERP components by personality traits regarding sensitivity to moral and social norms as well as cognitive-motivational conflict processing (Leue et al., 2012). Ambach and colleagues failed to find a similar influence of interindividual differences in psychopathy but reported higher detection accuracy of autonomic measures when the CIT procedure included the face of the fictive interrogator and verbal instead of textual question presentation (Ambach et al., 2012). Two further studies explored the validity of novel physiological and ocular measures in the CIT. Park and colleagues successfully used facial temperature in a periorbital region as determined by thermal imaging to detect concealed knowledge (Park et al., 2013); Seymour and colleagues were able to accurately determine hidden knowledge by differences in pupil responses and blink rates (Seymour et al., 2012). Using a slightly different interrogation protocol, Marchak (2013) could show that eye blink measures even allow for differentiating truthful and false intent. These studies collectively demonstrate that a number of behavioral and physiological variables are susceptible to concealed information and therefore allow for detecting individuals with crime-related knowledge. Moreover, several of these studies reported enhanced classification accuracy when combining different indices of concealed information (Ambach et al., 2012; Seymour et al., 2012; Jang et al., 2013).

Besides using new measures or combining different behavioral and physiological indices, it has been suggested to increase cognitive load during the examination to facilitate the detection of liars. Walczyk and colleagues provided a review and a detailed theoretical outline of this approach (Walczky et al., 2013). Visu-Petra and colleagues explored such strategy empirically by asking participants to carry out a secondary task simultaneously to the CIT. They could show that such interference facilitates the detection of concealed information based on behavioral measures (Visu-Petra et al., 2013). In a seminal study, Meijer and colleagues developed a novel dynamic questioning approach that does not concern individual responses of single examinees but instead the global responsiveness of a group of suspects. Such method was shown to allow for an identification of specific details of a collectively planned mock terrorist attack (Meijer et al., 2013). Finally, Agosta and Sartori (2013) summarized recent studies on the so-called autobiographical Implicit Association Test; this promising new development allows for accurately determining whether an autobiographical memory is encoded in a given suspect. Taken together, these studies document the substantial advancement of current research on the CIT regarding the potential of novel methods as well as situational and personality factors that are modulating the response pattern. Beyond these efforts, Ben-Shakhar (2012) identified a number of open questions regarding practical aspects and outlined future directions for research on the CIT. These issues are highly relevant for the field implementation of the CIT in police investigations. Such procedure, which is currently only adopted in Japan, is discussed in great detail by Matsuda et al. (2012). The vital international research on the CIT along with the large body of practical experience with this

technique in Japan holds promise for further implementations of the CIT as an advancement of currently applied deception detection techniques.

One major problem of current techniques is their susceptibility to countermeasures. Thus, guilty examinees might be able to deliberately alter their pattern of responses to appear innocent. Similarly, certain groups of suspects might have less difficulty in lying as compared to others because of frequent lying in general. Two studies in the current Research Topic explored these issues using variants of the differentiation-of-deception paradigm. Increasing the proportion of deceptive as compared to truthful responses led to reduced reaction time differences between truth and lie, which might indicate that lies require less cognitive effort in frequent liars and are therefore more difficult to detect (Van Bockstaele et al., 2012). Hu and colleagues showed that the instruction to selectively speed up deceptive answers along with a short training substantially altered the pattern of response times such that truthful and deceptive responses became indistinguishable (Hu et al., 2012). However, it cannot be generalized from these results that merely emphasizing that an examination aims at detecting deception necessarily reduces lie detection efficacy. By contrast, a study using a variant of the differentiation-of-deception paradigm in conjunction with functional magnetic resonance imaging revealed larger differences between deceptive and truthful answers in the neural activation of different brain areas when participants believed that a lie-detector was activated (Sip et al., 2013). In line with the majority of neuroimaging studies in this domain (Gamer, 2011), activity in the right inferior frontal gyrus was also modulated by deception. This region was frequently supposed to reflect the recruitment of response inhibition processes. However, temporary disruption of the inferior frontal gyrus by means of continuous theta-burst stimulation did not significantly alter the pattern of behavioral responses in a variant of the differentiation-of-deception paradigm (Verschueren et al., 2012). These results thus question the frequently assumed functional role of the inferior frontal gyrus in deception.

Besides exploring specific cues of deceptive behavior in highly standardized situations and with highly standardized interrogation techniques, it also seems interesting to examine deception in more naturalistic settings. For example, Spence and colleagues asked participants to provide relatively unrestricted honest and deceptive accounts of their opinion regarding social issues. For these accounts, speech parameters were extracted and the authors found a significantly reduced speech rate along with increased response latency during deception compared with truth-telling (Spence et al., 2012). In a similar vein, Duran and colleagues examined movement dynamics accompanying deceptive and truthful accounts. Instead of searching for specific discrete cues such as the rise of a brow, they examined the whole time course of movements and provided preliminary evidence for unique dynamic signatures of deception in these kinetic variables (Duran et al., 2013). Finally, Mackinger and Jonas (2012) explored determinants of deception in advisor-client interactions and provided evidence for the use of explicit and implicit strategic deceptive behavior in advisors aiming to receive an incentive.

PERSPECTIVE

Research on deception has a long tradition in psychology and related fields. On the one hand, the drive for detecting deception has inspired research, teaching, and application over many decades. On the other hand, research on deception as a process or phenomenon is characterized by manifold interactions with other areas of psychological research such as attention, memory, executive control, or motor behavior. It remains to be debated whether deception and its detection should be studied as a key topic which entails addressing these other fields, or rather as a particular, illustrative manifestation of them. We regard the present Research Topic as clearly underlining the scientific benefits arising from the broad and multidisciplinary perspective that characterizes deception research today and we hope that it will enrich and inspire future research in this domain.

REFERENCES

- Agosta, S., and Sartori, G. (2013). The autobiographical IAT: a review. *Front. Psychol.* 4:519. doi:10.3389/fpsyg.2013.00519
- Ambach, W., Assmann, B., Krieg, B., and Vaitl, D. (2012). Face and voice as social stimuli enhance differential physiological responding in a concealed information test. *Front. Psychol.* 3:510. doi:10.3389/fpsyg.2012.00510
- Ben-Shakhar, G. (2012). Current research and potential applications of the concealed information test: an overview. *Front. Psychol.* 3:342. doi:10.3389/fpsyg.2012.00342
- Duran, N. D., Dale, R., Kello, C. T., Street, C. N. H., and Richardson, D. C. (2013). Exploring the movement dynamics of deception. *Front. Psychol.* 4:140. doi:10.3389/fpsyg.2013.00140
- Furedy, J. J., Davis, C., and Gurevich, M. (1988). Differentiation of deception as a psychological process: a psychophysiological approach. *Psychophysiology* 25, 683–688.
- Gamer, M. (2011). “Detecting of deception and concealed information using neuroimaging techniques,” in *Memory Detection: Theory and Application of the Concealed Information Test*, eds B. Verschuere, G. Ben-Shakhar, and E. H. Meijer (Cambridge: University Press), 90–113. doi: 10.1017/CBO9780511975196.006
- Gamer, M., and Berti, S. (2012). P300 amplitudes in the concealed information test are less affected by depth of processing than electrodermal responses. *Front. Hum. Neurosci.* 6:308. doi:10.3389/fnhum.2012.00308
- Ganis, G., and Schendan, H. E. (2013). Concealed semantic and episodic autobiographical memory electrified. *Front. Hum. Neurosci.* 6:354. doi:10.3389/fnhum.2012.00354
- Hu, X., Chen, H., and Fu, G. (2012). A repeated lie becomes a truth? The effect of intentional control and training on deception. *Front. Psychol.* 3:488. doi:10.3389/fpsyg.2012.00488
- Jang, K.-W., Kim, D.-Y., Cho, S., and Lee, J.-H. (2013). Effects of the combination of P3-based GKT and reality monitoring on deceptive classification. *Front. Hum. Neurosci.* 7:18. doi:10.3389/fnhum.2013.00018
- Leue, A., Lange, S., and Beauducel, A. (2012). “Have you ever seen this face?” - Individual differences and event-related potentials during deception. *Front. Psychol.* 3:570. doi:10.3389/fpsyg.2012.00570
- Lykken, D. T. (1959). The GSR in the detection of guilt. *J. Appl. Psychol.* 43, 385–388.
- Mackinger, B., and Jonas, E. (2012). How do incentives lead to deception in advisor-client interactions? Explicit and implicit strategies of self-interested deception. *Front. Psychol.* 3:527. doi:10.3389/fpsyg.2012.00527
- Marchak, F. M. (2013). Detecting false intent using eye blink measures. *Front. Psychol.* 4:736. doi:10.3389/fpsyg.2013.00736
- Matsuda, I., Nittono, H., and Allen, J. J. B. (2012). The current and future status of the concealed information test for field use. *Front. Psychol.* 3:532. doi:10.3389/fpsyg.2012.00532
- Meijer, E. H., Bente, G., Ben-Shakhar, G., and Schumacher, A. (2013). Detecting concealed information from groups using a dynamic questioning approach: simultaneous skin conductance measurement and immediate feedback. *Front. Psychol.* 4:68. doi:10.3389/fpsyg.2013.00068
- Park, K. K., Suk, H. W., Hwang, H., and Lee, J.-H. (2013). A functional analysis of deception detection of a mock crime using infrared thermal imaging and the concealed information test. *Front. Hum. Neurosci.* 7:70. doi:10.3389/fnhum.2013.00070
- Seymour, T. L., Baker, C. A., and Gaunt, J. T. (2012). Combining blink, pupil, and response time measures in a concealed knowledge test. *Front. Psychol.* 3:614. doi:10.3389/fpsyg.2012.00614
- Sip, K. E., Carmel, D., Marchant, J. L., Li, J., Petrovic, P., Roepstorff, A., et al. (2013). When Pinocchio's nose does not grow: belief regarding lie-detectability modulates production of deception. *Front. Hum. Neurosci.* 7:16. doi:10.3389/fnhum.2013.00016
- Spence, K., Villar, G., and Arciuli, J. (2012). Markers of deception in Italian speech. *Front. Psychol.* 3:453. doi:10.3389/fpsyg.2012.00453
- Van Bockstaele, B., Verschuere, B., Moens, T., Suchotzki, K., Debey, E., and Spruyt, A. (2012). Learning to lie: effects of practice on the cognitive cost of lying. *Front. Psychol.* 3:526. doi:10.3389/fpsyg.2012.00526
- Verschuere, B., Ben-Shakhar, G., and Meijer, E. H. (2011). *Memory Detection: Theory and Application of the Concealed Information Test*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511975196
- Verschuere, B., Schuhmann, T., and Sack, A. (2012). Does the inferior frontal sulcus play a functional role in deception? A neuronavigated theta-burst transcranial magnetic stimulation study. *Front. Hum. Neurosci.* 6: 284. doi:10.3389/fnhum.2012.00284
- Visu-Petra, G., Varga, M., Miclea, M., and Visu-Petra, L. (2013). When interference helps: increasing executive load to facilitate deception detection in the concealed information test. *Front. Psychol.* 4:146. doi:10.3389/fpsyg.2013.00146
- Walczuk, J. J., Igou, F. P., Dixon, A. P., and Tcholakian, T. (2013). Advancing lie detection by inducing cognitive load on liars: a review of relevant theories and techniques guided by lessons from polygraph-based approaches. *Front. Psychol.* 4:14. doi:10.3389/fpsyg.2013.00014

Received: 28 February 2014; accepted: 10 March 2014; published online: 25 March 2014.

*Citation: Gamer M and Ambach W (2014) Deception research today. *Front. Psychol.* 5:256. doi: 10.3389/fpsyg.2014.00256*

*This article was submitted to Cognitive Science, a section of the journal *Frontiers in Psychology*.*

Copyright © 2014 Gamer and Ambach. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Current research and potential applications of the concealed information test: an overview

Gershon Ben-Shakhar *

The Hebrew University of Jerusalem, Jerusalem, Israel

Edited by:

Matthias Gamer, University Medical Center Hamburg-Eppendorf, Germany

Reviewed by:

Donald Krapohl, National Center for Credibility Assessment, USA

William Iacono, University of Minnesota, USA

***Correspondence:**

Gershon Ben-Shakhar, Department of Psychology, The Hebrew University of Jerusalem, Jerusalem 91905, Israel.

e-mail: mskpugb@mssc.huji.ac.il

Research interest in psychophysiological detection of deception has significantly increased since the September 11 terror attack in the USA. In particular, the concealed information test (CIT), designed to detect memory traces that can connect suspects to a certain crime, has been extensively studied. In this paper I will briefly review several psychophysiological detection paradigms that have been studied, with a focus on the CIT. The theoretical background of the CIT, its strength and weaknesses, its potential applications as well as research findings related to its validity (based on a recent meta-analytic study), will be discussed. Several novel research directions, with a focus on factors that may affect CIT detection in realistic settings (e.g., memory for crime details; the effect of emotional stress during crime execution) will be described. Additionally, research focusing on mal-intentions and attempts to detect terror networks using information gathered from groups of suspects using both the standard CIT and the searching CIT will be reviewed. Finally, implications of current research to the actual application of the CIT will be discussed and several recommendations that can enhance the use of the CIT will be made.

Keywords: the concealed information test, psychophysiological detection of deception, the guilty knowledge test, memory detection, the searching CIT

Deception is a frequent, perhaps essential, feature of human behavior, which may be expressed in a variety of situations (e.g., Saxe, 1991). The frequent use of deception in social contexts highlights the importance of detecting deception. However, research on perceivers' ability to differentiate between truthful and deceptive messages has indicated that, in most cases, people, including professionals whose tasks involve detection of deceit, perform this task at chance levels (see Vrij, 2008 for a review). Consequently it is not surprising that the idea of using physiological measures for detecting deception has been very appealing to law-enforcement agencies (e.g., Marston, 1917, 1938; Larson, 1932; Reid, 1947; Reid and Inbau, 1977). Indeed, several psychophysiological methods (popularly labeled, "polygraph techniques") have been developed since the beginning of the twentieth century and the study of psychophysiological detection of deception has attracted a great deal of interest among researchers as well as practitioners and has become an important area of applied psychology (e.g., Reid and Inbau, 1977; Raskin, 1989; Ben-Shakhar and Furedy, 1990; Lykken, 1998; National Research Council, 2003). This interest has considerably increased since the September 11 terror attack in the United States and the subsequent terror activities in Europe (for a review of recent research, see Verschueren et al., 2011; Rosenfeld et al., 2012). Furthermore, the increased need to detect suspects involved in planning and executing terror activities has raised new questions that require new research directions. One of the main goals of this paper is to describe and discuss these new directions.

METHODS OF PSYCHOPHYSIOLOGICAL DETECTION

The various psychophysiological detection methods that have been developed can be broadly classified into two categories: (1)

Methods designed to detect deception, which rely on physiological responses to direct questions (e.g., "did you break into the Jewelry store on Thursday night?"); and (2) methods designed to detect concealed knowledge (e.g., "was the stolen jewel, a golden watch?", "was it a diamond ring?"). The detection method, most closely associated with the first category, has been labeled the Control (or more recently, comparison) Questions Technique (CQT). The CQT has been the preferable detection method used by law-enforcement agencies in the United States and it has been exported to various other countries. Yet, the CQT has been severely criticized and nowadays it is considered by most researchers as lacking scientific foundation (e.g., Ben-Shakhar, 2002; Iacono and Lykken, 2002; National Research Council, 2003). The major obstacle in any attempt to detect deception directly is that there is no specific and unique response associated with deception and under realistic police investigations both deceptive and honest suspects are highly aroused by the relevant ("Did you do it?") questions and thus may show similar physiological responses to these questions.

The method designed to detect concealed knowledge was traditionally labeled the guilty knowledge test (GKT, see Lykken, 1959, 1960), but more recently it has been referred to as the concealed information test (CIT, see Verschueren et al., 2011). This test utilizes a series of multiple-choice questions, each having one relevant alternative, also labeled as Probe (e.g., a feature of the crime under investigation) and several neutral (control) alternatives, chosen so that an innocent suspect would not be able to discriminate them from the probe (Lykken, 1998). The relevant alternatives are significant only for knowledgeable (guilty) individuals and there is ample evidence, mostly from psychophysiological research on orienting responses (ORs), indicating that significant stimuli elicit

enhanced ORs (e.g., Sokolov, 1963; Gati and Ben-Shakhar, 1990; Siddle, 1991). Thus, if the suspect's physiological responses to the relevant alternative are consistently larger than to the neutral (or irrelevant) alternatives, knowledge about the event (e.g., crime) is inferred. As long as information about the event has not leaked out to innocent suspects, the probability that an innocent suspect would produce consistently stronger responses to the relevant than to the neutral alternatives depends only on the number of questions and the number of alternative answers per question, and hence it can be controlled such that maximal protection for the innocent is provided. Clearly the detection of concealed information does not necessarily imply that the suspect is deceptive, as other explanations may be offered for the possession of guilty knowledge. Thus, deception or guilt can only be inferred indirectly and they require additional investigation. Although the CIT does rely on solid scientific grounds (e.g., Verschuere and Ben-Shakhar, 2011) it is very rarely used in practice in Western countries and in fact it is routinely used as the standard psychophysiological detection method only in Japan (see Osugi, 2011).

This paper will focus only on the CIT because it is the only psychophysiological method that is properly grounded in scientific research and theory. Both the strength and weaknesses of this technique will be briefly described as well as possible reasons for its limited usage. Finally, I will discuss current and future research directions as well as attempts to increase the usage of the CIT.

A BRIEF REVIEW OF CIT RESEARCH

Concealed information test research can be traced back to the early 1940s and 1950s (e.g., Geldreich, 1941, 1942; Ellson et al., 1952), but two articles published by Lykken (1959, 1960) were the first to make a real impact on the field and enhance interest in the CIT among various research groups. This early research relied on just a single physiological measure, namely skin conductance response (SCR) and demonstrated an impressive ability to detect concealed information. Specifically, Lykken (1959) employed a mock-crime procedure where some subjects committed one or two mock-crimes (the "guilty" subjects) while others (the "innocents") did not commit any. The results revealed that 88% of the "guilty" subjects were detected while none of the "innocent" subjects were misclassified as "guilty." Lykken's (1960) second study relied on a personal items paradigm and used 25 biographical details of 20 subjects, all of whom were correctly detected.

Concealed information test research has expanded in several directions in the following decades. First, the validity of additional autonomic measures, such as changes in respiration and heart rate, was examined (e.g., Thackray and Orne, 1968; Cutrow et al., 1972). For a recent review of CIT studies based on autonomic nervous system (ANS) measures, see Gamer (2011a). Furthermore, in the past two decades, much research interest has been devoted to the use of brain evoked potentials (see Rosenfeld, 2011 for a review) and brain imaging (see Gamer, 2011b; Rosenfeld et al., 2012) for the detection of concealed information. Second, attempts were made to shed light on the theoretical basis of the CIT effect – the enhanced responses elicited by the significant stimuli (e.g., Gustafson and Orne, 1963, 1965; Lieblich et al., 1970; Ben-Shakhar, 1977; Ben-Shakhar and Lieblich, 1982; Verschuere et al., 2004, 2007). Third, many studies examined the effects of various

factors on the outcomes of the CIT (e.g., the effect of type of verbal responses to the CIT questions, Kugelmass et al., 1967; Horneman and O'Gorman, 1985; the effect of drugs, Waid et al., 1981a; Iacono et al., 1984). Finally, factors that may limit the applicability of the CIT have been examined (e.g., the vulnerability of the CIT to countermeasures, Ben-Shakhar and Dolev, 1996; Honts et al., 1996; the effect of leakage of critical CIT items to innocent suspects, Bradley and Warfield, 1984; Bradley and Rettinger, 1992).

THE THEORETICAL FOUNDATION OF THE CIT

Recently, Verschuere and Ben-Shakhar (2011) reviewed the various theoretical approaches proposed to account for the enhanced autonomic responses to the relevant CIT alternatives. In this paper I will discuss only the main theoretical accounts. As the autonomic measures used in the CIT are components of the OR (see Sokolov, 1963; Lynn, 1966), it is not surprising that this concept has been proposed to account for the CIT effect. Furthermore, Sokolov (1963) and his followers noted that significant stimuli ("signal-value stimuli," to use Sokolov's terminology) elicit enhanced ORs with slower habituation and this can account for the enhanced responses to the crime-relevant stimuli observed among knowledgeable (guilty) individuals. The relationship between the CIT effect and OR was highlighted by Lykken (1974) who wrote that, "... for the guilty subject only, the 'correct' alternative will have a special significance, an added 'signal value' which will tend to produce a stronger orienting reflex than that subject will show to other alternatives (p. 728)."

There is ample evidence supporting the OR account for the CIT effect. First, the physiological response pattern elicited by the relevant CIT items in knowledgeable individuals (e.g., increased SCR, Lykken, 1959; heart-rate deceleration, Verschuere et al., 2004; respiratory suppression, Timm, 1982; and increased pupil dilation, Lubow and Fein, 1996) is typical for the OR. Second, several features characteristic of the OR have been demonstrated, using the CIT paradigm. For example, response habituation has been observed in several CIT studies (e.g., Ben-Shakhar et al., 1975; Balloun and Holmes, 1979; Verschuere et al., 2005). In addition, as predicted by OR theory, the CIT effect has been demonstrated to increase when the critical items are less frequently presented (e.g., Ben-Shakhar, 1977). Forth, the information processing view of orienting states that the OR serves to allow more elaborate processing of the OR-eliciting stimulus (Kahneman, 1973; Wagner, 1978; Öhman, 1992). Research demonstrating positive correlations between OR and later recall of the stimulus material supports this view (e.g., Corteen, 1969). Indeed, several CIT studies found a positive association between recall and detection efficiency (e.g., Waid et al., 1978, 1981b; Iacono et al., 1984; Carmel et al., 2003; Verschuere et al., 2007).

On the other hand, some research findings are hard to reconcile with the OR theory. For example, heart-rate deceleration elicited by relevant CIT items may last for 15 s, whereas according to OR theory heart rate typically decelerates 1–5 s after the onset of the OR-eliciting stimulus, and then returns to baseline (Richards and Casey, 1992). In addition, although OR theory predicts greater startle modulation to the relevant than to the irrelevant items, Verschuere et al. (2007) failed to support this prediction and proposed an alternative hypothesis, namely response inhibition, to explain

the startle data. Processes other than orienting may contribute to physiological responding in the CIT, and response inhibition seems a reasonable candidate. This account is also supported by recent fMRI research (see Gamer et al., 2007).

THE VALIDITY OF THE CIT

Although the initial studies reported by Lykken (1959, 1960) produced impressive validity estimates for the CIT based on SCR, the results of subsequent studies that used both SCR and additional ANS measures were less uniform. The best method for evaluating research results across many studies is meta analysis (e.g., Hunter and Schmidt, 1990). Indeed two meta analytic studies published last decade (MacLaren, 2001; Ben-Shakhar and Elaad, 2003) demonstrated a relatively large mean effect size (Cohen's d) for the CIT based on SCRs. For example, Ben-Shakhar and Elaad (2003) covered 80 laboratory studies, which included 169 experimental conditions with a total of 5198 participants tested under a variety of CIT paradigms (e.g., card test, mock-crime) and reported an overall average effect size of 1.55. They further showed that studies relying on the mock-crime paradigm, which seems more relevant for field applications than other paradigms, produced an average effect size of 2.1.

However both meta analyses relied only on a single measure and as more studies using additional measures were published during the last decade, it is more informative to describe a more recent meta analysis that included four measures (Meijer et al., 2012). In addition to SCR, this meta analysis included studies that measured respiration line length (RLL, see e.g., Timm, 1987), heart-rate deceleration (e.g., Ambach et al., 2011), and the P300 component of the event-related potential (e.g., Rosenfeld et al., 1988; Farwell and Donchin, 1991). Meijer et al. (2012) included in their meta analysis two CIT paradigms (the mock-crime and the personal items paradigms) and several measures of detection efficiency. In addition to the average Cohen's d they computed the variance of d across studies and subtracted from it the variance that would be expected from sampling errors. The residual variance represents true differences among the studies.

The main results of this meta analysis indicated that the four measures differ significantly in their detection efficiency. Specifically, the P300 measure outperformed all three ANS measures, with an average d of 2.55, but it should be noted that 80% of the P00 studies, included in this meta analysis, came from a single laboratory (of J. P. Rosenfeld) that has been most active in the past two decades. The HR measure was the least effective of all four measures that have been examined (with an average d of 0.88), but it is important to note that even this d value is considered a large effect size (see Cohen, 1988). Moreover, several studies demonstrated that a combination of several ANS measures outperforms the best single measure (e.g., Ben-Shakhar and Dolev, 1996; Ben-Shakhar and Elaad, 2002; Gamer et al., 2008) and from this respect the d value of 1.73 obtained for the SCR measure can be considered as an underestimate of detection efficiency with ANS measures. The results of this meta analysis also revealed a considerable residual variance for the SCR and the P300, which means that real differences between studies exist for these measures.

Indeed, several moderating factors were identified for the SCR measure. Specifically, two factors that were also identified by

Ben-Shakhar and Elaad (2003), namely motivation to avoid detection and the number of CIT questions mediate the SCR effect size. In experimental conditions that employed either incentive or instructions to avoid detection, the average d was 1.89 as compared with an average of 1.45 observed under law motivation conditions. In addition, when the number of CIT questions used was at least six the average d was 1.99 as compared with 1.45 when a smaller number of questions were used. These two moderators may be very important for the application of the CIT because real-life investigations are clearly associated with very high levels of motivation to avoid detection and because investigators can increase detection efficiency by making efforts to identify as many appropriate critical items as possible. The number of ERP studies (32) was too small to allow for an analysis of moderators. In addition, motivation was not manipulated in ERP studies and the number of questions used was more or less uniform.

However, although this meta analysis, as well the previous meta analytic studies (MacLaren, 2001; Ben-Shakhar and Elaad, 2003), demonstrated very large effect sizes, it should be emphasized that only laboratory experiments were analyzed and it is questionable whether the results of CIT experiments can be generalized to realistic criminal investigations. Unfortunately, only two field CIT studies were reported in the scientific literature (Elaad, 1990; Elaad et al., 1992). The results of these studies, which were based on criminal cases investigated by the Israeli Police, showed that while the rates of false-positive errors were as low as those reported in laboratory experiments (2% in the former study, which relied only on the electrodermal measure, and 5% in the latter study, which utilized a combination of electrodermal and respiration measures), the rates of false-negative errors were much larger (42% in the former study and 20% in the latter). This may imply that CIT experiments have a weak external validity, but it should be noted that the use of the CIT in the criminal cases studied by Elaad (1990) and Elaad et al. (1992) was not optimal. In particular, the mean number of questions used in these field studies (2 and 1.8 in Elaad, 1990 and Elaad et al., 1992, respectively), was much smaller than recommended. In addition, the two field studies were based on CITs that were administered immediately after a CQT, and this might attenuate the sensitivity of the physiological measures due to habituation. Thus, it is possible that the relatively high rates of false-negative errors and lower detection efficiency obtained in these field studies resulted from a non-optimal usage of the CIT.

WEAKNESSES AND POTENTIAL LIMITATIONS OF THE CIT

So far, I have listed several advantages of the CIT over alternative detection methods, namely its solid theoretical foundation, the impressive validity estimates obtained for the CIT in experimental settings and its potential for protecting innocent suspects against false classification. Unfortunately, the CIT has several weaknesses and in this section I will discuss factors that may limit its application. As indicated above, the bulk of CIT studies were conducted in artificial laboratory settings where volunteering participants were requested to commit a mock-crime, with no consequences for their well-being. It is important therefore, to examine the factors that differentiate the experimental setting from real criminal investigations.

LEAKAGE OF CRITICAL ITEMS

Implementation of the CIT depends on a successful concealment of the critical items. Whereas in mock-crime studies concealment is perfectly guaranteed, in real-life this is not necessarily the case and critical items may leak to innocent suspects, either through the media, or during the course of police interrogations.

Several studies examined the effect of information leakage on the CIT accuracy and particularly on false-positive outcomes. Most of these studies were conducted by Bradley and his colleagues (Bradley and Warfield, 1984; Bradley and Rettinger, 1992; Bradley et al., 1996; see Bradley et al., 2011, for a recent review of the leakage literature). Generally, these studies demonstrated that although informed innocent participants show larger relative responses to the critical items, as compared with uninformed innocents, they could be differentiated from guilty participants. However, two recent studies demonstrated that informed innocents were not differentiated from guilty participants when the CIT was administered immediately after the mock-crime (Gamer et al., 2010; Nahari and Ben-Shakhar, 2011). But when the test was delayed (as is usually the case in realistic criminal investigations), informed innocents showed smaller differential responses to the critical items, as compared with guilty participants. This was mediated in both studies by the fact that informed innocents forgot critical items more than guilty participants.

Several means to reduce the damaging effects of information leakage (in addition to improving police practices) were examined by some researchers. Ben-Shakhar et al. (1999) used target items to which participants had to respond in addition to the critical and the control items. Under this procedure, the rate of false-positive outcomes among informed innocents was somewhat reduced.

Bradley and Warfield (1984) proposed a modified version of the CIT, labeled the guilty action test (GAT), in which the formulation of the questions emphasize actions rather than knowledge (e.g., "Did you kill Mr. X with a gun?, knife? . . .," rather than "Was Mr. X killed with a gun?, knife? . . ."). Under the GAT guilty suspects are deceptive when giving negative answers to these questions, whereas informed innocents are telling the truth. Bradley et al. (1996) directly compared the CIT and the GAT and showed that the GAT significantly reduced the false-positive rates, although these rates were still very high (50%). On the other hand, a more recent study by Gamer (2010) failed to find any differences between the two test formats: In both formats informed innocents were undifferentiated from guilty participants.

Previewing the CIT questions has also been offered as a means to prevent the usage of items that might have leaked. Presenting the CIT questions prior to the test may provide examinees with an opportunity to explain that they are familiar with certain items (e.g., they were mentioned in prior interrogations). Verschueren and Crombez (2008) demonstrated that previewing CIT items does not reduce the test's validity. Clearly, leakage of critical information is a major threat to the validity of the CIT and the test should not be used when critical items were leaked. No information is available about the extent to which critical items are being leaked in police investigations, but the results of the two field studies reported by Elaad and his colleagues (Elaad, 1990; Elaad et al., 1992) were encouraging with this respect, as in both studies the false-positive

rates were small, indicating that at least in these criminal cases critical information did not leak.

THE EFFECTS OF COUNTERMEASURES

While leakage of critical information may affect false-positive rates, other factors that can increase false-negatives were also identified in previous research. Specifically, several studies demonstrated that the CIT is vulnerable to countermeasures, namely deliberate techniques that might be used by suspects to alter their physiological reactions in order to avoid detection. Several countermeasure techniques have been experimentally examined (e.g., Kubis, 1962; Elaad and Ben-Shakhar, 1991; Ben-Shakhar and Dolev, 1996; Honts et al., 1996; see a recent review of the countermeasure literature in Ben-Shakhar, 2011), but countermeasures were most effective when subjects attempted to create or enhanced responses to the neutral items. This can be achieved either by physical (subjects can bite their tongue to inflict pain when the control items are presented) or by mental means (recalling exciting and emotional memories, or exercising mental activities during presentation of control items). Mental countermeasures may be most harmful because they cannot be detected by the examiners. Two studies examined the effects of mental countermeasures on the outcomes of the CIT (Ben-Shakhar and Dolev, 1996; Honts et al., 1996) and demonstrated a significant reduction in SCR detection efficiency when these countermeasures were applied. However, no countermeasure effects were observed in these studies when the RLL was used as the detection measure.

Clearly, both physical and mental countermeasures require some sophistication and certain knowledge. However, there is an extensive literature in which ANS-based polygraph procedures including effective countermeasure techniques are described in great detail. Thus, the danger that interested individuals may gain the necessary understanding in order to use countermeasures is a real one.

Several researches reported that even CIT based on the P300 component of the event-related potential may be vulnerable to countermeasures (e.g., Rosenfeld et al., 2004; Mertens and Allen, 2008). To overcome this difficulty, Rosenfeld et al. (2008) developed a novel P300 protocol called the *Complex Trial Protocol* which temporally separates the presentation of probe or irrelevant from target or non-target. Several studies reported by Rosenfeld and his colleagues demonstrated that this protocol was indeed highly resistant against both mental and physical countermeasures (Rosenfeld et al., 2008; Meixner and Rosenfeld, 2010; Rosenfeld and Labkovsky, 2010; Winograd and Rosenfeld, 2011). Clearly these studies should be replicated in other laboratories, but they indicate that CIT based ERPs may be immune against countermeasures and as ERPs are associated with very large effect size (see Meijer et al., 2012) they may have an excellent potential as an applied detection method.

THE ROLE OF PERCEPTION AND MEMORY OF CRIME-RELATED ITEMS ON CIT VALIDITY

A successful implementation of the CIT in the criminal investigation context depends on the identification of a sufficient number of salient features of the crime, features that are likely to be

noticed by the perpetrator and stored in memory. Unfortunately, the bulk of CIT research has been conducted in artificial settings where it was guaranteed that participants memorized all critical features of a mock-crime. Furthermore, the CIT is typically administered immediately after participants committed the mock-crime, whereas in realistic criminal investigations polygraph tests are administered after a relatively long delay. Thus, the external and ecological validity of mock-crime studies seem highly questionable. Recently, three studies examined the role of memory for critical items on the CIT's outcomes (Carmel et al., 2003; Gamer et al., 2010; Nahari and Ben-Shakhar, 2011). These studies revealed that when the CIT is administered one or two weeks after the mock-crime, certain critical items are not recalled and do not elicit differential responses. However, consistent with memory research (e.g., Kensinger, 2007), memory loss occurs mostly with peripheral items (features that are not directly related to the execution of the crime, such as a picture on the wall of the crime scene). Central features, such as type of weapon used, are capable of eliciting large responses even when the test is delayed. Clearly, this line of research that has important practical implications for constructing proper CITs should be continued and extended.

THE EFFECTS OF EMOTIONAL STRESS AND MOTIVATION ON CIT VALIDITY

Another important difference between the typical experimental setup and realistic criminal investigations is the level of stress experienced by the examinees as well as their motivation to avoid detection. However, there are several indications that these factors are not interfering with the external validity of CIT experiments. First, as indicated above, motivation to avoid detection was manipulated in several studies and was generally associated with an increased CIT effect (Ben-Shakhar and Elaad, 2003; Meijer et al., 2012). Thus from this perspective, it seems that the CIT should have even larger detection efficiency in realistic investigations than in laboratory experiments.

Second, two studies (Kugelmass and Lieblich, 1966; Bradley and Janisse, 1981) manipulated the level of stress experienced by subjects while taking the CIT and included levels that seem to resemble realistic situations. Both studies demonstrated that the level of stress had no effect on the outcomes of the CIT. It was concluded that, "within a considerable range of stress no necessary decrease in the detection efficiency of the GSR channel need be expected" (Kugelmass and Lieblich, 1966, p. 215). Thus, on the basis of these two studies it seems that detection efficiency estimated in laboratory experiments can be generalized to situations characterized by much higher levels of motivation and stress.

Third, recently Peth et al. (2012) manipulated the level of stress during mock-crime execution and found that level of stress did not affect the relative responses to the critical CIT items with electrodermal, respiration, and cardiovascular measures. Furthermore, the data revealed that under the high arousal level, detection efficiency based on central items tended to be unaffected by delaying the test. The authors concluded that, "emotional arousal might facilitate the detection of concealed information sometime after the crime occurred" (Peth et al., 2012, p. 381).

CURRENT USAGE OF THE CIT IN PRACTICE

As mentioned above, despite its many advantages, the CIT is hardly used in criminal investigations in the West, whereas the much more controversial, CQT is used extensively in the United States and several other countries. The limitations of the CIT, listed in the previous section have been offered as an explanation for this state of affairs. Krapohl (2011) discussed various factors that limit the applicability of the CIT and classified them into two categories, practical and cultural limitations. The practical factors relate to the difficulty in identifying a sufficient number of salient features of a crime and protecting them from leaking as well as the vulnerability of the CIT to countermeasures (although the CQT is as vulnerable to countermeasures as the CIT, e.g., Honts et al., 1994). Podlesny (1993) made similar arguments and estimated that the CIT might have been used in only 13.1% of FBI cases for which polygraphs have been used. This estimate is based on the assumption that at least four different CIT questions are required to construct a CIT.

However, it is difficult to reconcile these arguments and estimates with the fact that the CIT has been used for many decades by the Japanese police as the standard polygraph method. Approximately 5000 CITs are administered annually in Japan and this method has even been used as admissible evidence in the Japanese criminal courts (Hira and Furumitsu, 2002; Nakayama, 2002; Osugi, 2011). Therefore, it seems more reasonable that the cultural factors may provide a better explanation for these differences in the application of the CIT. Indeed, Krapohl (2011) suggested that even if the practical difficulties were resolved, "the expanded use of the CIT would still face resistance from some experienced polygraph examiners who, wedded to the methods they learned in polygraph school, find such a radical departure from the CQT protocol unsettling and unnecessary" (Krapohl, 2011, p. 160). He added that only 5 out of the 20 certified polygraph schools in the U.S. formally teach the CIT.

There is a huge gap between scientists and practitioners in this area and while the bulk of the scientific community regard the CQT as a non-scientific method, most practitioners believe it is highly accurate. A possible explanation for this gap was offered by Ben-Shakhar (1991) who argued that the belief of practitioners in the validity of the CQT reflects a biased decision process. Specifically, polygraph examiners are affected by the confirmation bias (e.g., Nisbett and Ross, 1980; Darley and Gross, 1983) when they administer the CQT and evaluate the physiological responses. As a result, the outcomes of the CQT are typically consistent with the examiners' *a priori* hypotheses and this creates a strong illusion of validity (see Einhorn and Hogarth, 1978). In addition, the CQT is often used to extract confessions (Furedy and Liss, 1986), and naturally investigators make efforts to extract confessions only when they believe that the suspect is guilty. Thus, confessions made after a CQT are typically associated with an incriminating CQT's outcome (Iacono, 1991) and this is another factor that contributes to the illusion of validity. Finally, Western practitioners may have been influenced by the positive results of controlled mock-crime experiments that generally supported the CQT's validity (although their weak external validity does not allow for generalizing their results, see Ben-Shakhar, 2002).

In addition to the strong belief of polygraph examiners in the CQT's validity, it should be noted that it is much easier to formulate

CQT questions than to identify salient features of a crime and as the CQT is a test of deception, it can be used in all types of criminal cases. Thus, practitioners in most countries do not feel that the CQT need to be replaced.

FUTURE DIRECTIONS IN RESEARCH AND PRACTICE

THE NEED FOR FIELD-VALIDITY STUDIES

In the previous section I discussed several factors differentiating the artificial experimental setting from that of realistic criminal investigations. Clearly, the best approach would be to examine the validity of the CIT as practiced with real suspects. However, as indicated above only two field-validity studies were published so far (Elaad, 1990; Elaad et al., 1992). This unfortunate situation may be explained by the difficulties involved in conducting proper field studies in this area. Specifically, a ground truth criterion is typically unavailable and the use of confessions is problematic because they may depend on the test's outcomes (see Iacono, 1991). Nevertheless, efforts must be made to overcome these difficulties and the natural setting for such studies seems to be Japanese criminal investigations arena because the CIT is the standard polygraph method used in Japan and because Japanese polygraph investigators have the proper scientific training (Osugi, 2011). The application of the CIT by Japanese Police meets very high standards. Specifically, it typically rests on five different questions (as opposed to an average of about two in the Israeli Police studies), each repeated five times and on four physiological measures (as opposed to one or two in the Israeli studies). Furthermore, from the description of how the CIT is conducted by the Japanese Police (Osugi, 2011), it seems that CITs are conducted independently of other criminal investigations and it is not used as a means to elicit confessions. Such studies would shed light on the validity of the CIT in practice.

EXAMINING ADDITIONAL PHYSIOLOGICAL AND BEHAVIORAL MEASURES

The use of brain imaging in the CIT

The validity of additional measures that can be incorporated into the CIT may also be important. A great research interest has recently been directed to the use of brain imaging for the detection of deception. These studies used a variety of research paradigms and were focused primarily on the search of brain regions that are differentially activated when subjects give deceptive versus truthful responses. For example, the differentiation of deception (DOD) paradigm, designed to isolate deception and examine processes associated with deception (e.g., Furedy et al., 1988), was often used in fMRI research. Other studies used variations of the CIT paradigm (primarily, the card test and the personal item paradigm), to examine brain activation when critical information is concealed. The results based on group data were not uniform and even studies using similar experimental procedures failed to fully replicate their findings, but most studies found regions in the prefrontal cortex being more activated when deceiving or concealing knowledge (see recent reviews by Gamer, 2011b; Rosenfeld et al., 2012). These studies are important from a theoretical perspective as they may shed light on brain mechanisms associated with deception, but from a practical perspective it is important to examine the efficiency of fMRI in classifying individuals as concealing critical

information. Only very few studies assessed the validity of the CIT with fMRI. The results of these studies, as summarized by Rosenfeld et al. (2012), indicate that the average sensitivity and specificity were, about 86 and 92%, respectively. These figures are more or less similar to those obtained with ERPs (Meijer et al., 2012) and also to those obtained with a combination of ANS measures (see Gamer et al., 2008). Thus, given the complexity of fMRI measurement relative to ANS and ERP measures, it is highly questionable whether fMRI would have a practical utility as a field detection method. In addition, detection of concealed information with fMRI is vulnerable to all the threats mentioned earlier and the generalizability of the few published studies is questionable. For example, Ganis et al. (2011) demonstrated that when subjects applied countermeasures CIT detection accuracy with fMRI dropped from 100 to only 33%.

The use of behavioral measures

Several behavioral measures can be used for detecting concealed information with the CIT, but these measures have received relatively little research attention and definitely should be more thoroughly explored. Examining response latency (or response time-RT) to critical and neutral items is a natural candidate for providing useful information that can distinguish between knowledgeable and unknowledgeable (innocent) individuals because significant stimuli capture attention and thus require more processing time. Indeed, RT has been included in many ERP studies using the oddball paradigm (e.g., Farwell and Donchin, 1991) and showed the expected effect (enhanced RTs to critical items among knowledgeable participants). Moreover, Allen et al. (1992) reported a slightly better performance of the behavioral measures (response time and number of errors) as compared with the ERP measures. Seymour et al. (2000) were the first to examine RTs as a sole index for information concealment and concluded that RTs can serve as a simple alternative to the physiological measures typically used in the CIT. However, the question of whether RTs have incremental validity over ANS or ERP measures has not been resolved yet and studies using different paradigms produced different results (e.g., Gronau et al., 2005; Verschuere et al., 2009). In their review of the research on the use of RTs in the CIT, Verschuere and De Houwer (2011) argued that paradigms based on a manipulation of relevant stimulus-response compatibility, such as the oddball task are effective, whereas tasks that do not manipulate relevant stimulus-response compatibility, such as the modified Stroop used by Gronau et al. (2005) have not produced robust response latency differences between concealed and control items. Clearly, this is an important hypothesis that deserves further research. Similarly, the vulnerability of RT to countermeasure manipulations should be thoroughly examined.

The symptom validity test

This test may be promising because it is based on an entirely different rationale than that underlying both physiological and RT measures. Specifically, the SVT is based on asking examinees, who deny knowledge of the critical items, to guess these items. Effective concealment is possible when guessing is random (i.e., where the critical alternative is guessed with the same probability as all other alternatives), but producing random guesses may be very difficult for those who are actually aware of the

true alternatives. Consequently the outcome of multiple guessing attempts may differentiate knowledgeable (who would not be able to produce random guessing) and unknowledgeable examinees (whose guesses will be random). The SVT has been used to detect malingering in various contexts (e.g., Merckelbach et al., 2002; Verschueren et al., 2008) and recently it has been adopted for the CIT (Meijer et al., 2007; Nahari and Ben-Shakhar, 2011). These studies demonstrated that the SVT can improve detection efficiency when combined with ANS measures. Once again, much more research is required to determine the practical utility of the SVT.

THE POTENTIAL USE OF THE CIT IN THE ANTI-TERROR CAMPAIGN

The increased terror activities during the last decade have raised an increased interest in detection methods in general, and particularly in the CIT. The use of the CIT to detect individuals and groups involved in terror activities has raised new questions. First, suspects in terror activities are often being interrogated about their plans, rather than about crimes already committed. Thus, one question that deserves careful research is whether detecting past *actions* is equivalent to detecting future *intentions*. Two initial studies have already examined this question. Meijer et al. (2010a) conducted a systematic comparison between committing a mock-crime and planning a mock-crime. These authors demonstrated that the CIT with the SCR measure was similarly effective in both conditions, suggesting that the CIT can be used to detect mal-intentions. This conclusion is also supported by recent findings reported by Meixner and Rosenfeld (2011) showing impressive detection efficiency of the P300-based-CIT with participants who planned a mock terrorist attack. Clearly, this line of research should be continued and elaborated.

A second, related question is whether the CIT can be applied to cases where the precise details are *not* available to the investigators. For example, the Japanese Police applies the CIT in some cases to retrieve information that is unavailable to the investigators (e.g., finding the location of a murder weapon). This application of the CIT, termed “the Searching CIT” (SCIT), is described in detail by Osugi (2011). The SCIT may be applied in the anti-terror campaign. For example, imagine a terrorist group planting a bomb in a certain location unknown to the investigators. Can this location be detected when suspects are identified and tested using the SCIT, to prevent an upcoming explosion? Clearly the use of the SCIT requires some *a priori* knowledge (e.g., possible terror targets) and therefore can be applied only when some intelligence information is available to the investigative authorities. Although the SCIT is being used by the Japanese Police, research examining the validity of this method, as used in Japan is unavailable.

However, initial research on the SCIT has recently emerged. Meijer et al. (2010b) examined the SCIT with the electrodermal measure. They tested 12 participants, who were informed about the details of a planned terror attack, where these details were not known to the investigator (though it was assumed that the terror-related details are among the different alternatives included in the test). Relying upon group averages, these researchers were able to identify the correct alternative in each of the three SCIT questions used. However, this study is of limited external validity because all participants were exposed to the critical items, whereas in most real-life cases, some suspects may be innocent (unaware

of the critical items). For example, in the terror attack example, some suspects may be only partially aware of the critical information, or they may be innocent altogether (not belonging to the terror organization). Therefore, it is important to test the SCIT validity under conditions in which suspects’ status (i.e., knowledgeable or unknowledgeable) is unknown to the investigator. Meixner and Rosenfeld (2011) were the first who examined the SCIT with both “guilty” and “innocent” participants. This study used the P300 component of the event-related brain potentials and compared the largest average P300 amplitude of each participant with the second largest response. Detection was made at the individual participant level and 10 out of the 12 knowledgeable participants were correctly detected with no false positives. This yielded an area under the receiver operating characteristic (ROC) curve of 0.979. Additionally, 58% (21 out of 36) critical items were correctly detected.

A different approach was recently adopted by Breska et al. (2012) who examined several algorithms designed to detect the critical items as well as differentiate between knowledgeable and unknowledgeable participants in the SCIT. They reanalyzed three data sets from previous, published CIT studies, assuming that the critical items are unknown to the investigators, but are included among the alternative items presented to the subjects. Specifically, they examined two classes of algorithms. The first class was based on averaging responses across subjects to identify critical items and then on averaging responses across the identified critical items to identify knowledgeable subjects. The second class was based on the correlations between the response profiles of all subject-pairs and applied a principle component analysis to decompose the correlation matrix into its principal components. The detection score was defined as the coefficient of each subject on the component explaining the largest portion of the variance. The results revealed that in most cases all critical items were correctly identified and the efficiency of differentiation between knowledgeable and unknowledgeable subjects in the SCIT (indexed by the area under the ROC curve) approached that of the standard CIT, for both classes of algorithms. In addition, the robustness of these results to variations in the number of knowledgeable and unknowledgeable subjects in the sample was examined. This analysis demonstrated that the performance of these algorithms is relatively robust to changes in the number of individuals examined in each group, provided that at least two (but desirably five or more) knowledgeable examinees are included. Although these results seem promising, the validity of the SCIT should be examined in new experiments involving groups planning illegal activities.

CONCLUSION AND RECOMMENDATIONS

This paper focused on the CIT and discussed its strength and weaknesses as well as several new potential applications of this method and future research directions. The limited application of the CIT was explained by several practical factors related to its weaknesses and by cultural factors. As the CIT seems to be the only scientifically based detection method, with impressive validity estimates observed in controlled, laboratory studies, it is important to suggest ways to overcome its difficulties and expand its usage. Thus, in this final section I will list several recommendations that may enhance the applicability of the CIT.

1. *Identifying a sufficient number of salient crime-features:* Recent research suggested that the CIT performs best with central features of the crime, especially when the test is delayed (Carmel et al., 2003; Gamer et al., 2010; Nahari and Ben-Shakhar, 2011). This poses a great challenge because it has been suggested that at least five different CIT questions should be formulated (Lykken, 1988; Ben-Shakhar and Elaad, 2003). Two approaches may be offered to overcome this difficulty. First, although multiple questions are definitely desired, two studies demonstrated that the CIT can be successfully used with much fewer questions, and even with a single question, provided that questions are repeated several times and that a combination of several physiological measures is used (Elaad and Ben-Shakhar, 1997; Ben-Shakhar and Elaad, 2002). Second, the criminal investigation process should be modified, such that polygraph examiners would be able to inspect the crime scene soon after a crime was committed, as practiced by the Japanese National Police.
2. *Protecting critical items and preventing leakage:* Although the results of the field studies reported by Elaad and his colleagues (Elaad, 1990; Elaad et al., 1992) suggest that leakage of crime-related information did not affect the results of CITs administered by the Israeli police, preventing leakage is essential for a wide application of the CIT. Some research results described earlier (Bradley et al., 1996; Ben-Shakhar et al., 1999) offered methods to reduce the effects of information leakage. However, even with these methods false-positive outcomes among knowledgeable innocent subjects were too high to tolerate. Thus, it seems that the only solution to this problem is to modify police practices, such that critical features of the event are identified and concealed at the outset of the investigation, as a standard investigative practice and that the CIT questions will be previewed by the suspects.
3. *Dealing with countermeasures:* A possible approach for dealing with countermeasure manipulations is the use of the CIT with event-related potentials, rather than autonomic measures. Although initial studies suggested that ERPs are vulnerable to countermeasures (Mertens and Allen, 2008; Rosenfeld et al., 2004), more recent studies using the complex trial protocol showed impressive detection efficiency both when participants applied physical and mental countermeasures and under a non-countermeasure condition (Rosenfeld et al., 2008; Meixner and Rosenfeld, 2010; Rosenfeld and Labkovsky, 2010). In addition, it is important to note that detection efficiency with ERP measures have been demonstrated to be significantly better than that obtained with ANS measures (Meijer et al., 2012). A different approach for dealing with countermeasures was adopted by Elaad and his colleagues who examined several covert respiration measures, with the idea that examinees who are unaware of the fact that they are connected to a polygraph will not be motivated to apply countermeasures (e.g., Elaad and Ben-Shakhar, 2008). However, this idea raises ethical questions that may severely limit or even prohibit its use (for a review of research on covert measures, see Elaad, 2011). More recently, two studies examined whether the CIT can be applied when the questions are presented subliminally and masked (Lui and Rosenfeld, 2009; Maoz et al., 2012). The rationale is similar to the use of covert measures, but it is unclear whether the potential advantage of using invisible stimuli in combatting countermeasures, outweighs the cost of reducing detection efficiency as observed by Maoz et al. (2012) under subliminal presentation conditions.
4. *Future research directions:* Clearly, all the above recommendations require additional research. For example, the complex trial protocol should be further examined in various laboratories. Similarly, the idea that memory of central crime details is stable over time and unaffected by emotional stress needs further research. Finally, it is essential to examine these factors under realistic conditions, with real criminal suspects.

ACKNOWLEDGMENTS

This research was funded by a grant from the Israel Science Foundation to Gershon Ben-Shakhar. I am grateful to Ewout Meijer for his constructive comments.

REFERENCES

- Allen, J., Iacono, W. G., and Danielson, K. D. (1992). The identification of concealed memories using the event-related potential and implicit behavioral measures – a methodology for prediction in the face of individual-differences. *Psychophysiology* 29, 504–522.
- Ambach, W., Stark, R., and Vaitl, D. (2011). An interfering n-back task facilitates the detection of concealed information with EDA but impedes it with cardiopulmonary physiology. *Int. J. Psychophysiol.* 80, 217–226.
- Baloun, K. D., and Holmes, D. S. (1979). Effects of repeated examinations on the ability to detect guilt with a polygraph examination: laboratory experiment with a real crime. *J. Appl. Psychol.* 64, 316–322.
- Ben-Shakhar, G. (1977). A further study of the dichotomization theory in detection of information. *Psychophysiology* 14, 408–413.
- Ben-Shakhar, G. (1991). Clinical judgment and decision-making in CQT-polygraphy. *Integr. Psychol. Behav. Sci.* 26, 232–240.
- Ben-Shakhar, G. (2002). “A critical review of the control questions test (CQT),” in *Handbook of Polygraph Testing*, ed. M. Kleiner (London: Academic Press), 103–126.
- Ben-Shakhar, G. (2011). “Countermeasures,” in *Memory Detection: Theory and Application of the Concealed Information Test*, eds B. Verschueren, G. Ben-Shakhar, and E. Meijer (Cambridge: Cambridge University Press), 200–214.
- Ben-Shakhar, G., and Dolev, K. (1996). Psychophysiological detection through the guilty knowledge technique: effects of mental countermeasures. *J. Appl. Psychol.* 81, 273–281.
- Ben-Shakhar, G., and Elaad, E. (2002). Effects of questions’ repetition and variation on the efficiency of the guilty knowledge test: a reexamination. *J. Appl. Psychol.* 87, 972–977.
- Ben-Shakhar, G., and Elaad, E. (2003). The validity of psychophysiological detection of deception with the Guilty Knowledge Test: a meta-analytic review. *J. Appl. Psychol.* 88, 131–151.
- Ben-Shakhar, G., and Furedy, J. J. (1990). *Theories and Applications in the Detection of Deception: A Psychophysiological and International Perspective*. New York: Springer-Verlag.
- Ben-Shakhar, G., Gronau, N., and Elaad, E. (1999). Leakage of relevant information to innocent examinees in the GKT: an attempt to reduce false-positive outcomes by introducing target stimuli. *J. Appl. Psychol.* 84, 651–660.
- Ben-Shakhar, G., and Lieblich, I. (1982). The dichotomization theory for differential autonomic responsivity reconsidered. *Psychophysiology* 19, 277–281.
- Ben-Shakhar, G., Lieblich, I., and Kugelmass, S. (1975). Detection of information and GSR habituation: an attempt to derive detection efficiency from two habituation curves. *Psychophysiology* 12, 283–288.
- Bradley, M. T., Barefoot, C. A., and Arsenault, A. M. (2011). “Leakage of information to innocent suspects,” in *Memory Detection: Theory and Application of the Concealed Information Test*, eds B. Verschueren, G. Ben-Shakhar, and E. Meijer (Cambridge: Cambridge University Press), 187–199.

- Bradley, M. T., and Janisse, M. P. (1981). Accuracy demonstration, threat, and the detection of deception: cardiovascular, electrodermal, and pupillary measures. *Psychophysiology* 18, 307–315.
- Bradley, M. T., MacLaren, V. V., and Carle, S. B. (1996). Deception and nondeception in guilty knowledge and guilty action polygraph tests. *J. Appl. Psychol.* 81, 153–160.
- Bradley, M. T., and Rettinger, J. (1992). Awareness of crime-relevant information and the guilty knowledge test. *J. Appl. Psychol.* 77, 55–59.
- Bradley, M. T., and Warfield, J. F. (1984). Innocence, information, and the guilty knowledge test in the detection of deception. *Psychophysiology* 21, 683–689.
- Breska, A., Ben-Shakhar, G., and Gronau, N. (2012). Algorithms for detecting concealed knowledge among groups when the critical information is unavailable. *J. Exp. Psychol. Appl.* (in press).
- Carmel, D., Dayan, E., Naveh, A., Raveh, O., and Ben-Shakhar, G. (2003). Estimating the validity of the guilty knowledge test from simulated experiments: the external validity of mock crime studies. *J. Exp. Psychol. Appl.* 9, 261–269.
- Cohen, J. E. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum.
- Corteen, R. S. (1969). Skin conductance changes and word recall. *Br. J. Psychol.* 60, 81–84.
- Cutrow, R. J., Parks, A., Lucas, N., and Thomas, K. (1972). The objective use of multiple physiological indices in the detection of deception. *Psychophysiology* 9, 578–587.
- Darley, J. M., and Gross, P. H. (1983). A hypothesis-confirming bias in labeling effects. *J. Pers. Soc. Psychol.* 44, 20–33.
- Einhorn, H. J., and Hogarth, R. M. (1978). Confidence in judgment: persistence of the illusion of validity. *Psychol. Rev.* 85, 395–416.
- Elaad, E. (1990). Detection of guilty knowledge in real-life criminal investigations. *J. Appl. Psychol.* 75, 521–529.
- Elaad, E. (2011). “New and old covert measures in the concealed information test,” in *Memory Detection: Theory and Application of the Concealed Information Test*, eds B. Verschueren, G. Ben-Shakhar, and E. Meijer (Cambridge: Cambridge University Press).
- Elaad, E., and Ben-Shakhar, G. (1991). Effects of mental countermeasures on psychophysiological detection in the guilty knowledge test. *Int. J. Psychophysiol.* 11, 99–108.
- Elaad, E., and Ben-Shakhar, G. (1997). Effects of items’ repetitions and variations on the efficiency of the guilty knowledge test. *Psychophysiology* 34, 587–596.
- Elaad, E., and Ben-Shakhar, G. (2008). Covert respiration measures for the detection of concealed information. *Biol. Psychol.* 77, 284–291.
- Elaad, E., Ginton, A., and Jungman, N. (1992). Detection measures in real-life criminal guilty knowledge tests. *J. Appl. Psychol.* 77, 757–767.
- Ellson, D. C., Burke, C. G., Davis, R. C., and Saltzman, I. J. (1952). A report of research on detection of deception. Contract NG onr-18011, Office of Naval Research.
- Farwell, L. A., and Donchin, E. (1991). The truth will out: interrogative polygraphy (“lie detection”) with event-related potentials. *Psychophysiology* 28, 531–547.
- Furedy, J. J., Davis, C., and Gurevich, M. (1988). Differentiation of deception as a psychological process: a psychophysiological approach. *Psychophysiology* 25, 683–688.
- Furedy, J. J., and Liss, J. (1986). Countering confessions induced by the polygraph: of confessionals and psychological rubber hoses. *Crim. Law Q.* 29, 92–114.
- Gamer, M. (2010). Does the guilty actions test allow for differentiating guilty subjects from informed innocents? A re-examination. *Int. J. Psychophysiol.* 76, 19–24.
- Gamer, M. (2011a). “Detecting concealed information using autonomic measures,” in *Memory Detection: Theory and Application of the Concealed Information Test*, eds B. Verschueren, G. Ben-Shakhar, and E. Meijer (Cambridge: Cambridge University Press), 27–45.
- Gamer, M. (2011b). “Detecting of deception and concealed information using neuroimaging techniques,” in *Memory Detection: Theory and Application of the Concealed Information Test*, eds B. Verschueren, G. Ben-Shakhar, and E. Meijer (Cambridge: Cambridge University Press), 84–92.
- Honts, C. R., Devitt, M. K., Winbush, M., and Kircher, J. C. (1996). Mental and physical countermeasures reduce the accuracy of the concealed knowledge test. *Psychophysiology* 33, 211–216.
- Honts, C. R., Raskin, D. C., and Kircher, J. C. (1994). Mental and physical countermeasures reduce the accuracy of polygraph tests. *J. Appl. Psychol.* 79, 252–259.
- Horneman, C. J., and O’Gorman, J. G. (1985). Detectability in the card test as a function of the subject’s verbal response. *Psychophysiology* 22, 330–333.
- Hunter, J. E., and Schmidt, F. L. (1990). *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. Newbury Park, CA: Sage Publications, Inc.
- Iacono, W. G. (1991). “Can we determine the accuracy of polygraph tests?” in *Advances in Psychophysiology*, eds P. K. Ackles, J. R. Jennings, and M. G. H. Coles (Greenwich, CT: JAI Press), 201–201.
- Iacono, W. G., Boisvenu, G. A., and Fleming, J. A. (1984). Effects of diazepam and methylphenidate on the electrodermal detection of guilty knowledge. *J. Appl. Psychol.* 69, 289–299.
- Iacono, W. G., and Lykken, D. T. (2002). “The scientific status of research on polygraph techniques: the case against polygraph tests,” in *Modern Scientific Evidence: The Law and Science of Expert Testimony*, Vol. 2, eds D. L. Faigman, D. H. Kaye, M. J. Saks, and J. Sanders (St. Paul, MN: West Publishing, Co.), 483–538.
- Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall.
- Kensinger, E. A. (2007). Negative emotion enhances memory accuracy. *Curr. Dir. Psychol. Sci.* 16, 213–218.
- Krapohl, D. J. (2011). “Limitations of the concealed information test in criminal cases,” in *Memory Detection: Theory and Application of the Concealed Information Test*, eds B. Verschueren, G. Ben Shakhar, and E. Meijer (Cambridge: Cambridge University Press), 151–170.
- Kubis, J. F. (1962). *Studies in Lie Detection: Computer Feasibility Considerations*. Technical Report #62-205, prepared for the Air Force Systems Command. Contract No. AF 30 (602)-2270, Project No. 5534. Fordham University.
- Kugelmass, S., and Lieblich, I. (1966). Effects of realistic stress and procedural interference in experimental lie detection. *J. Appl. Psychol.* 50, 312–316.
- Kugelmass, S., Lieblich, I., and Bergman, Z. (1967). The role of “lying” in psychophysiological detection. *Psychophysiology* 3, 312–315.
- Larson, J. A. (1932). *Lying and Its Detection: A Study of Deception and Deception Tests*. Chicago, IL: University of Chicago Press.
- Lieblich, I., Kugelmass, S., and Ben Shakhar, G. (1970). Efficiency of GSR detection of information as a function of stimulus set size. *Psychophysiology* 6, 601–608.
- Lubow, R. E., and Fein, O. (1996). Pupillary size in response to a visual guilty knowledge test: new technique for the detection of deception. *J. Exp. Psychol. Appl.* 2, 164–177.
- Lui, M., and Rosenfeld, P. J. (2009). The application of subliminal priming in lie detection: scenario for identification of members of a terrorist ring. *Psychophysiology* 46, 889–903.

- Lykken, D. T. (1959). The GSR in the detection of guilt. *J. Appl. Psychol.* 43, 385–388.
- Lykken, D. T. (1960). The validity of the guilty knowledge technique: the effects of faking. *J. Appl. Psychol.* 44, 258–262.
- Lykken, D. T. (1974). Psychology and the lie detector industry. *Am. Psychol.* 29, 725–739.
- Lykken, D. T. (1988). Detection of guilty knowledge: a comment on Forman and McCauley. *J. Appl. Psychol.* 73, 303–304.
- Lykken, D. T. (1998). *A Tremor in the Blood: Uses and Abuses of the Lie Detector*. New York: Plenum Trade.
- Lynn, R. (1966). *Attention, Arousal and the Orienting Reaction*. New York: Pergamon.
- MacLaren, V. V. (2001). A quantitative review of the guilty knowledge test. *J. Appl. Psychol.* 86, 674–683.
- Maoz, K., Breska, A., and Ben-Shakhar, G. (2012). Orienting response elicitation by personally significant information under subliminal stimulus presentation: a demonstration using the concealed information test. *Psychophysiology* (in press).
- Marston, W. M. (1917). Systolic blood pressure changes in deception. *J. Exp. Psychol.* 2, 143–163.
- Marston, W. M. (1938). *The Lie Detector Test*. New York: Smith.
- Meijer, E., Smulders, F., and Merckelbach, H. (2010a). Extracting concealed information from groups. *J. Forensic Sci.* 55, 1607–1609.
- Meijer, E. H., Verschueren, B., and Merckelbach, H. (2010b). Detecting criminal intent with the concealed information test. *Open Criminol.* J. 3, 44–47.
- Meijer, E. H., Ben-Shakhar, G., Elber, L., and Klein-Selle, N. (2012). Skin conductance, respiration, heart rate and P300 in the concealed information test: a meta analysis. *A Paper Presented at the Conference of the European Association for Psychology and Law*, Nicosia.
- Meijer, E. H., Smudlers, F. T. Y., Johnston, J. E., and Merckelbach, H. L. G. J. (2007). Combining skin conductance and forced choice in the detection of concealed information. *Psychophysiology* 44, 814–822.
- Meixner, J. B., and Rosenfeld, J. P. (2010). Countermeasure mechanisms in a P300-based concealed information test. *Psychophysiology* 47, 57–65.
- Meixner, J. B., and Rosenfeld, J. P. (2011). A mock terrorism application of the P300-based concealed information test. *Psychophysiology* 48, 149–154.
- Merckelbach, H. L. G. J., Hauer, B., and Rassin, E. (2002). Symptom validity testing of feigned dissociative amnesia: a simulation study. *Psychol. Crime Law* 8, 311–318.
- Mertens, R., and Allen, J. J. (2008). The role of psychophysiology in forensic assessments: deception detection, event-related potentials, and virtual reality mock crime scenarios. *Psychophysiology* 45, 286–298.
- Nahari, G., and Ben-Shakhar, G. (2011). Psychophysiological and behavioral measures for detecting concealed information: The role of memory for crime details. *Psychophysiology* 48, 733–744.
- Nakayama, M. (2002). “Practical use of the concealed information test for criminal investigation in Japan,” in *Handbook of Polygraph Testing*, ed. M. Kleiner (London: Academic Press), 49–86.
- National Research Council. (2003). *The Polygraph and Lie Detection*. Washington, DC: The National Academies Press.
- Nisbett, R. E., and Ross, L. (1980). *Human Inference: Strategies and Shortcomings of Social Judgment*. Englewood Cliffs, NJ: Prentice-Hall.
- Öhman, A. (1992). “Orienting and attention: preferred preattentive processing of potentially phobic stimuli,” in *Attention and Information Processing in Infants and Adults*, eds B. A. Campbell, H. Hayne, and R. Richardson (Hillsdale, NJ: Erlbaum), 263–295.
- Osugi, A. (2011). “Daily application of the concealed information test: Japan,” in *Memory Detection: Theory and Application of the Concealed Information Test*, eds B. Verschueren, G. Ben Shakhar, and E. Meijer (Cambridge: Cambridge University Press), 253–275.
- Peth, J., Vossel, G., and Gamer, M. (2012). Emotionsl arousal modulates the encoding of crime-related details and corresponding physiological responses in the concealed information test. *Psychophysiology* 49, 381–390.
- Podlesny, J. A. (1993). Is the guilty knowledge polygraph technique applicable in criminal investigations? A review of FBI case records. *Crime Lab. Dig.* 20, 57–61.
- Raskin, D. C. (1989). “Polygraph techniques for the detection of deception,” in *Psychological Methods in Criminal Investigation and Evidence*, ed. D. C. Raskin (New York: Springer).
- Reid, J. E. (1947). A revised questioning technique in lie-detection tests. *J. Crim. Law Criminol.* 37, 542–547.
- Reid, J. E., and Inbau, F. E. (1977). *Truth and Deception: The Polygraph (“Lie detector”) Technique*, 2nd Edn. Baltimore: Williams & Wilkins.
- Richards, J. E., and Casey, B. J. (1992). “Development of sustained visual attention in the human infant,” in *Attention and Information Processing in Infants and Adults: Perspectives from Human and Animal Research*, eds B. A. Campbell, H. Hayne, and R. Richardson (Hillsdale, NJ: Lawrence Erlbaum Associates), 30–60.
- Rosenfeld, J. P. (2011). “P300 in detecting concealed information,” in *Memory Detection: Theory and Application of the Concealed Information Test*, eds B. Verschueren, G. Ben Shakhar, and E. Meijer (Cambridge: Cambridge University Press), 63–89.
- Rosenfeld, J. P., Ben-Shakhar, G., and Ganis, G. (2012). “Detection of concealed stored memories with psychophysiological and neuroimaging methods,” in *Memory and Law*, eds L. Nadel and W. Sinnott-Armstrong (Oxford University Press), in press.
- Rosenfeld, J. P., Cantwell, G., Nasman, V. T., Wojdac, V., Ivanov, S., and Mazzeri, L. (1988). A modified, eventrelated potential-based guilty knowledge test. *Int. J. Neurosci.* 24, 157–161.
- Rosenfeld, J. P., and Labkovsky, E. (2010). New P300-based protocol to detect concealed information: resistance to mental countermeasures against only half the irrelevant stimuli and a possible ERP indicator of countermeasures. *Psychophysiology* 47, 1002–1010.
- Rosenfeld, J. P., Labkovsky, E., Winograd, M., Lui, M. A., Vandenboom, C., and Chedid, E. (2008). The Complex Trial Protocol (CTP): a new, countermeasure-resistant, accurate P300-based method for detection of concealed information. *Psychophysiology* 45, 906–919.
- Rosenfeld, J. P., Soskins, M., Bosh, G., and Ryan, A. (2004). Simple, effective countermeasures to P300-based tests of detection of concealed information. *Psychophysiology* 41, 205–219.
- Saxe, L. (1991). Lying: thoughts of an applied social psychologist. *Am. Psychol.* 46, 409–415.
- Seymour, T. L., Seifert, C. M., Mosmann, A. M., and Shafto, M. G. (2000). Using response time measures to assess guilty knowledge. *J. Appl. Psychol.* 85, 30–37.
- Siddle, D. A. T. (1991). Orienting,habituation, and resource allocation: an associative analysis. *Psychophysiology* 28, 245–259.
- Sokolov, E. N. (1963). *Perception and the Conditioned Reflex*. New York: Macmillan.
- Thackray, R. I., and Orne, M. T. (1968). A comparison of physiological indices in detection of deception. *Psychophysiology* 4, 329–339.
- Timm, H. W. (1982). Effect of altered outcome expectancies stemming from placebo and feedback treatments on the validity of the guilty knowledge technique. *J. Appl. Psychol.* 67, 391–400.
- Timm, H. W. (1987). Effect of Biofeedback on the detection of deception. *J. Forensic Sci.* 32, 736–746.
- Verschueren, B., and Ben-Shakhar, G. (2011). “Theory of the concealed information test,” in *Memory Detection: Theory and Application of the Concealed Information Test*, eds B. Verschueren, G. Ben-Shakhar, and E. Meijer (Cambridge: Cambridge University Press), 128–148.
- Verschueren, B., Ben-Shakhar, G., and Meijer, E. (eds.). (2011). *Memory Detection: Theory and Application of the Concealed Information Test*. Cambridge: Cambridge University Press.
- Verschueren, B., and Crombez, G. (2008). Déjà vu! The effect of previewing test items on the validity of the Concealed Information polygraph test. *Psychol. Crime Law* 14, 287–297.
- Verschueren, B., Crombez, G., De Clercq, A., and Koster, E. (2005). Psychopathic traits and autonomic responding to concealed information in a prison sample. *Psychophysiology* 42, 239–245.
- Verschueren, B., Crombez, G., Declercq, A., and Koster, E. (2004). Autonomic and behavioral responding to concealed information: differentiating defensive and orienting responses. *Psychophysiology* 41, 461–466.
- Verschueren, B., Crombez, G., Degroote, T., and Rosseel, Y. (2009). Detecting concealed information with reaction times: validity and comparison with the polygraph. *Appl. Cogn. Psychol.* 23, 1–11.
- Verschueren, B., Crombez, G., Koster, E., Van Bockstaele, B., and De Clercq, A. (2007). Startling secrets: startle eye blink modification by concealed crime information. *Biol. Psychol.* 76, 52–60.
- Verschueren, B., Crombez, G., Koster, E., and De Houwer, J. (2011). “Detecting concealed information in less than a second: response latency-based measures,” in *Memory Detection: Theory and Application of the Concealed Information Test*, eds B. Verschueren, G. Ben-Shakhar, and E. Meijer

- (Cambridge: Cambridge University Press), 128–148.
- Verschueren, B., Meijer, E., and Crombez, G. (2008). Symptom validity testing for the detection of simulated amnesia: not robust to coaching. *Psychol. Crime Law* 14, 523–528.
- Vrij, A. (2008). *Detecting Lies and Deceit: Pitfalls and Opportunities*. Chichester: John Wiley and Sons.
- Wagner, A. R. (1978). "Expectancies and the priming of STM," in *Cognitive Processes in Animal Behavior*, eds S. H. Hulse, H. Fowler, and W. K. Honig (Hillsdale, NJ: Lawrence Erlbaum Associates), 177–209.
- Waid, W. M., Orne, E. C., Cook, M. R., and Orne, M. T. (1978). Effects of attention, as indexed by subsequent memory, on electrodermal detection of information. *J. Appl. Psychol.* 63, 728–733.
- Waid, W. M., Orne, E. C., Cook, M. R., and Orne, M. T. (1981a). Meprobamate reduces accuracy of physiological detection of deception. *Science* 212, 71–73.
- Waid, W. M., Orne, E. C., and Orne, M. T. (1981b). Selective memory for social information, alertness, and physiological arousal in the detection of deception. *J. Appl. Psychol.* 66, 224–232.
- Winograd, M. R., and Rosenfeld, J. P. (2011). Mock crime application of the complex trial protocol P300 based concealed information test. *Psychophysiology* 48, 155–161.

Citation: Ben-Shakhar G (2012) Current research and potential applications of the concealed information test: an overview. *Front. Psychology* 3:342. doi: 10.3389/fpsyg.2012.00342

This article was submitted to Frontiers in Cognitive Science, a specialty of Frontiers in Psychology.

Copyright © 2012 Ben-Shakhar. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.



Effects of the combination of P3-based GKT and reality monitoring on deceptive classification

Ki-Won Jang[†], Deok-Yong Kim[†], Sungkun Cho and Jang-Han Lee*

Clinical Neuro-psychology Laboratory, Department of Psychology, Chung-Ang University, Seoul, South Korea

Edited by:

Wolfgang Ambach, Institute for Frontier Areas of Psychology and Mental Health (IGPP), Germany

Reviewed by:

Elena Rusconi, University College London, UK

Giorgio Ganis, Plymouth University, UK

John B. Meixner, Northwestern University, USA

***Correspondence:**

Jang-Han Lee, Clinical Neuro-psychology Laboratory, Department of Psychology, Chung-Ang University, 221 Heukseok-dong, Dongjak-gu, Seoul 156-756, South Korea.
e-mail: clipsy@cau.ac.kr

[†]These authors equally contributed to this work.

The study aimed to investigate whether a combination of the P3-based Guilty Knowledge Test (GKT) and reality monitoring (RM) distinguished between individuals who are guilty, witnesses, or informed, and using both tests provided more accurate information than did the use of either measure alone. Participants consisted of 45 males that were randomly and evenly assigned to three groups (i.e., guilty, witness, and informed). The guilty group conducted a mock crime where they intentionally crashed their vehicle into another vehicle in a virtual environment (VE). As those in the witness group drove their own vehicles, they observed the guilty groups' vehicle crash into another vehicle. The informed group read an account and saw screenshots of the accident. All participants were instructed to insist that they were innocent. Subsequently, they performed the P3-based GKT and wrote an account of the accident for the RM analysis. A higher P3 amplitude corresponded to how well the participants recognized the presented stimulus, and a higher RM score corresponded to how well the participants reported vivid sensory information and how much less they reported uncertain information. Findings for the P3-based GKT indicated that the informed group showed lower P3 amplitude when presented with the probe stimulus than did the guilty and witness groups. Regarding the RM analysis, the informed group obtained higher RM scores on visual, temporal, and spatial details and lower scores on cognitive operations than the guilty and witness groups. Finally, discriminant analysis revealed that the combination of the P3-based GKT and RM more accurately distinguished between the three groups than the use of either measure alone. The findings suggest that RM may build upon a weakness of the P3-based GKT's. More specifically, it may build upon its susceptibility to the leakage of information about the crime, therefore helping protect innocent individuals who have information about a crime from being perceived as guilty.

Keywords: lie detection, Guilty Knowledge Test, reality monitoring, P3, leakage of knowledge

INTRODUCTION

Deception occurs in a variety of interpersonal situations. Individuals are able to detect deception using several methods. When an individual tells a lie, that person unconsciously displays potential cues that may be behavioral, verbal, and psychophysiological (Vrij, 2000; Gamer et al., 2006). Typically, lie detection tools are designed to detect a lie using these cues. One commonly used lie detection tool is the arousal-based polygraph (Vrij, 2000). The arousal-based polygraph detects a lie based on differences in psychophysiological responses (e.g., electrodermal response, cardiovascular, and respiratory) to crime-relevant and crime-irrelevant questions (Kircher and Raskin, 1988; Richardson et al., 1990; Ben-Shakhar et al., 1999). Despite its usefulness in the field, the arousal-based polygraph has some limitations. For example, the arousal-based polygraph indirectly detects a lie by measuring variables related to the lie (e.g., guilty, anxiety). However, these emotions can appear not only in a situation where individuals tell a lie but also in an uncertain situation where innocent individuals are false accused (Allen and Mertens, 2009).

Another method of lie detection is based on recognition of crime-related information that is stored in memory. This method

is called the Guilty Knowledge Test (GKT). Individuals who commit a crime have specific knowledge or memories about the crime, whereas innocent individuals do not. The GKT examines whether suspects possess this specific knowledge. If a guilty suspect recognizes the crime-related evidence presented, he or she is more likely to produce higher physiological responses than will non-guilty suspects (Vrij, 2000). One of the methods used to detect such recognition is the event-related potentials (ERPs). The ERPs provides considerably accurate information regarding temporal changes in brain activity in response to the processing of a particular stimulus. Of the ERPs components, the P300 (P3) component is evoked in response to attentive, recognized, and meaningful stimuli (Polich and Kok, 1995; Polich, 2000).

The P3-based GKT is a tool that uses temporal changes in brain activity to detect deception (Farwell and Donchin, 1991; Rosenfeld et al., 1991; Allen et al., 1992; Abootalebi et al., 2006). In the P3-based GKT, three types of stimuli, target, probe, and irrelevant, are presented to participants. Probe stimuli contain concealed crime-related information that only the guilty individuals possess (Ben-Shakhar and Elaad, 2002). Irrelevant stimuli are

stimuli with information that is unrelated to the crime. Target stimuli are also stimuli that contain information unrelated to the crime. Target stimuli are essentially a list of stimuli the participants are instructed to respond to by pressing a button when the stimuli are presented (Farwell and Donchin, 1991). The basic assumption of the P3-based GKT is that guilty individuals will recognize the probe stimuli, thus evoking higher P3 amplitude in their brain potentials compared to that evoked by the irrelevant stimuli. Conversely, innocent individuals will show no differences in P3 amplitude in brain potentials in response to the probe and irrelevant stimuli. Prior studies of ERPs-based GKT demonstrated that accurate rates of deception detection are relatively high, ranging from 89 to 90% (Farwell and Donchin, 1991; Rosenfeld et al., 1991; Allen et al., 1992). Despite its usefulness, the P3-based GKT is weak in natural environments. For example, probe stimuli may be disclosed to the public through mass media or disseminated verbally by individuals participating in the investigation (e.g., a criminal investigator). The witnesses may also have information about the crime even though they did not conduct a crime. Thus, they may not be classified into the innocent groups because of their knowledge or memories of the crime, which discourages the use of the P3-based GKT (Ben-Shakhar et al., 1999).

The quality of information about the crime may be different for guilty vs. innocent individuals even though they both have information about a particular crime. Given that they conducted the crime, guilty individuals may have more vivid sensory information about the crime as compared to innocent individuals. Such differences can be revealed by reality monitoring (RM), one type of statement analyses conducted in a crime setting. RM is based on differences in the quality of information contained in an individual's memory for experienced vs. imagined events (Johnson and Raye, 1981; Memon et al., 2010). The assumption of RM is that experienced memories differ from fictional memories (e.g., Vrij et al., 2001). The RM evaluates the quality of information contained in memories using several criteria, including visual, auditory, temporal, and spatial details as well as cognitive operations. The visual and auditory details involve perceptual and sensory information. The temporal details provide information about timing or duration of events. The spatial details include information regarding where the event took place and how objects and people were situated in relation to each other. Memories of experienced events may involve more perceptual, sensory, spatial, and temporal details than the memories of imagined events. However, imagined memories that are obtained through an internal source are likely to contain thoughts and reasoning in one's testimony, called cognitive operations. Cognitive operations are usually vague and not concrete (Vrij, 2008) and therefore are less frequent in memories of the experienced events as compared to memories of imagined events (Vrij et al., 2004). In criminal situations, an individual's statement about an experienced event is likely to contain the truth, whereas a statement pertaining to an imagined event is likely to contain false information (Vrij, 2008). Thus, the RM analysis would differentiate statements between those who actually experience a crime (guilty individuals) and those who receive information about the crime only (witness

or informed individuals). By distinguishing between those who have experienced a crime and those who only have information about a crime, RM can determine whether individuals are innocent. The individuals who witnessed the crime are more likely to describe perceptual and sensory details of the experienced event in their statements, whereas the guilty and informed individuals may describe the imagined event without actually experiencing it. For example, the guilty and informed individuals are more likely to make up a story either because they did not actually experience the crime or because had to pretend being innocent. Given these, RM may be used to complement and overcome a possible weakness of the P3-based GKT.

The aim of this study was to investigate whether a combination of the P3-based GKT and RM would more effectively differentiate among guilty individuals who conduct a crime, witnesses who experience the crime, and informed individuals who did not experience the crime but have information about the crime. Several hypotheses guide the current study. First, we predicted that in the P3-based GKT, the guilty group would show higher P3 amplitude in response to the probe stimuli than the other two groups. Second, we predicted that the witness group would meet the RM criteria more frequently than the other two groups. Third, we predicted that the combination of the P3-based GKT and RM would discriminate the groups more accurately than the P3-based GKT or RM alone.

MATERIALS AND METHODS

The participants consisted of 45 male undergraduates (15 per group), and the mean age of the sample was 23.07 years ($SD = 2.41$). They were randomly and evenly assigned to three groups: guilty, witness, and informed. This study used male participants because gender may affect ERPs amplitude (e.g., Cahill and Polich, 1992; Polich and Martin, 1992; Reinvang, 1999). We used the Machiavellianism Scale (Christie and Geis, 1970), the Social Adroitness Scale (Jackson, 1994), and Levenson's Self-Report Psychopathy Scale (LSRP) (Levenson et al., 1995) to control for the manipulativeness of the participants. No significant differences were found among groups; the Machiavellianism Scale $F_{(2, 42)} = 0.68$, *n.s.*; the Social Adroitness Scale $F_{(2, 42)} = 0.43$, *n.s.*; the LSRP $F_{(2, 42)} = 0.44$, *n.s.* (**Table 1**).

A 3D visual display was presented on dual monitors through an Olympus FMD-250W head-mounted display with a resolution of 800×600 pixels. A computer game (Grand Theft Auto

Table 1 | Demographic data and the questionnaire scores.

	Mean (SD)		
	Guilty group (n = 15)	Witness group (n = 15)	Informed group (n = 15)
Age	23.27 (1.94)	22.93 (2.84)	23.00 (2.51)
Machiavellianism	50.87 (5.18)	51.13 (6.10)	48.80 (6.68)
Social Adroitness	12.60 (2.32)	11.67 (3.39)	11.87 (2.85)
LSRP	36.47 (6.97)	37.60 (6.58)	35.40 (5.65)

LSRP Levenson's self-report psychopathy scale.

San Andreas: GTASA) that involved a third-person shooter and a driving simulator in a virtual environment (VE) was used.

When the participants arrived at the laboratory, they were asked to sign a consent form and then the experimenter explained the objectives of the study. Each participant was randomly assigned to one of the groups (i.e., guilty, witness, or informed). The guilty or witness group wore a head-mounted device while experiencing the VE. Two vehicles were driven in the VE. One vehicle was driven by the guilty or witness group, and the other one was driven by an experimenter. The guilty group intentionally crashed their vehicle into the experimenter's vehicle (i.e., head-on collision). The crash was severe enough to blow the hood off the vehicles. The witness group was instructed to drive safely and watch the crash that was caused by the experimenter. The informed group was instructed to watch screenshots of the car accident caused by the guilty group and read the description of the accident. The duration of the VE was approximately 15 min. All of the groups were instructed to write statements about the car accident and to insist that they were innocent.

Subsequently, all participants were given the GKT to evaluate whether they recognized the vehicle of assailant. Three types of stimuli (irrelevant, probe, and target stimuli) were used in the GKT. The irrelevant stimuli consisted of four screenshots of a vehicle not related to the accident, and the probe stimulus was a screenshot of the assailant's vehicle. All participants were instructed to respond by pressing a spacebar when the target stimulus (i.e., a vehicle unrelated to the car accident) was presented. The vehicles used in this experiment were similar in size and color but different in shape. Each trial consisted of one probe, four irrelevant, and one target stimuli (total 40 trials). After presenting a fixation for 500 ms, the probe, irrelevant, and target stimuli were randomly presented for 1000 ms, and then a blank screen followed for 1500 ms with an inter-stimulus-interval (ISI) of 3000 ms (visual angle being 16° in width and 13° in height; 518 × 370 pixels). After attaching the electrodes to the head of each participant, the participants were instructed to press a "yes" button when the target stimuli were presented and a "no" button when the others were presented. The total duration of the P3-based GKT was approximately 12 min.

EEG data were recorded from 28 sites (Fpz, Fz, FCz, Cz, CPz, Pz, Oz, Fp2, F3/4, F7/8, FC3/4, C3/4, CP3/4, P3/4, P7/8, O1/2, T7/8, and FT7/8) with reference electrodes on the earlobes using the Neuroscan System (Neuroscan Labs, Sterling, VA, USA). EOGs were recorded from the outer canthus of each eye to measure horizontal electro-oculograms (HEOG) activity and from the left eye to measure vertical electro-oculograms (VEOG) activity. All impedances were maintained at 5 kΩ or less. The data were digitized at a rate of 512 Hz for 800 ms and recorded with a band-pass of 0.01–100 Hz. Epochs were created from -100 to 898 ms around the stimulus and baseline corrected using the 100 ms pre-stimulus period. Artifacts in which the EEG or EOG exceeded ±100 µV were rejected (Semlitsch et al., 1986). The bandpass filter was applied 0.05–10 Hz (24 dB octave/slope). The P3 component was typically defined as the largest positive peak occurring between 300 and 1000 ms at each electrode (Abootalebi et al., 2006). Amplitude was measured as the difference between the mean pre-stimulus baseline and maximum peak amplitude at Pz,

because the Pz site is usually reported to be maximal among the other sites. Peak detection was done automatically but verified manually.

The RM analysis consisted of five domains: visual, auditory, temporal, and spatial details as well as cognitive operations. First, a statement met the criteria of having visual details if it contained a vivid or clear description. Second, a statement met the criteria for having auditory details if it encompassed auditory information. Third, a statement met the criteria for temporal details if it included the order in which the accident occurred. Fourth, a statement met the criteria for spatial details if the statement encompassed locational information on humans or objects. Fifth, a statement met the criteria for cognitive operations if it contained descriptions of imagined events from internal source, such as thoughts and reasoning. Cognitive operations were scored dichotomously: a score of 0 when cognitive operations were not present and a score of 1 when cognitive operations were presented. For the RM, a score for each domain was calculated by summing of frequency of meeting the criteria in their statements by two independent raters (Vrij et al., 2004).

RESULTS

P3-BASED GKT

A 3 (group: guilty, witness, and informed) × 3 (stimulus: probe, irrelevant, and target) repeated-measures ANOVA was performed to analyze the P3 data. The results indicated that there was a significant interaction effect between group and stimulus, $F_{(4, 84)} = 3.27, p < 0.05, \eta^2 = 0.14$. Subsequently, we performed One-Way ANOVA for each stimulus to investigate differences among the groups. There was a significant main effect of group membership on the P3 amplitudes for the probe stimuli, $F_{(2, 42)} = 8.42, p = 0.001, \eta^2 = 0.29$, but there were no significant group differences for the target and irrelevant stimuli, $F_{(2, 42)} = 0.31; F_{(2, 42)} = 0.37, n.s.$. In the pairwise comparison test, the informed group showed lower P3 amplitudes in response to the probe stimuli than the guilty, $t_{(28)} = 4.13, p < 0.001$, and witness groups, $t_{(28)} = 2.68, p = 0.01$. The difference between the guilty and the witness groups, however, did not reach statistical significance, $t_{(28)} = -1.33, n.s.$ (**Figure 1**). Additionally, there was a significant main effect of stimulus, $F_{(2, 84)} = 49.48, p < 0.001, \eta^2 = 0.54$. A pairwise comparison test indicated that P3 amplitudes in response to the target stimuli were higher than those in response to the probe, $t_{(44)} = 4.43, p < 0.001$, and irrelevant stimuli, $t_{(44)} = 8.99, p < 0.001$, and that the P3 amplitudes in response to the probe stimuli were significantly higher than the irrelevant stimuli, $t_{(44)} = 5.89, p < 0.001$. However, there was no main effect of group (**Figure 2**).

REALITY MONITORING

For the RM analysis, a 2 (group: guilty, witness, and informed) × 5 (criteria: visual, auditory, temporal, spatial, and cognitive operations) MANOVA was used. A MANOVA revealed a significant multivariate main effect for group, Wilks' = 0.13, $F_{(10, 76)} = 13.61, \eta^2 = 0.64$. At a univariate level, there were significant main effects for visual, temporal, and spatial details as well as cognitive operations: visual details, $F_{(2, 42)} = 57.10, p < 0.001, \eta^2 = 0.73$; temporal details, $F_{(2, 42)} = 6.43, p < 0.01$,

$\eta^2 = 0.23$; spatial details, $F_{(2, 42)} = 12.57, p < 0.001, \eta^2 = 0.37$; and cognitive operations, $F_{(2, 42)} = 34.20, p < 0.001, \eta^2 = 0.62$. A pairwise comparison tests indicated that the witness group reported significantly more visual, temporal, and spatial details than did the guilty [visual details: $t_{(28)} = 8.70, p < 0.001$; temporal details: $t_{(28)} = 2.84, p < 0.01$; spatial details: $t_{(28)} = 3.72, p = 0.001$] and informed groups [visual details: $t_{(28)} = 8.87, p < 0.001$; temporal details: $t_{(28)} = 3.04, p < 0.01$; spatial details: $t_{(28)} = 3.80, p = 0.001$]. In terms of the auditory details, however, there were no significant differences among the groups.

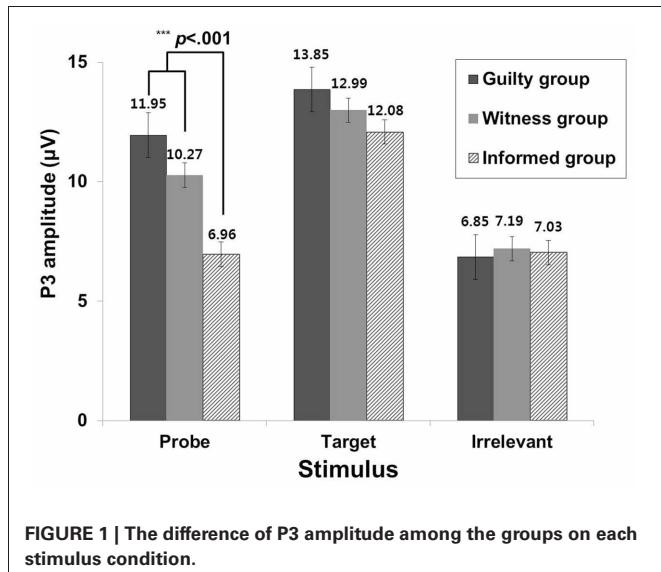


FIGURE 1 | The difference of P3 amplitude among the groups on each stimulus condition.

For cognitive operations, the witness group had significantly less than the guilty, $t_{(28)} = -8.30, p < 0.001$, and informed group, $t_{(28)} = -3.55, p = 0.001$. Furthermore, the informed group also reported cognitive operations significantly less than the guilty group, $t_{(28)} = -4.63, p < 0.001$ (Figure 3).

DISCRIMINANT ANALYSES FOR P3-BASED GKT AND REALITY MONITORING

A discriminant analysis was conducted to investigate whether a combination of the P3-based GKT and RM was better at distinguishing between the guilty, witness, and informed groups than

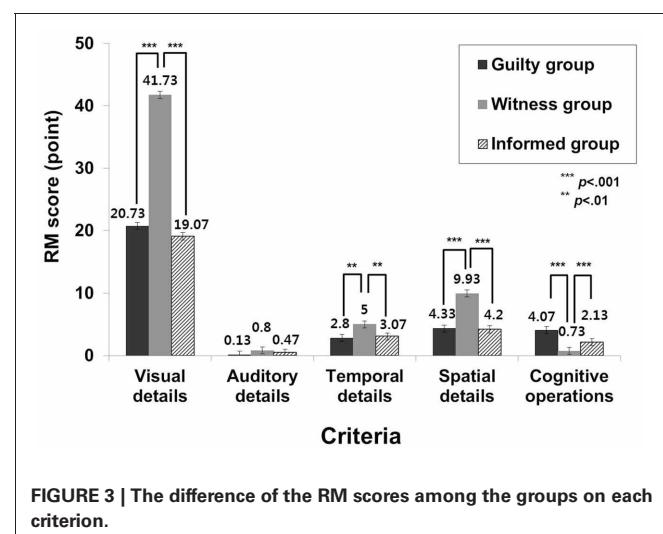


FIGURE 3 | The difference of the RM scores among the groups on each criterion.

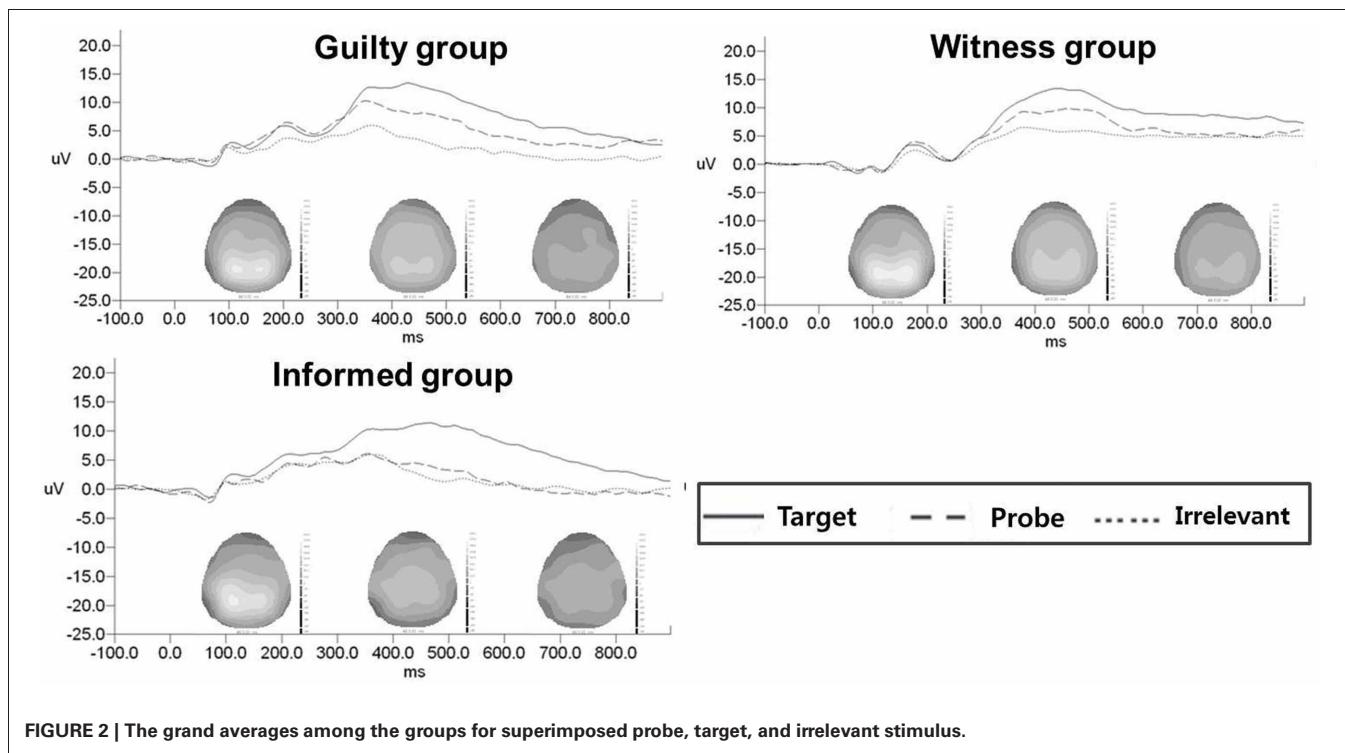


FIGURE 2 | The grand averages among the groups for superimposed probe, target, and irrelevant stimulus.

the P3-based GKT (i.e., probe, irrelevant, and target stimuli) or RM (i.e., visual, auditory, temporal, and spatial details as well as cognitive operations). To develop an optimum classifier to discriminate among the groups, two discriminant analyses were conducted. One analysis was performed to find a discriminant function to maximize the separation between the guilty and witness groups, and the other analysis was performed to distinguish between the guilty and informed groups.

First, we conducted a discriminant analysis to determine whether the P3-based GKT, RM, and a combination of the P3-based GKT and RM distinguished between the guilty and witness groups. For the results of the P3-based GKT, univariate *F* tests showed that the discriminant function was not significant, $\chi^2_{(3, N=30)} = 5.20, p = 0.16$, and indicated that 82.2% of total variance was not explained, $\lambda = 0.82$. In the RM, univariate *F* tests indicated significant differences for the visual, auditory, temporal, and spatial details and cognitive operations, $\chi^2_{(5, N=30)} = 49.93, p < 0.001$. The mean classification accuracy was 100.0%. **Table 2** shows that 100.0% of the guilty group and 100.0% of witness group were classified correctly in the present study. An internal validation of the discriminant analysis (jack-knife) also indicated 100.0% correct classifications. The result for the combination of P3-based GKT and RM showed significant differences for the predicted variables, $\chi^2_{(8, N=30)} = 48.60, p < 0.001$. The mean classification accuracy was 100.0%. In these analyses, 100.0% of the guilty group and 100.0% of the witness group were classified correctly both in the present study and in the internal validation of the discriminant analysis (**Table 2**).

A discriminant analysis was performed to distinguish between the guilty and informed groups in the P3-based GKT, RM, and combination of the two. The discriminant function analysis results for the P3-based GKT were significant, $\chi^2_{(3, N=30)} = 25.53, p < 0.001$, and showed an 86.7% overall

correct classification accuracy. A total of 86.7% of the guilty group and 86.7% of the informed group were classified correctly. An internal validation of the discriminant analysis also indicated 86.7% correct classifications. For the results of the RM, the discriminant function indicated significant differences, $\chi^2_{(5, N=30)} = 23.57, p < 0.001$. The mean classification accuracy was 86.7%. For these analyses, 86.7% of the guilty group and 86.7% of the informed group were classified correctly. An internal validation of the discriminant analysis showed 76.7% correct classifications, and 66.7% of the guilty group and 86.7% of the informed group were correctly classified. The result of a combination between the P3-based GKT and RM showed significant differences in the predicted variables, $\chi^2_{(8, N=30)} = 34.22, p < 0.001$. The overall classification accuracy was 93.3%. In this result, 86.7% of the guilty group and 100.0% of the witness group were classified correctly both in the present study. An internal validation of the discriminant analysis was 90.0% (**Table 3**).

DISCUSSION

The purpose of this study was to investigate differences in the P3-based GKT or RM among individuals in the guilty, witness, and informed groups. Additionally, we investigated whether the combination of the P3-based GKT and RM would more accurately discriminate among the groups than either test alone.

The results indicated that the informed group showed lower P3 amplitude in response to the probe stimulus than did the guilty and witness groups. These results partly support the first hypothesis. Indeed, the results suggest that the P3-based GKT may differentiate individuals who do not experience the crime but who have information about the crime (i.e., informed individuals), from those who do experience it (i.e., witnesses, and guilty individuals). The informed individuals may have less specific memories surrounding their crime compared to individuals

Table 2 | Discriminant analyses with P3-based GKT and reality monitoring between the guilty and witness group.

Cross validated classification rates (original classification)					
	Hit rates			Eigenvalue	λ
	Guilty group	Witness group	Total		
P3-GKT	60.0 (73.3)	40.0 (26.7)	73.3 (80.0)	0.22	0.82
RM	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)	6.09	0.14
P3-GKT + RM	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)	6.58	0.13

Table 3 | Discriminant analyses with P3-based GKT and reality monitoring between the guilty and informed group.

Cross validated classification rates (original classification)					
	Hit rates			Eigenvalue	λ
	Guilty group	Informed group	Total		
P3-GKT	86.7 (86.7)	86.7 (86.7)	86.7 (86.7)	1.62	0.38
RM	66.7 (86.7)	86.7 (86.7)	76.7 (86.7)	1.52	0.40
P3-GKT + RM	86.7 (86.7)	93.3 (100.0)	90.0 (93.3)	3.16	0.24

who experience the crime because they did not directly experience the crime. It may be difficult to identify the source of their crime-related knowledge (Zvi et al., 2012). More specifically, the guilty and witness groups recognized the assailants' better than the informed group, presumably because these two groups had direct experiences with the accident. The P3-based GKT, however, did not reveal significant differences between the guilty and witness groups regarding responses to the probe stimuli. Given these findings, the P3-based GKT appears to be weak in terms of its ability to discriminate between groups when knowledge of a crime is leaked (Ben-Shakhar and Elaad, 2003). The witness and guilty groups may have had similar amounts of vivid information about the crime stored in memory. If this is the case, the P3-based GKT may not have been able to differentiate between the guilty and witness groups. Thus, the P3-based GKT may need another tool to overcome this weakness. More specifically, another tool is needed to differentiate between guilty individuals and witnesses.

Regarding the RM, there were significant differences among the groups, in the visual, temporal, and spatial details as well as cognitive operations. The statements from the witness group included more visual details, temporal information, spatial information, and less cognitive operations than those from the guilty and informed groups. These results may be due to differences in the quality of the crime-related information stored in memory (Vrij et al., 2004). Most likely, the witness group recalled the sensory, perceptual, and contextual memories that they experienced in the VE. The guilty and informed groups, however, described a story that they did not experience. More specifically, the guilty group had to provide a false statement. The informed group had to provide factual statements about the accident. Thus, they were less likely to have limited sensory, perceptual, and contextual memories of the accident. These results suggest that RM may differentiate the experienced-driven true memories and false or imagined memories. Given these findings, the RM seems to be weak in terms of its ability to accurately distinguish among the three groups.

The discriminant analyses revealed that the combination of the P3-based GKT and RM showed higher accuracy rates compared to two methods independently. In terms of the results of the P3-based GKT, the discriminant function was unable to distinguish between the guilty individuals and witnesses, whereas it was able to distinguish between the guilty and informed individuals. These results indicate that the P3-based GKT may not differentiate between witnesses and guilty individuals. Thus, the results suggest the P3-based GKT is weak when knowledge of a crime is leaked. In other words, the witnesses may be falsely accused of committing a crime when the P3-based GKT is employed for the purpose of lie detecting. The results pertaining to RM analysis revealed a high discrimination rate for the witness group (100.0% of the witness group). The combination of the P3-based GKT and RM correctly classified 100.0% of the witness group. This result implies that RM may overcome a weakness of the P3-based GKT. Additionally, the discriminant analysis of the P3-based GKT revealed that it was highly able to discriminate between the guilty and informed individuals, whereas the discriminant analysis of the RM showed moderate discrimination (an internal validation of discrimination analysis: 76.7%). These results

highlight a limitation of RM because both the guilty and informed individuals should possess an imagined memory of the crime that is reflected in their statement. Regarding the overall classification rate, the combination of the P3-based GKT and RM also showed a higher rate of classifications than either the P3-based GKT or RM alone, although differences in correct-classification rates among the techniques were not examined. The results of the present study are comparable to a previous study in which the combination of the Criteria-Based Content Analysis and RM correctly classified 80.8% of the participants (Vrij et al., 2000). However, the aim of the previous study was to discriminate between the guilty group and innocent group that had no information about the crime. In the present study, the combined method showed a higher classification rate (an internal validation between the guilty and witness group of 100.0% and an internal validation between the guilty and informed group of 90.0%). By combining the P3-based GKT and RM, each method builds upon the weaknesses of the other method. In conclusion, the present study suggests that the combination of the P3-based GKT and RM may differentiate among individuals who are guilty, witnesses, and informed.

In the present study, the RM was used to build upon a weakness of the P3-based GKT. Although the combination of the P3-based GKT and RM was adequately able to differentiate between the guilty, witness, and informed groups, this combination also has some possible weakness. If the guilty individuals know the RM criteria, they may be able to manipulate the quality of their report by intentionally changing the balance among cognitive operations, visual, auditory, temporal, and spatial details. Therefore, future studies need to identify the optimal combination of the P3-based GKT and other various methods to differentiate between the guilty individuals and the witnesses or informed individuals.

The present study has several implications. First, the present study suggests that a combination of the P3-based GKT and RM may build upon the weakness of the P3-based GKT because the test is susceptible to the leakage of information about the crime. Therefore, the method may help protect innocent individuals from perceived as guilty when they have information about the crime that was disclosed to the public through mass media or by participating in the investigation.

Second, the present study showed that the GKT using the image stimulus that participants experienced in a VE can discriminate between the groups. Previous study of ERP-based deception detection using a mock crime in a VE indicated that the hit rates were quite low (Mertens and Allen, 2008). Possible reason for such low hit rates would be due to the feature of stimulus. For example, they used stimulus consisting of words, but not images. Many studies have suggested the picture superiority effect (Buckner et al., 2000) that pictures are better recalled than words and better recollected when cued with a fragment only (McBride and Dosher, 2002; Cutmore et al., 2009). Given these, it appears to be reasonable to use image stimuli for a GKT in a VE than word stimuli. Although the findings on the detection of deception using the ERPs-based GKT in a VE have been acceptable in a laboratory experiment, we suggest that future research apply the ERPs-based GKT in a real forensic situation. For example, in a

real forensic situation, there could be delay between conducting a crime and assessing deception. However, in the laboratory setting, the deception is assessed right after conducting a mock crime. Therefore, we suggest that future studies compare memories about a crime both immediately after the crime and after a delayed period.

REFERENCES

- Abootalebi, V., Moradi, M. H., and Khalilzadeh, M. A. (2006). A comparison of methods for ERP assessment in a P300-based GKT. *Int. J. Psychophysiol.* 62, 309–320.
- Allen, J., Iacono, W. G., and Danielson, K. D. (1992). The identification of concealed memories using the event-related potential and implicit behavioral measures: a methodology for prediction in the face of individual differences. *Psychophysiology* 29, 504–522.
- Allen, J., and Mertens, R. (2009). Limitations to the detection of deception: true and false recollections are poorly distinguished using an event-related potential procedure. *Soc. Neurosci.* 4, 473–490.
- Ben-Shakhar, G., and Elaad, E. (2002). “The Guilty Knowledge Test (GKT) as an application of psychophysiology: future prospects and obstacles,” in *Handbook of Polygraph Testing*, ed M. Kleiner (New York, NY: Academic Press), 87–102.
- Ben-Shakhar, G., and Elaad, E. (2003). The validity of psychophysiological detection of information with the Guilty Knowledge Test: a meta-analytic review. *J. Appl. Psychol.* 88, 131–151.
- Ben-Shakhar, G., Gronau, N., and Elaad, E. (1999). Leakage of relevant information to innocent examinees in the GKT: an attempt to reduce false-positive outcomes by introducing target stimuli. *J. Appl. Psychol.* 84, 651–660.
- Buckner, R. L., Logan, J., Donaldson, D. I., and Wheeler, M. E. (2000). Cognitive neuroscience of episodic memory encoding. *Acta Psychol.* 105, 127–139.
- Cahill, J. M., and Polich, J. (1992). P300, probability, and introverted/extroverted personality types. *Biol. Psychol.* 33, 23–35.
- Christie, R., and Geis, F. (1970). *Studies in Machiavellianism*. New York, NY: Academic Press.
- Cutmore, T. R. H., Djakovic, T., Kebbell, M. R., and Shum, D. H. K. (2009). An object cue is more effective than a word in ERP-based detection of deception. *Int. J. Psychophysiol.* 71, 185–192.
- Farwell, L. A., and Donchin, E. (1991). The truth will out: interrogative polygraphy (lie detection) with event-related potentials. *Psychophysiology* 28, 531–547.
- Gamer, M., Rill, H., Vossel, G., and Gödert, H. W. (2006). Psychophysiological and vocal measures in the detection of guilty knowledge. *Int. J. Psychophysiol.* 60, 76–87.
- Jackson, D. N. (1994). *Jackson Personality Inventory-revised Manual*. Port Huron, MI: Sigma Assessment Systems.
- Johnson, M. K., and Raye, C. L. (1981). Reality monitoring. *Psychol. Rev.* 88, 67–85.
- Kircher, J. C., and Raskin, D. C. (1988). Human versus computerized evaluations of polygraph data in a laboratory setting. *J. Appl. Psychol.* 73, 291–302.
- Levenson, M. R., Kiehl, K. A., and Fitzpatrick, C. M. (1995). Assessing psychopathic attributes in a noninstitutional population. *J. Pers. Soc. Psychol.* 68, 151–158.
- McBride, D. M., and Dosher, B. A. (2002). A comparison of conscious and automatic memory processes for picture and word stimuli: a process dissociation analysis. *Conscious. Cogn.* 11, 423–460.
- Memon, A., Fraser, J., Colwell, K., Odinot, G., and Mastroberardino, S. (2010). Distinguishing truthful from invented accounts using reality monitoring criteria. *Leg. Crim. Psychol.* 15, 177–194.
- Mertens, R., and Allen, J. J. B. (2008). The role of psychophysiology in forensic assessments: deception detection, ERPs, and virtual reality mock crime scenarios. *Psychophysiology* 45, 286–298.
- Polich, J. (2000). “P300 in clinical applications,” in *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, 4th Edn. eds E. Niedermeyer, and F. Lopes Da Silva (Baltimore, Maryland: Lippincott Williams and Wilkins), 1073–1091.
- Polich, J., and Kok, A. (1995). Cognitive and biological determinants of P300: an integrative review. *Biol. Psychol.* 41, 103–146.
- Polich, J., and Martin, S. (1992). P300, cognitive capability and personality: a correlational study of university undergraduates. *Pers. Individ. Dif.* 13, 533–543.
- Reinvang, I. (1999). Cognitive event-related potentials in neuropsychological assessment. *Neuropsychol. Rev.* 9, 231–248.
- Richardson, D. C., Carlton, B. L., and Dutton, D. W. (1990). Blind analysis of skin conductance response (SCR) recordings from a number test. *Polygraph* 19, 9–20.
- Rosenfeld, J. P., Angell, A., Johnson, M., and Qian, J. H. (1991). An ERP-based, control-question lie detector analog: algorithms for discriminating effects within individuals’ average waveforms. *Psychophysiology* 28, 319–335.
- Semlitsch, H. V., Anderer, P., Schuster, P., and Presslich, O. (1986). A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP. *Psychophysiology* 23, 695–703.
- Vrij, A. (2000). *Detecting Lies and Deceit: The Psychology of Lying and the Implications for Professional Practice*. Chichester: John Wiley and Sons.
- Vrij, A. (2008). Nonverbal dominance versus verbal accuracy in lie detection: a plea to change police practice. *Crim. Justice Behav.* 35, 1323–1336.
- Vrij, A., Akehurst, L., Soukara, S., and Bull, R. (2004). Let me inform you how to tell a convincing story: CBCA and reality monitoring scores as a function of age, coaching and deception. *Can. J. Behav. Sci.* 36, 113–126.
- Vrij, A., Edward, K., and Bull, R. (2001). Stereotypical verbal and nonverbal responses while deceiving others. *Pers. Soc. Psychol. Bull.* 27, 899–909.
- Vrij, A., Edward, K., Roberts, K. P., and Bull, R. (2000). Detection deceit via analysis of verbal and nonverbal behavior. *J. Nonverbal Behav.* 24, 239–263.
- Zvi, L., Nachson, I., and Elaad, E. (2012). Effects of coping and cooperative instructions on guilty and informed innocents’ physiological responses to concealed information. *Int. J. Psychophysiol.* 84, 140–148.

ACKNOWLEDGMENTS

National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2012-M3A2A1051124). This research was presented in poster form at the 15th World Congress of Psychophysiology of International Organization of Psychophysiology.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 April 2012; **accepted:** 15 January 2013; **published online:** 31 January 2013.

Citation: Jang K-W, Kim D-Y, Cho S and Lee J-H (2013) Effects of the combination of P3-based GKT and reality monitoring on deceptive classification. *Front. Hum. Neurosci.* 7:18. doi: 10.3389/fnhum.2013.00018

Copyright © 2013 Jang, Kim, Cho and Lee. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.



Concealed semantic and episodic autobiographical memory electrified

Giorgio Ganis^{1,2,3*} and Haline E. Schendan^{1,2}

¹ School of Psychology, Cognition Institute, University of Plymouth, Plymouth, UK

² Massachusetts General Hospital, Athinoula A. Martinos Center for Biomedical Imaging, Charlestown, MA, USA

³ Department of Radiology, Harvard Medical School, Boston, MA, USA

Edited by:

Matthias Gamer, University Medical Center Hamburg-Eppendorf, Germany

Reviewed by:

Izumi Matsuda, National Research Institute of Police Science, Japan
John B. Meixner, Northwestern University, USA

***Correspondence:**

Giorgio Ganis, School of Psychology, Cognition Institute, University of Plymouth, Plymouth, UK.
e-mail: ganis@nmr.mgh.harvard.edu

Electrophysiology-based concealed information tests (CIT) try to determine whether somebody possesses concealed information about a crime-related item (probe) by comparing event-related potentials (ERPs) between this item and comparison items (irrelevants). Although the broader field is sometimes referred to as "memory detection," little attention has been paid to the precise type of underlying memory involved. This study begins addressing this issue by examining the key distinction between semantic and episodic memory in the autobiographical domain within a CIT paradigm. This study also addresses the issue of whether multiple repetitions of the items over the course of the session habituate the brain responses. Participants were tested in a 3-stimulus CIT with semantic autobiographical probes (their own date of birth) and episodic autobiographical probes (a secret date learned just before the study). Results dissociated these two memory conditions on several ERP components. Semantic probes elicited a smaller frontal N2 than episodic probes, consistent with the idea that the frontal N2 decreases with greater pre-existing knowledge about the item. Likewise, semantic probes elicited a smaller central N400 than episodic probes. Semantic probes also elicited a larger P3b than episodic probes because of their richer meaning. In contrast, episodic probes elicited a larger late positive complex (LPC) than semantic probes, because of the recent episodic memory associated with them. All these ERPs showed a difference between probes and irrelevants in both memory conditions, except for the N400, which showed a difference only in the semantic condition. Finally, although repetition affected the ERPs, it did not reduce the difference between probes and irrelevants. These findings show that the type of memory associated with a probe has both theoretical and practical importance for CIT research.

Keywords: concealed information, deception, deception detection, ERPs (event-related potentials), semantic memory, episodic memory

INTRODUCTION

The logic of concealed information tests (CIT) is that stimuli that are known or familiar to people should elicit a different response relative to comparable stimuli that are new (Lykken, 1959). Such tests could have various forensic applications, for example, to determine whether a person who denies having information about certain crime details or certain sensitive information actually possesses such information. CITs have been studied for many decades using several dependent variables, including long-standing, peripheral psychophysiological measures (Ben-Shakhar and Elaad, 2003) and, more recently, electrophysiological (event-related potential, ERP) (Rosenfeld et al., 1988, 1991, 2008; Farwell and Donchin, 1991; Allen et al., 1992) and hemodynamic ones (functional magnetic resonance imaging, fMRI) (Langleben et al., 2002; Phan et al., 2005; Christ et al., 2009; Nose et al., 2009; Ganis et al., 2011).

ERP-based CITs have garnered increased attention lately due to several advantages (e.g., Rosenfeld et al., 2008). (1) They have shown high accuracy rates reliably in detecting concealed

information in mock crime scenario paradigms, at least in the laboratory conditions tested. (2) They are relatively inexpensive to implement. (3) The data can be acquired relatively quickly by using a few recording sites on the head. However, the underlying neural mechanisms are largely undetermined. A critical, under-studied issue in the field is that people can learn and remember information about an event in many ways.

For example, memory theories distinguish between semantic and episodic memory (Tulving, 1972), and different brain systems have been implicated in each. Episodic memory depends on mediotemporal lobe structures, especially the hippocampus, whereas semantic memory does so much less, if at all, and depends on association cortex, such as anterior temporal cortex (Vargha-Khadem et al., 1997; Schmolck et al., 2002; Eichenbaum et al., 2007; Patterson et al., 2007; Bayley et al., 2008). That different brain systems support episodic and semantic memory raises the important issue that the brain signatures should differ when concealed information revealed on a CIT relies to different degrees on episodic vs. semantic memory. For example, evidence

from developmental amnesia patients, who have hippocampal damage and impaired episodic but spared semantic memory, suggests that even residual hippocampal function (despite 50% volume loss or more) is necessary and sufficient to support relative sparing of the ability to imagine false events (Maguire et al., 2010), which is a necessary episodic memory ability for effective deception; neural signatures of such hippocampal activity would thus be expected to be greater for a CIT based on episodic memory relative to one based on semantic memory.

Indeed, the episodic-semantic distinction extends also to the kind of autobiographical memory typically tested in CITs (e.g., Martinelli et al., 2012), the focus of this paper. There are episodic and semantic forms of autobiographical memory. An episodic memory encompasses concrete and unique details associated with distinct events that were experienced by a person in a specific spatiotemporal context and, critically, becomes an episodic autobiographical memory (EAM) when this memory also refers to the self in relation to that context (Tulving, 2002). For example, details about a specific experience that happened at a certain time and place that caught one by surprise. In contrast, semantic autobiographical memory (SAM) encompasses personal information, including general knowledge of personal facts not associated with a specific time and place of acquisition (e.g., “my name is Pat” or “my birthday is December 5th”) and non-specific events, including both repeated and extended events (e.g., schema and script knowledge about “birthdays” not associated with any specific time and place, such as that birthdays are fun and involve friends and family) (Schank and Abelson, 1977). Studies in neurological patients confirm this distinction. For example, amnesic patient K.C. (Tulving, 1993) could report semantic knowledge, such as his own date of birth, but not any autobiographical episodic information (e.g., autobiographical details about any specific birthday). An important question is whether autobiographical probes associated with high semantic vs. episodic memory are associated with different neural processes in the context of a CIT, as would be predicted by neurocognitive studies of these two types of memories (e.g., Tulving et al., 1988; Martinelli et al., 2012). This question also has applied relevance because it could provide information about the brain signatures of these different types of memories that can inform how to maximize detecting concealed information in specific cases. It is important to note that, although there may be distinct neural systems supporting EAM and SAM (e.g., Martinelli et al., 2012), most information in real life is often associated with both EAM and SAM, though with different relative strengths. Note that, for simplicity, in the rest of the paper we will often omit the attribute “autobiographical” and refer simply to semantic and episodic memory.

The main previous ERP study that addressed a related question with an explicitly applied focus is one by Rosenfeld et al. (2006). “High-impact” and “low-impact” probes were compared that differed in semantic and episodic memory content. The high-impact probe was the participant’s name, whereas the low-impact probe was the experimenter’s name (i.e., “JULIE”). The ERP differences between high-impact probes and a set of random control names (referred to as “irrelevants” in the CIT literature) were much larger than those between the low-impact probes and the irrelevants (i.e., the CIT effect was larger for high than low-impact

probes). However, important issues about this finding need to be resolved. First, the same low-impact probe was used for all participants (i.e., the experimenter’s name was always “JULIE”). This raises the concern that there could be something intrinsically special, and consistently so across participants, about this name (e.g., frequency, length, associations). This confound was not present for the high-impact probes, as they varied across participants. Furthermore, it is unclear whether the female name used for everybody in the low-impact condition might have been processed differently by male and female participants (i.e., Julie is a female name), as well as individuals (i.e., different people named Julie that each one knows), increasing variability in the results. The ERPs were also recorded from only three sites, limiting assessment of spatial distribution differences between conditions. Finally, the study examined only the P3, leaving it open what effects other ERPs might show, such as the centroparietal N400 marker of semantic memory (Kutas and Federmeier, 2011) or the parietal late positive complex (LPC) associated with episodic recollection (Rugg and Curran, 2007). We would argue that the better way to describe the high- and low-impact probes is in terms of how they activate different kinds of memory. For example, both probes activate semantic and episodic memory, but in different ways for the participant’s name (“high-impact”) and the experimenter’s name (“low-impact”). Specifically, the participant’s name could activate semantic memory more automatically than episodic memory, on average, because people are overlearned experts at responding to their own name, whereas most episodic memories associated with their name would be remote and many would be highly similar and so not distinctly memorable, such as people calling their name, potentially resulting in a lot of interference for recalling associated episodic memories and making them effortful to activate (Soderlund et al., 2012). Thus, semantic memory would be exceptionally automatic for the participant’s name, consistent with evidence for a large auditory N400 for one’s own name relative to other proper names and no evidence for a posterior LPC effect, suggesting little difference in episodic memory for one’s own name and other proper names (Muller and Kutas, 1996). However, by telling subjects that the experimenter’s name is “Julie,” subjects acquire a recent episodic memory, which is less effortful to activate than the more remote memories associated with one’s own name (Soderlund et al., 2012), predicting a larger LPC for the experimenter’s than participant’s name, but this has not yet been examined to date. In summary, we would suggest that in the study by Rosenfeld et al. (2006), the participant’s name would predominantly activate SAM, whereas the experimenter’s name would predominantly activate recent EAM, but such ideas have not yet been systematically addressed.

Thus, the first goal of the current study was to address the question of concealed information based on different types of memory more directly while getting around the limitations in the previous work. First, comparable stimuli without a gender component were used for the semantic (the participant’s date of birth) and episodic (a “secret” date given to the participant just before the study) autobiographical memory conditions. Second, all probes and irrelevants varied by person, eliminating any systematic biases in the group average. Third, 32 recording sites

were used, enabling potential scalp distribution differences in the ERPs elicited by the two conditions to be determined. Fourth, and related to the previous point, not only the P3 but also other ERPs were evaluated, including the frontal N2, the N400, and the LPC.

A second important issue that has not been addressed systematically in the ERP literature is the effect of stimulus repetition. Because of the relatively low signal-to-noise ratio achievable with all behavioral and psychophysiological measures employed, the typical CIT paradigm averages several tens of trials in which probes and irrelevants repeat many times. Differences between probes and irrelevants using psychophysiological measures, such as skin conductance, decrease rapidly with stimulus repetition because of habituation (e.g., Ben-Shakhar et al., 1975; Ben-Shakhar and Elaad, 2002). However, the same effect may not be present with ERP measures because they may tap into different mechanisms. Furthermore, potential differences between semantic and episodic probes may change over the course of the experimental session. For example, repeated presentation will reactivate semantic and/or episodic memories associated with a probe but do less so if at all for irrelevants, since no distinct semantic or episodic information is available about them. This could result in a difference between probes and irrelevants that becomes larger over time, as ERP repetition effects can be greater for meaningful than meaningless items (Schendan and Maher, 2009; Voss et al., 2010). Another possibility is that repetition of the probes might alter the activation of the semantic and/or episodic memory underlying each. For example, the episodic probe might develop increasing associations with the experimental context, resulting in development of semantic memory (Gratton et al., 2009). This might reduce the N400 (which is smaller when semantic memory activates more successfully) (Voss et al., 2010; Voss and Federmeier, 2011), thereby reducing differences between semantic and episodic probes. On the other hand, all stimuli, including the semantic probe, might develop additional episodic memories with each exposure in the experiment, resulting in additional episodic memories that might increase the LPC (which is larger for more episodic memory), thereby also reducing differences between semantic and episodic probes and associated CIT effects.

The key idea in classical CIT theories is that probes will generate an orienting response associated with, for example, increased skin conductance (e.g., Sokolov, 1963; Gati and Ben-Shakhar, 1990). Although these theories may be adequate to explain autonomic nervous system findings, they cover only a subset of the central nervous system processes engaged by a probe during the CIT (relative to irrelevants) and implicitly assume that probes activate only one kind of memory. However, in the framework described here, semantic and episodic probes may be associated with different neural processes.

Current theories of memory predict that semantic probes would primarily activate semantic memories stored in the neocortex and indexed by ERPs such as the N400 and P3b, whereas episodic probes would primarily activate episodic memory stored in mediotemporal and linked cortical structures, indexed by late parietal potentials, such as the LPC (Paller and Kutas, 1992; Rugg et al., 1998; Dien et al., 2004; Voss and Paller, 2006). In practice, most stimuli are associated with both semantic and episodic

memories, and so they would elicit some combination of these effects. For the stimuli in this study, one's date of birth is associated with strong SAM, activating meaning-related processes about oneself in semantic memory but also activating episodic memories incidentally (e.g., events during a specific birthday party, although this may be reduced by providing only the day and month of each date). Prior to the experiment, the birth date is also associated with relatively remote episodic memories of birthday events and other experiences involving one's birth date, such as filling out applications (e.g., for jobs, insurance, taxes). In addition, as the birth date is repeatedly experienced over the course of the experiment, each of these experiences may be encoded as a new (1) episodic memory (Paller and Wagner, 2002) and/or (2) constructed memory that combines new and old (i.e., due to incidental recollection of various birth date memories) episodic elements as well as semantic memory (Hassabis and Maguire, 2009).

New, recent episodic memory encoding can also occur for a different date with no semantic or episodic memory associated with it before the experiment, such as the secret date. Importantly, while multiple trace theory proposes that the hippocampus supports all episodic memories, regardless of how long ago they were encoded (Nadel et al., 2000), some evidence suggests that different parts of the hippocampus support more recent vs. remote episodic memory (Kesner and Hunsaker, 2010; Mankin et al., 2012). Further, between 3 days and 3 months after the learning episode, episodic memories may become semantic by increasing connectivity between cortical areas while decreasing connectivity with the hippocampus (Harand et al., 2012), and a study comparing episodic memories for events ranging in time from very recent (3–14 days old) to very remote (10 years old) found evidence that the hippocampus and the EAM cortical network are integrated more strongly for recent than remote memories (Soderlund et al., 2012). Consequently, more remote memories require more strategic top-down processes in prefrontal cortex for them to be retrieved than do more recent memories. This predicts that ERP effects related to EAM will be greater for the secret date, which involves very recent episodic memory, than the birth date, which involves mostly much more remote episodic memory.

On the other hand, the secret date is minimally meaningful (i.e., low in semantic memory) relative to the birth date. Repeated experiences with any date could potentially begin to construct new semantic memory about that date (Curran et al., 2002; Gratton et al., 2009), but the ability to do so would be minimal because little meaningful information is provided about any dates within the experiment. Notably, the information that the probe is a secret date to be kept concealed during the experiment is meaningful and could lead to learning this as new semantic memory due to repeated experiences with it; knowledge and semantic memory typically require multiple experiences to acquire (Glisky and Schacter, 1987; Verfaellie and Cermak, 1994). Another important way that all these semantic and episodic memory processes could affect the CIT is by inducing standard oddball effects thought to be related to ongoing contextual updating processes in working memory (Kutas et al., 1977; Donchin and Coles, 1988; Dien et al., 2004; Polich, 2007). This could result in a larger P3b to the probes than irrelevants. Further, the P3b to probe

conditions could differ as a function of the relative combination of associated semantic and episodic memory. In sum, the birth date potentially activates a combination of high semantic memory and remote episodic memory for multiple birthdays related events, whereas the secret date potentially activates a combination of low semantic memory and recent episodic memory for a single event. Despite reflecting a combination of memory influences, the birthdate and secret date provide an interesting and important starting point for assessing the role of semantic and episodic memory in CITs.

The focus of this paper is on the frontal N2, N400, P3, and LPC components. The frontal N2 is important because recent studies suggest that concealed information in CITs modulate this component with visual (Gamer and Berti, 2010) and auditory stimuli (Matsuda et al., 2009), with probes eliciting a larger frontal N2 than irrelevants. This would be predicted by orienting reflex theory (Ben-Shakhar and Elaad, 2003), as the probe is more meaningful than the irrelevants and occurs infrequently (it is “novel” within the local stimulus sequence). If the frontal N2 reflects primarily an orienting reflex to meaningful information, the N2 should be larger for (1) probes and targets than irrelevants, and (2) semantic autobiographical information, such as one’s date of birth, relative to recently acquired episodic information, such as a random (secret) date seen just before the study. However, the frontal N2 is known to be modulated by other variables as well, including the extent to which a stimulus matches to memory (e.g., Folstein and Van Petten, 2008; Folstein et al., 2008); the less a stimulus matches memory, the larger the N2. The precise type of memory involved is usually not specified, but knowledge (e.g., of an object category) and working memory have been mainly studied so far. Thus, an alternative prediction can be made based on the idea that match to knowledge is relevant for N2 modulation. The numbers and month abbreviations used as stimuli will activate knowledge about numbers and months, respectively. This predicts that the N2 will be larger to the irrelevants (minimal memory: people have minimal knowledge about the numbers in random dates that have no task relevance) than a meaningful item (e.g., birth date with rich semantic and remote episodic memories). In addition, depending upon how much new memory is encoded for the episodic item (e.g., a “secret” probe date will be associated with new episodic memory and possibly new knowledge induced by repetition within the experimental context), the N2 to this item may be in-between that to irrelevants and the semantic item.

The centroparietal N400 is larger when an item activates semantic memory less relative to more successfully (Kutas and Federmeier, 2011). Although people know the numbers and month abbreviations used to denote dates, an arbitrary date is not very rich in meaning. In contrast, one’s birth date is personally meaningful because it is rich in SAM. This predicts that the N400 will be larger for irrelevant dates than the semantic item (birthdate). In addition, as with the frontal N2, depending upon the extent to which new semantic memory is encoded for the episodic item, its N400 may be in-between that to irrelevants and the semantic item. However, the N2, which merely requires new knowledge to be acquired, may be more sensitive to the memory manipulations in this experiment than the N400,

which requires the more demanding encoding of a meaningful representation. After all, the episodic manipulation can induce new knowledge to be learned, but this new information is minimal in meaning, and meaningful representations would typically require a stronger induction event than that used in this experiment (Gratton et al., 2009). For example, acquisition of category knowledge with minimal associated meaning modulates a fronto-central N2 but not necessarily the N400 (Folstein et al., 2008). The N400 may thus show little or no difference between irrelevants and episodic items, instead differing primarily between irrelevants and semantic items.

The effect of concealed information on the P3 has been investigated in numerous ERP studies (e.g., Rosenfeld et al., 1988; Allen et al., 1992; Rosenfeld et al., 2004), but almost all used fewer than five recording sites and so differences between the spatial distribution of the P3 in the different conditions may have been missed. Indeed, the P3 is a family of components, and what has usually been referred to as P3 in previous studies is most likely an instance of the P3b, which has been dissociated from the P3a (Dien et al., 2004; Polich, 2007; Verleger, 2008). The P3b is known to be modulated by many factors, including the subjective probability of items in a perceived category, the complexity of the task and stimuli, and stimulus value (e.g., Johnson, 1986, 1993). We predicted that the P3b to probes would be larger than to irrelevants, replicating previous findings (e.g., Rosenfeld et al., 2004). Further, the semantic probes might elicit a larger P3b than the episodic probes in part because they were the only items associated with strong semantic memory and so they may stand out more in the stream of irrelevants, which are associated only with episodic information acquired during the study.

Finally, the LPC is typically larger during tasks that entail the reactivation of episodic memories (Rugg and Curran, 2007) and so we expected the LPC to be larger to probes, for which episodic memories have been clearly associated, than to irrelevants, for which episodic memory is minimal, and larger to probes in the episodic than semantic condition.

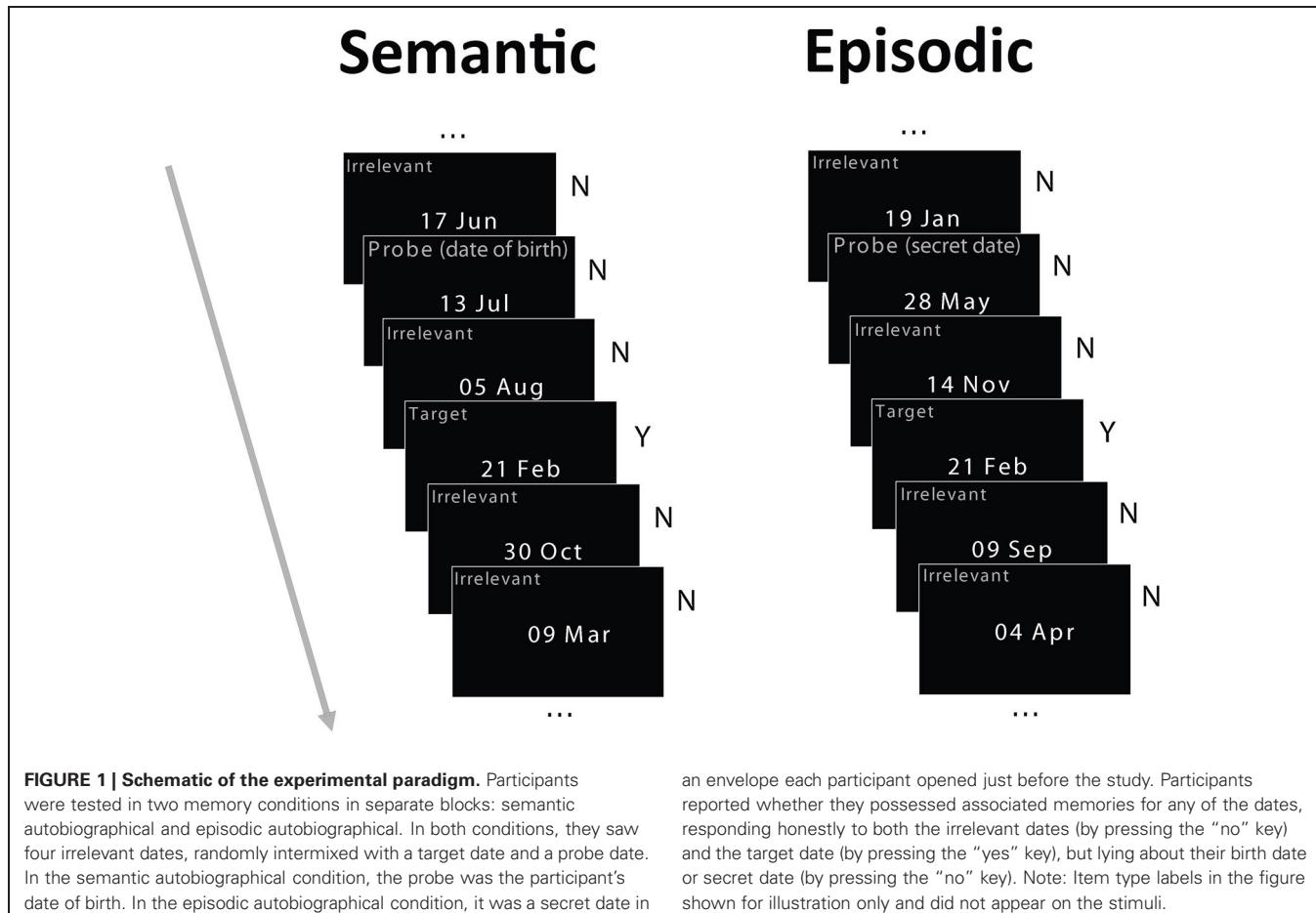
MATERIALS AND METHODS

SUBJECTS

Twenty-five naïve healthy volunteers (18 females, between 18 and 35 years of age, mean = 21, SD = 3.5 average age: z years), recruited from the University of Plymouth (UoP), took part in for course credit. Data from eight participants were excluded due to excessive artifacts (7) or failure to carry out the task as instructed (1). Participants had normal or corrected vision, and no history of neurological or psychiatric disease. All procedures were approved by the UoP Ethics Board.

STIMULI

The stimuli were dates in the format “day month” (e.g., 15 Apr, **Figure 1**) commonly used by our European participants, subtending about $3 \times 2^\circ$ of visual angle. Three types of dates were used in each condition: irrelevants, probe, and target. During the week preceding the study, at the same time detailed and demographics and health questionnaires were administered, participants were asked over the phone to provide their own date of birth (only the day and month were required) and a list of



other important dates (dates of birth of close relatives and friends, anniversaries and so on), so that a set of irrelevant dates could be generated for each participant that excludes these personally important dates. For the semantic autobiographical condition, the probe was the birth date of each participant. For the episodic autobiographical condition, the probe was a date that differed from all other dates used in the study and was not on the participant's list of important dates. The irrelevant dates used for the episodic and semantic conditions were always different. Irrelevant dates never shared the day or the month of the probe or target dates, and they were never famous dates. Furthermore, the target never shared the day or month of the probe.

PROCEDURE

Before beginning the EEG setup, participants were shown a target date and then were unexpectedly taken into an adjacent fire refuge area by an assistant and the experimenter and they were given an envelope containing their "secret" date. Next, the experimenter left the room, and participants were told by the assistant to open the envelope and to memorize the secret date contained in it, ensuring not to do anything that could reveal they knew this date to the experimenter. Participants were also told that this was their own secret date, different from everyone else's, and that they should keep the note it was written

on in their pocket or purse. After setting up the EEG cap and electrodes, participants were seated on a comfortable chair in front of a computer screen (about 114 cm away) in a dark room. Two conditions were administered in separate blocks, the semantic and episodic conditions, with order counterbalanced across participants. In the semantic condition, the probe date was the individual's birth date whereas, in the episodic condition, it was the "secret date." This secret date varied by participant to match the between-participant variability of the date of birth. In both semantic and episodic conditions, participants were instructed to deny possessing any memory for the probe date (birth date or secret date, respectively) throughout the session by giving a deceptive "no" response. They were also instructed to give an honest "yes" response about knowing the target date. Thus, participants had to report honestly whether they knew each date, but they had to lie about the probe date. In sum, participants responded honestly to both the target (pressing "yes") and the irrelevants (pressing "no") but deceptively to the probe (pressing "no"). Participants responded by pressing one of two buttons with the index and middle finger of their dominant hand. They were instructed to respond as fast as possible without sacrificing accuracy. Each item was presented for 800 ms with an inter-trial interval of 3000 ms. In each condition, each item (four irrelevants, one probe, and one target) was presented 35 times in a

pseudo-random order for a total of 210 trials. The constraints on the pseudo-random sequence were that a probe and a target could never appear in temporally adjacent trials, and any individual irrelevant could only repeat for a maximum of three times in the sequence. The same abstract sequence (i.e., the sequence of irrelevant, probe and target types of items) was used for the two conditions to eliminate potential differences due to sequence statistics. Each condition was split into two blocks of ~7 min each, to test the effect of stimulus repetition. There was a short practice session (10 trials) before the experimental trials. Finally, at the end of the study, participants were asked to recall the target and the secret dates and indicate if they had any pre-experiment memory associated with any of the other dates. Since there was 100% recall accuracy in all cases, the recall data were not further analyzed.

ELECTROPHYSIOLOGICAL DATA ACQUISITION

The electroencephalogram (EEG) was sampled at 250 Hz from Ag/AgCl electrodes (gain = 20,000, bandpass filtering = 0.01–100 Hz). EEG data were collected from 32 electrodes arranged in a geodesic array (**Figure 3**) and additional electrodes placed below the right eye referenced to left mastoid to monitor eye blinks, on the tip of the nose, and the right mastoid, all of which were referenced to the left mastoid. Note that in this configuration, Fz is just posterior to site 27, Cz coincides with site 28, and Pz is just posterior to site 29. Horizontal eye movements were monitored using two electrodes placed on the outer canthi of the right and left eyes, referenced to each other. Electrode impedance was below 5 kΩ for all channels.

ANALYSES

Performance measures were submitted to ANOVAs with three factors: item type (average of irrelevants; probe; target), memory condition (semantic and episodic), and repetition (first and second half). To ensure participants carried out the task, follow-up ANOVAs also contrasted targets with irrelevants and targets with probes. However, the main comparison of interest for each memory condition was between probes and irrelevants because the same response (“no”) was associated with both. As this comparison was the main focus of this experiment, and targets received a different response (“yes”) from all other items, confounding their comparison with other items, ERP analyses focus only on an item factor that include probes and irrelevants; note, preliminary analyses that included ERPs to the targets confirms expected target P3b effects. In the following, significant differences between probes and irrelevants (in the behavioral or ERP data) will be referred to as the CIT effect.

Response times

Response times (RTs) and accuracy rates were analyzed in the omnibus ANOVA and planned comparisons.

ERPs

ERPs were averaged off-line for an epoch of 1000 ms, including a 100 ms baseline. Trials affected by blinks, eye movements, muscle activity or amplifier blocking were rejected off-line. An average of 31 artifact-free trials per item type per participant went into the

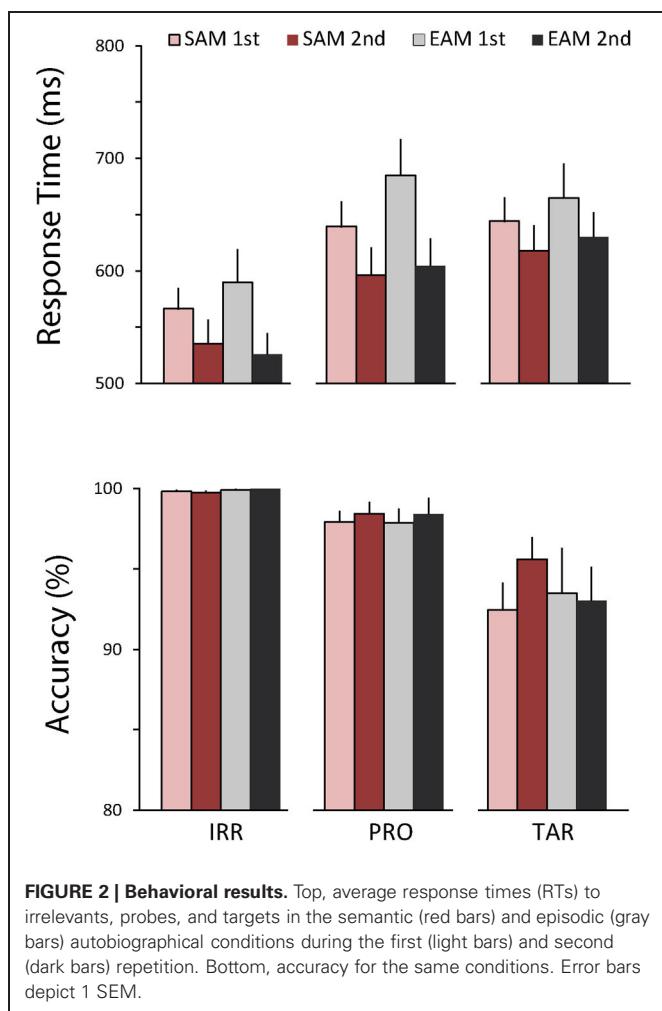
analyses ($MIN = 16, SD = 4.2$). A One-Way ANOVA showed no differences in the number of trials across conditions (including both repetitions), $F_{(5, 85)} = 1.05, p > 0.1, \eta^2 = 0.06$. Data were analyzed unfiltered but shown filtered at low-pass 30 Hz in the figures. Repeated measures ANOVAs on the mean amplitude of the average ERPs assessed the effects of item type and condition on the N2, N400, P3, and LPC components. The time windows used for the main analyses centered arounds the mean peak latency of the N2 (250–350 ms), the N400 (350–500 ms), the P3 (400–600 ms), and the LPC (750–900 ms). To assess the overall pattern of results, a “lateral” ANOVA assessed lateral sites (13 pairs, see electrode montage in **Figure 3**) using factors of Item Type (probes vs. irrelevants), Site, and Hemisphere. A second, “midline” ANOVA assessed the midline sites (six electrodes) using factors of Item Type and Site.

Planned focal analyses were also conducted at frontal sites 1 and 2 for the N2, central site 28 (Cz) for the N400, and parietal site 30 for the P3b and LPC, where these components were maximal. These analyses compared (1) probes and irrelevants (i.e., the CIT effect) in both memory conditions, since we predicted differences between probes and irrelevants in both cases, and (2) probes between the two conditions, since we predicted differences between the semantic and episodic probes. Note that we did not carry out amplitude-latency analyses on the P3b because the overlapping N400 made it difficult to determine P3b peak latency in single participants. The focal analysis was carried out on the mean amplitude data within the time windows used in the main analyses. At focal sites, onset of the CIT effect (i.e., probes vs. irrelevants) and the difference between semantic and episodic probes was determined. For the N2 and N400, 25 ms time windows were used, between 100 ms and 350 ms, and 300 and 550 ms, respectively. A paired *t*-test between the conditions of interest was carried out on each time window until a significant difference was found in three successive time windows. The time window preceding the first significant time window was used as an estimate of the onset time of the effect. For the P3, the same logic was used with 25 ms time windows between 200 and 600 ms.

RESULTS

BEHAVIOR

Figure 2 shows the behavioral results. RTs varied by item type, $F_{(1, 16)} = 54.09, p < 0.001, \eta^2 = 0.77$. Furthermore, RTs were faster in the second than first half of each memory condition block, $F_{(1, 16)} = 21.49, p < 0.001, \eta^2 = 0.57$, and this repetition effect was modulated by item type, $F_{(2, 32)} = 4.73, p < 0.05, \eta^2 = 0.23$. Follow-up analyses to parse this effect compared each item type with the other two. RTs were slower to probes than irrelevants, $F_{(1, 16)} = 73.90, p < 0.001, \eta^2 = 0.82$, and both RTs were faster in the second than first half, $F_{(1, 16)} = 25.32, p < 0.001, \eta^2 = 0.61$, but the repetition effect tended to be marginally larger in the episodic than semantic condition, $F_{(1, 16)} = 3.68, p = 0.07, \eta^2 = 0.19$. Similarly, RTs were also slower to targets than irrelevants, $F_{(1, 16)} = 85.84, p < 0.001, \eta^2 = 0.84$, and both RTs were faster in the second than first half, $F_{(1, 16)} = 12.35, p < 0.005, \eta^2 = 0.44$, but the repetition effect tended to be marginally larger in the episodic than semantic condition, $F_{(1, 16)} = 3.22, p = 0.09, \eta^2 = 0.17$. In contrast, RTs to targets and probes were similar, and



both RTs were faster in the second than first half, $F_{(1, 16)} = 24.11$, $p < 0.001$, $\eta^2 = 0.60$, but probes were slower than targets in the first half, whereas the opposite held in the second half, $F_{(1, 16)} = 8.30$, $p < 0.05$, $\eta^2 = 0.34$. Accuracy showed only a main effect of item type, $F_{(1, 16)} = 15.36$, $p < 0.001$, $\eta^2 = 0.49$. Follow-up analyses revealed that accuracy was lower for targets than both irrelevants, $F_{(1, 16)} = 17.11$, $p < 0.001$, $\eta^2 = 0.52$, and probes, $F_{(1, 16)} = 13.36$, $p < 0.005$, $\eta^2 = 0.46$. Accuracy was also lower for probes than irrelevants, $F_{(1, 16)} = 11.67$, $p < 0.005$, $\eta^2 = 0.42$. Notably, there were no significant main effects of memory on RTs and accuracy and no significant repetition effects on accuracy.

EVENT-RELATED POTENTIALS (ERPs)

Qualitatively, the ERP waveform showed an occipitotemporal P1 and a corresponding anterior N1, followed by a frontocentral P2 and N2, and a centroparietal N400, P3b, and LPC (Figures 3–7). Four main differences between items and memory conditions are evident in the ERPs. The first difference is on the N2, maximal at frontal sites between 250 and 350 ms (Figures 3, 4 and 5A). Second, a clear N400 component overlapping the first part of the P3b is present in the episodic condition, maximal at central sites, and to a lesser extent in the semantic condition

(Figures 3 and 5B). The third difference is on the P3b, maximal at centroparietal sites between 400 and 600 ms (Figures 4 and 5C). Fourth, LPC differences appear later at the same sites, lasting until the end of the epoch (Figures 4 and 5C). Omnibus statistics are shown in Tables 1–3 and described below.

N2 (250–350 ms) and N400 (350–500 ms)

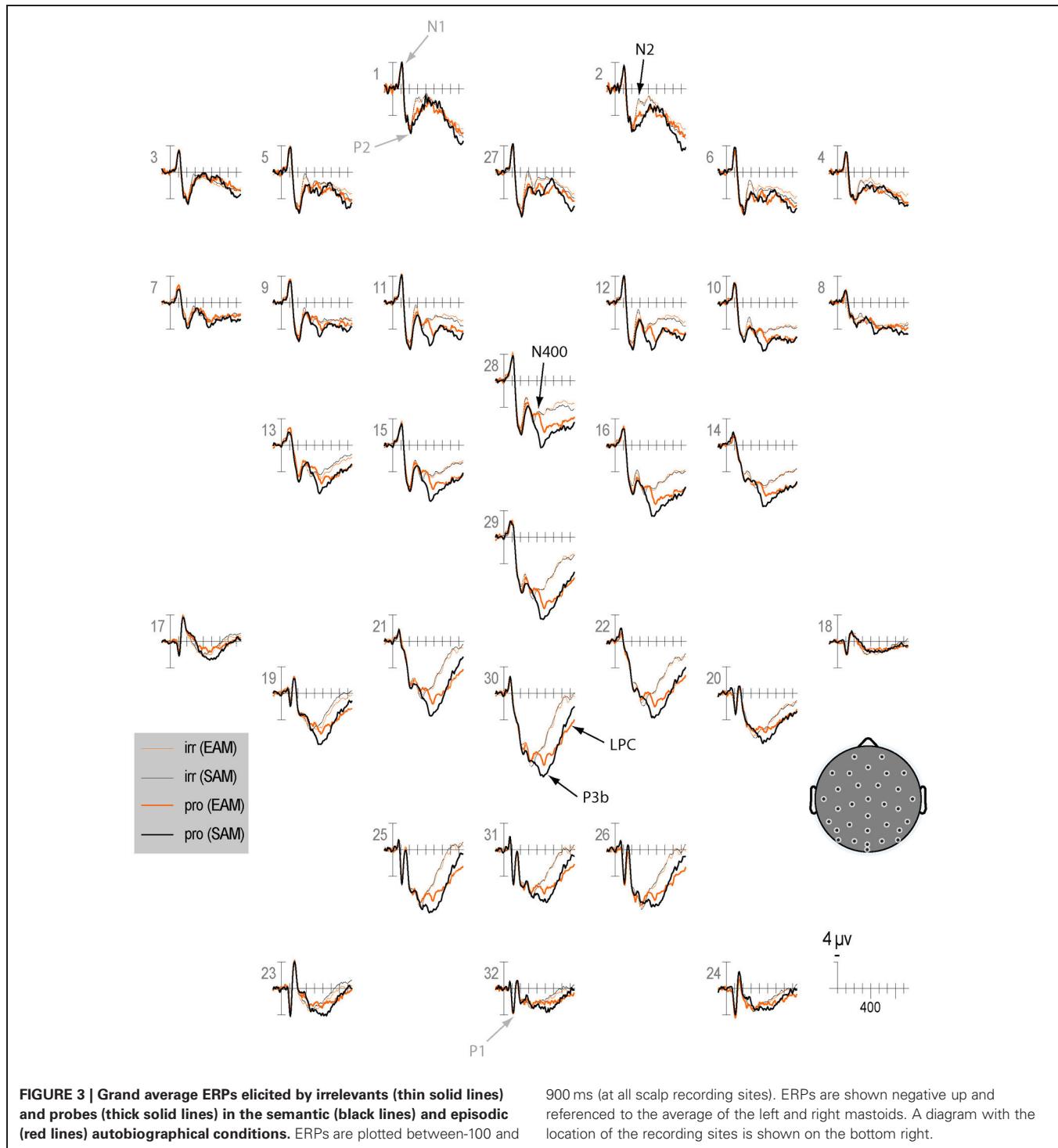
N2. Omnibus results at lateral and midline sites (Table 1) showed a larger N2 for irrelevants than probes at frontal and frontocentral sites ($I \times S$), and ERPs were more positive during the second than the first half (R, lateral sites, 3.81 vs. 3.31 μ V, respectively; midline sites, 5.55 vs. 4.79 μ V, respectively, Figures 6 and 7). At lateral sites, repetition effects were maximal at centroparietal sites and larger over the right hemisphere at frontocentral sites, but symmetric or larger over the left hemisphere at more posterior sites (R \times S \times H).

Planned focal analyses on frontal pair 1 and 2 showed that probes were more positive than irrelevants (4.35 vs. 2.10 μ V, respectively), $F_{(1, 16)} = 40.01$, $p < 0.001$, $\eta^2 = 0.71$, and this CIT effect tended to be larger on the right than the left (3.46 vs. 2.99 μ V, respectively), $F_{(1, 16)} = 4.01$, $p = 0.063$, $\eta^2 = 0.20$. Importantly, this CIT effect was larger in the semantic than episodic condition, $F_{(1, 16)} = 5.53$, $p < 0.05$, $\eta^2 = 0.26$, due to the probes in the semantic condition being more positive than those in the episodic condition (4.87 vs. 3.83 μ V, respectively), $F_{(1, 16)} = 4.50$, $p < 0.05$, $\eta^2 = 0.22$. This result is the opposite of the hypothesis that the N2 CIT effects reflect orienting to novelty but consistent with the alternative hypothesis that the N2 is sensitive to match to knowledge. No repetition effects were significant.

The onset of the CIT effect in the two memory conditions was determined at right frontal site 2, where the differences were largest. Results showed that the CIT effect onset between 200 and 225 ms in the semantic probe condition, and slightly later, between 225 and 250 ms, in the episodic probe condition. A second onset analysis showed that the onset of the difference between probes in the two memory conditions was also between 225 and 250 ms.

N400. The N400 is the only ERP component to show a CIT effect only in the semantic condition. The N400 is smallest for the semantic probe, relative to the episodic probe and all irrelevants, which are indistinguishable from each other (Figure 3). Figure 4 (middle) shows an overall centroparietal scalp distribution between 400 and 600 ms due to the combination of the central CIT effect on the earlier N400 and the parietal CIT effect on the later P3b. Figure 5B shows the memory effect around central sites where the N400 overlaps least with the frontal N2 and parietal P3b, illustrating that the N400 is more negative to episodic than semantic probes and has a central maximum and overall centroparietal scalp distribution, which is characteristic of the N400 index of semantic memory (Kutas and Federmeier, 2011). The omnibus analyses on the N2 and P3b capture the early and late part of the N400, so the focus was on planned focal analyses.

A focal analysis on Cz (site 28) showed that probes were less negative than irrelevants (7.29 vs. 5.16 μ V, respectively),



$F_{(1, 16)} = 7.05, p < 0.05, \eta^2 = 0.31$, and this effect was larger in the semantic than episodic condition (3.27 vs. 0.96 μ V, respectively), $F_{(1, 16)} = 6.16, p < 0.05, \eta^2 = 0.28$. A follow-up analysis showed that the difference between probes and irrelevants was only significant in the semantic condition, $t_{(16)} > 2.35, p < 0.05$, for both repetitions. ERPs were more positive during the second than first repetition, $F_{(1, 16)} = 4.58, p < 0.05, \eta^2 = 0.22$, but

this effect did not interact with any other factors. Finally, ERPs were more positive during the semantic than episodic conditions, $F_{(1, 16)} = 8.73, p < 0.01, \eta^2 = 0.35$.

Finally, an onset analysis of the CIT effect in both memory conditions was carried out at Cz. Results showed that the CIT effect onset between 400 and 425 ms in the semantic conditions, whereas it onset between 475 and 500 ms in the episodic

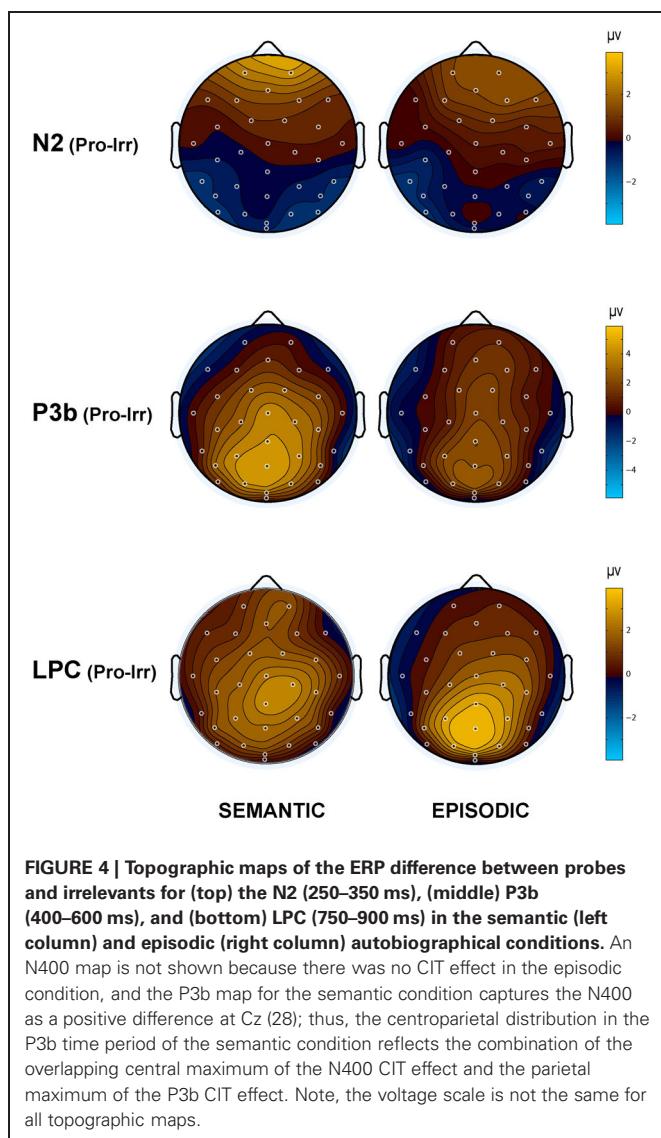


FIGURE 4 | Topographic maps of the ERP difference between probes and irrelevants for (top) the N2 (250–350 ms), (middle) P3b (400–600 ms), and (bottom) LPC (750–900 ms) in the semantic (left column) and episodic (right column) autobiographical conditions. An N400 map is not shown because there was no CIT effect in the episodic condition, and the P3b map for the semantic condition captures the N400 as a positive difference at Cz (28); thus, the centroparietal distribution in the P3b time period of the semantic condition reflects the combination of the overlapping central maximum of the N400 CIT effect and the parietal maximum of the P3b CIT effect. Note, the voltage scale is not the same for all topographic maps.

condition. A second onset analysis showed that the onset of the difference between probes in the two memory conditions was between 400 and 425 ms.

P3b (400–600 ms)

Omnibus results (**Table 2**) showed a larger P3b for probes than irrelevants at lateral (I, 6.84 vs. 4.30 μ V, respectively) and midline sites. This CIT effect was maximal at lateral and midline centroparietal sites (I \times S), and lateral results showed that this effect was larger on the right at frontocentral sites but on the left at more posterior sites (I \times S \times H). Importantly, the difference between probes and irrelevants was larger in the semantic than episodic condition at lateral and midline sites (M \times I), and this interaction was largest at centroparietal sites (lateral, M \times I \times S). ERPs at this time tended to be more positive during the second than the first half (R). At lateral sites, this repetition effect was larger on the right at frontal and posterior sites, but symmetrical at central sites, and maximal at

right centroparietal sites (R \times S \times H). The lateral repetition effect was also modulated by item type, as it was larger for probes than irrelevants (I \times R \times S \times H), and by memory type, as it was larger at centroparietal sites in the semantic condition, but at fronto-central sites in the episodic condition (M \times R \times S).

Planned focal analyses were conducted at parietal site 30 where the P3b was maximal. Consistent with the omnibus analysis, the P3b was larger for probes than irrelevants, $F_{(1, 16)} = 38.35, p < 0.001, \eta^2 = 0.71$ (10.77 vs. 7.15 μ V, respectively). Importantly, this CIT effect was larger in the semantic than episodic condition, $F_{(1, 16)} = 5.26, p < 0.05, \eta^2 = 0.25$ (4.80 vs. 2.45 μ V, respectively) because the P3b was more positive for the semantic than episodic probes, $F_{(1, 16)} = 4.53, p < 0.05, \eta^2 = 0.22$ (11.84 vs. 9.7 μ V, respectively). There was a non-significant trend for the P3b to be larger during the second than the first half, $F_{(1, 16)} = 3.56, p = 0.08, \eta^2 = 0.18$, and the CIT effect was numerically larger during the second than the first half, but this interaction of item and repetition was also not significant, $F_{(1, 16)} = 2.47, p = 0.14, \eta^2 = 0.13$ (4.11 vs. 3.14 μ V, respectively). Thus, the CIT effect did not change significantly as a function of repetition (if anything, it became slightly larger).

The onset of the CIT effect in the two memory conditions was determined at parietal site 30 where the differences were largest. Results showed that the CIT effect onset between 375 and 400 ms in the semantic probe condition, and, later, between 450 and 475 ms in the episodic probe condition. The onset of the difference between the probes in the two memory conditions was also analyzed, revealing an onset between 375 and 400 ms.

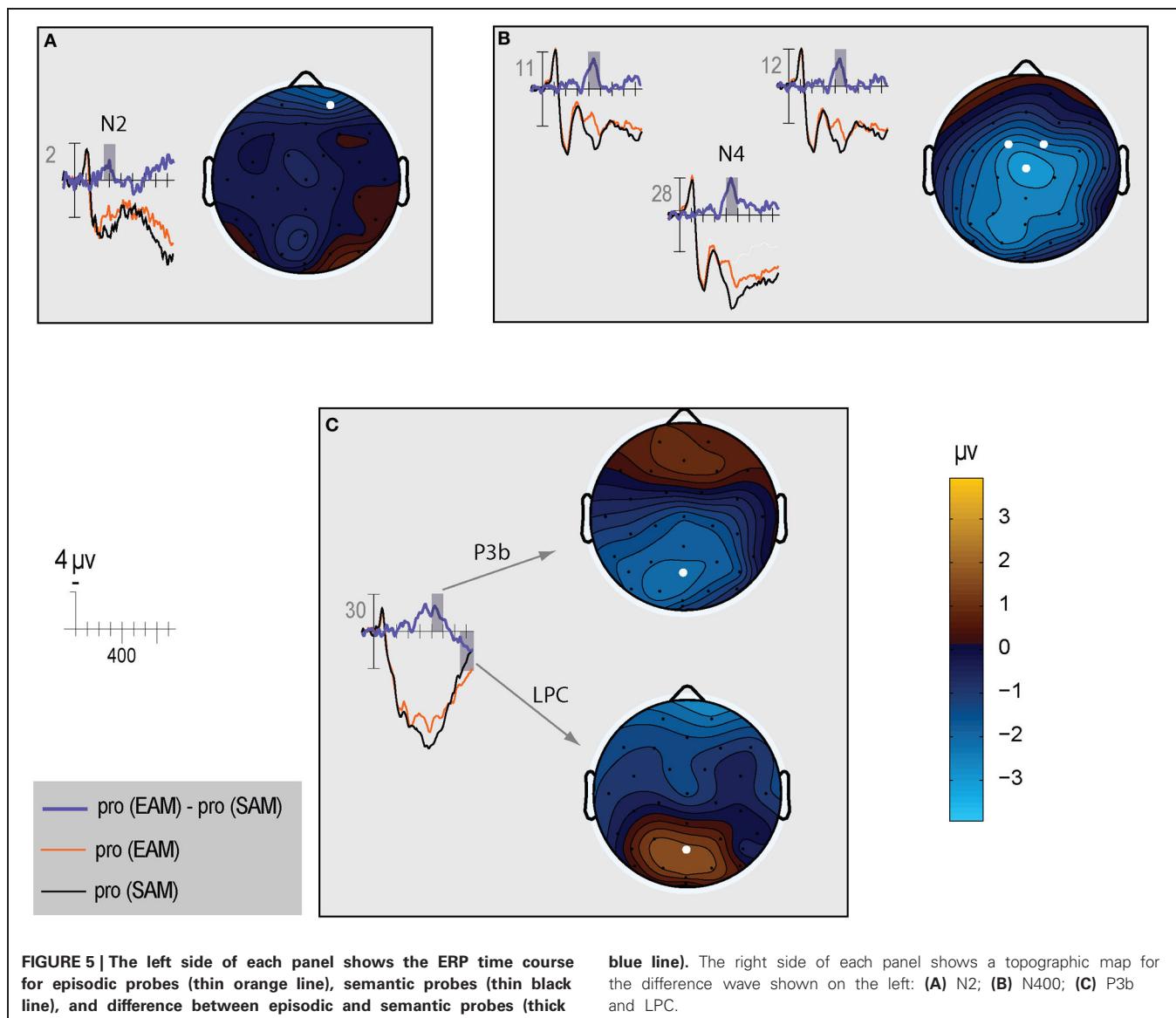
LPC (750–900 ms)

Omnibus analyses (**Table 3**) showed that the LPC was more positive for probes than irrelevants at lateral (I, 3.77 vs. 2.29 μ V, respectively) and midline sites, and these CIT effects were largest at lateral and midline centroparietal sites (I \times S). The lateral ANOVA also revealed that the LPC was larger in the second than first half, and more so over the right hemisphere (R \times H), and a significant four-way interaction indicated that the CIT effect was further modulated by repetition and hemisphere (I \times R \times S \times H). In contrast, the midline ANOVA also revealed that the CIT effect was larger in the semantic than episodic condition at centroparietal sites (M \times I \times S).

Planned focal analyses conducted at parietal site 30 where the LPC was maximal confirmed that the LPC was more positive for probes than irrelevants, $F_{(1, 16)} = 30.19, p < 0.001, \eta^2 = 0.65$ (3.79 vs. 0.64 μ V, respectively), and in the episodic than semantic condition, $F_{(1, 16)} = 4.97, p < 0.05, \eta^2 = 0.28$ (2.70 vs. 1.72 μ V, respectively). Follow-up analyses showed that probes in the episodic condition elicited a larger LPC than probes in the semantic condition $F_{(1, 16)} = 5.25, p < 0.05, \eta^2 = 0.25$ (4.67 vs. 2.91 μ V, respectively). No repetition effects were significant.

DISCUSSION

In summary, performance is consistent with previous CIT studies using the 3-stimulus paradigm with speeded responses in that responses for probes and targets are slower and less accurate than



for irrelevants (e.g., Gamer et al., 2007; Gamer and Berti, 2010). The present results also provide evidence for repetition priming, as responses to all items are faster in the second than the first half of each memory condition block, on average. ERPs show multiple effects. First, the frontal N2 is larger to irrelevants than both types of probes, and the CIT effect on the N2 starts by 225 ms in the semantic condition but slightly later, by 250 ms, in the episodic condition. Second, semantic and episodic probes begin to be processed differently by 250 ms, and this early effect is maximal at frontal sites, where the N2 is larger for episodic than semantic probes. Third, the N400 shows a CIT effect only in the semantic condition, as a central N400 is smaller for the semantic probe relative to the episodic probe and irrelevants, which resemble each other. Fourth, probes generate a larger P3b than irrelevants, and this CIT effect starts by 400 ms and is larger for semantic than episodic probes. Fifth, episodic probes generate a larger LPC than semantic probes. Sixth, although ERPs became more positive in

the second half of the trials, the CIT effect on the P3b remains similar. Next, we discuss these findings in turn.

PERFORMANCE

The behavioral results indicate that probes and targets are more difficult to process than irrelevants. The typical explanation for this finding is that both infrequent probes and targets stand out in the stream of irrelevants but require different responses. This creates a conflict that takes some time to resolve (Gamer et al., 2007). Note that, since targets were the only items requiring a “yes” response, the direct comparison between targets and irrelevant is not very informative. The pattern of behavioral effects was the same for both memory conditions. This indicates that the ERP differences between these conditions do not reflect RT or accuracy differences, and suggests that the probes in these two conditions were similar in terms of saliency. The overall repetition effect on the RTs, but the lack of a repetition effect

Table 1 | Results of the omnibus lateral (Lat) and midline (Mid) ANOVAs for the N2 (probe vs. irrelevants, 250–350 ms).

N2								
Lat			Mid					
F	p	η^2	F	p	η^2	F	p	η^2
I	3.87	0.07	0.19	◦	1.68	0.21	0.09	
I × S	14.14	0.00	0.47	***	9.57	0.00	0.37	***
I × S × H	1.47	0.19	0.08					
R	6.42	0.02	0.29	*	8.78	0.01	0.35	**
R × S	1.84	0.16	0.10		2.62	0.07	0.14	◦
R × S × H	2.16	0.03	0.12	*				

I, Item; R, Repetition; S, Site; H, Hemisphere. ◦ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table 2 | Results of the omnibus lateral (Lat) and midline (Mid) ANOVAs for the P3b (probe vs. irrelevants, 400–600 ms).

P3b								
Lat			Mid					
F	p	η^2	F	p	η^2	F	p	η^2
M × S	1.39	0.25	0.08		6.33	0.02	0.28	*
I	15.17	0.00	0.49	***	26.20	0.00	0.62	***
I × S	6.64	0.00	0.29	**	7.12	0.00	0.31	**
I × S × H	3.06	0.00	0.16	**				
R	2.99	0.10	0.16		4.39	0.05	0.22	◦
R × H	4.17	0.06	0.21	◦				
R × S × H	2.05	0.03	0.11	*				
M × I	7.45	0.02	0.32	*	6.33	0.02	0.28	*
M × I × S	3.49	0.04	0.18	*	2.33	0.09	0.13	
M × R × S	3.08	0.03	0.16	*	2.49	0.07	0.13	
I × R × S × H	2.66	0.01	0.14	**				

M, Memory; I, Item; R, Repetition; S, Site; H, Hemisphere. ◦ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

on the difference between probes and irrelevants, indicates that repetition had mostly a generic effect independent of item type.

FRONTAL N2 AND CENTROPARIETAL N400: KNOWLEDGE AND SEMANTIC MEMORY

N2

At least two types of frontal N2 components have been distinguished, a cognitive control N2 and a memory (mis)match N2, whose amplitude is modulated by different factors (Folstein and Van Petten, 2008; Folstein et al., 2008). Concealed information studies have focused on the cognitive control N2 and most used an orienting reflex account. Clearly, the pattern of effects on the frontal N2 found here is not consistent with a simple orienting reflex explanation (Sokolov, 1963). The N2 is largest for the frequent irrelevants which, according to an orienting reflex account, should be the least salient stimuli and so should be associated instead with the smallest N2.

Table 3 | Results of the omnibus lateral (Lat) and midline (Mid) ANOVAs for the LPC (probe vs. irrelevants, 750–900 ms).

LPC									
Lat			Mid						
F	p	η^2	F	p	η^2	F	p	η^2	
M × S			4.15	0.01	0.21	*	0.52	0.48	0.03
I			31.24	0.00	0.66	***	29.80	0.00	0.65
I × S			4.79	0.00	0.23	**	9.09	0.00	0.36
R × H			4.72	0.05	0.23	*			
M × I × S			1.57	0.17	0.09		2.93	0.02	0.15
M × R × S × H			1.73	0.09	0.10	◦			
I × R × H			3.52	0.08	0.18	◦			
I × R × S × H			2.47	0.02	0.13	*			

M, Memory; I, Item; R, Repetition; S, Site; H, Hemisphere. ◦ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Most previous CIT studies using ERPs have focused exclusively on the P3b component, making it difficult to compare our results with those of the previous literature. More troublesome, the low-pass filtering employed in some studies is so extreme (around 4 Hz in some cases) that any effects on fast changing components like the frontal N2 would be wiped out (Rosenfeld et al., 2006). However, two recent CIT studies examined the effect of the experimental manipulations on components of the N2-family (Matsuda et al., 2009; Gamer and Berti, 2010). The study by Matsuda and collaborators used a 2-stimulus paradigm (i.e., no targets were present), auditory stimuli of an episodic nature (single digits), long interstimulus intervals (22 s) to enable peripheral psychophysiological recordings, and a common average reference montage, making it difficult to compare the results with those of the current study with visual stimuli, fast intertrial intervals, and average mastoid reference. Their findings showed a slightly larger central N2 for probes than irrelevants, which the authors suggest is an N2b reflecting the redirection of attentional resources to salient stimuli (Matsuda et al., 2009). The study by Gamer and Berti (2010) is perhaps more comparable to the present study since it employed visual stimuli, a 3-stimulus paradigm, and a right mastoid reference. This study reported a larger frontal N2 to probes than irrelevants, and attributed such an effect to cognitive control processes required for response monitoring. Such an explanation would predict a larger frontal N2 for probes than irrelevants in the current study as well, but the opposite was found. It is possible that differences in the paradigms could account for this discrepancy: The stimuli differed (playing cards were used in that study compared to dates here), different interstimulus intervals were used (8 s, on average in that study vs. 2 s here), and stimuli were not counterbalanced across participants in the earlier work, leaving open the possibility of item-specific confounds. These differences in the paradigm clearly resulted in ERP differences compared with standard CIT results as, for example, there was no P3b effect. Furthermore, a subsequent CIT study by the same group (Gamer and Berti, 2012) failed to find any N2 effects. Although further work is required to fully characterize the

factors that affect the frontal N2 in CIT paradigms, the current study shows that concealed information is not necessarily associated with a larger frontal N2 in CIT paradigms and that the literature is inconsistent.

The pattern of N2 effects suggests that the degree to which an item matches memory, a factor known to modulate frontal N2 amplitude (i.e., larger N2 for memory mismatch), is the key factor modulating the N2 in this study. This interpretation is supported by the finding that the frontal N2 is smaller for the date of birth, the item associated with the most semantic (and remote episodic) memory, followed by the secret date, which is associated with recent episodic memory for the learning experience in which the participant received the envelope with this date (and perhaps some newly acquired semantic memory, e.g., the fact that the date is a secret), and by the irrelevants, which have very little associated semantic or episodic memory. Based on this finding alone, we cannot rule out that this N2 effect reflects both semantic and episodic memory, but prior evidence implicates semantic memory more. The N2 memory match effect has primarily been found when knowledge is manipulated, not in episodic memory experiments. This knowledge is not necessarily semantic (meaning) *per se*, because frontal effects, especially frontopolar ones where the N2 is maximum here, are not always found with semantic manipulations (Ganis and Kutas, 2003; Kutas and Federmeier, 2011). A visual knowledge interpretation is also indicated by evidence that a frontal N3(00) complex from 200 to 500 ms, which includes the memory match N2 as an early component of this waveform, is specific to processing visual images (Barrett and Rugg, 1989, 1990; McPherson and Holcomb, 1999) and modulated according to how successfully visual knowledge is activated for a category decision (e.g., dog, cat, car) (Schendan and Kutas, 2002, 2003; Schendan and Maher, 2009). Finally, it is noteworthy that ERPs in the N2 time window became more positive with repetition, but the effect did not vary by item type. Such an increase over multiple repetitions may reflect accumulation of knowledge with each repetition, as in category learning, which can modulate the frontal N2, the N400, and other ERPs (Curran et al., 2002; Folstein and Van Petten, 2004; Folstein et al., 2008; Scott et al., 2008; Gratton et al., 2009).

N400

This interpretation of the N2 is consistent with the modulation of the N400 index of semantic memory, which is clearest at central sites [see site Cz(28) in **Figure 5B**] where the N400 overlaps least with the frontal N2 and the parietal P3b. The N400 is more negative for episodic than semantic probes because the amplitude of the N400 is inversely proportional to the amount of semantic memory associated with both linguistic and non-linguistic stimuli (e.g., Kutas and Federmeier, 2011): A secret random date acquired just before the study has little or no semantic memory associated with it, compared to one's birth date, which is by far the most meaningful stimulus. Consequently, meaning activates most successfully for this semantic probe, and its N400 is smallest. In contrast, the N400 is larger to the episodic probe (secret date), but comparably as large to the irrelevants. The similarity between the N400 to the episodic probe and the irrelevants is consistent with the fact that the meanings of all these items are

minimal and about the same (i.e., just the meaning of the numbers and months but no other richly meaningful facts). This also indicates that the new information about the secret date acquired before the study did not result in a sufficiently meaningful representation to affect the semantic memory processes underlying the N400. Notably, in contrast, the secret date information did result in new knowledge, such as visual knowledge, as demonstrated by the smaller frontal N2 for the episodic probe relative to the irrelevants. This is consistent with evidence that certain types of newly acquired knowledge result in sensitivity of the frontal N2 (and similar frontal negativities between 200 and 500 ms, e.g., frontal N3 complex, N300, N350, N390 components) to this knowledge (Curran et al., 2002; Ganis and Kutas, 2003; Folstein and Van Petten, 2004, 2008; Folstein et al., 2008; Schendan and Maher, 2009), but additional more richly meaningful information needs to be provided for the centroparietal N400 to become sensitive to newly acquired facts about an item (Gratton et al., 2009). Importantly for deception detection, this means that the N400 shows a CIT effect for semantic autobiographical information, and quite a robust one, but minimal to no CIT effect for episodic autobiographical information.

Previous CIT studies have not reported N400 effects for several reasons. First, the N400 effect is largest at Cz(28) but overlaps to some extent the P3b spatiotemporally at this site and parietal sites. A number of CIT studies have used only the three midline sites Pz, Cz, and Fz, or reported data only for those sites (e.g., Rosenfeld et al., 1988, 2004, 2006; Gamer and Berti, 2010), and so may have missed N400 effects or analyzed them as part of the P3b effects. Second, as mentioned in the context of the frontal N2, extreme low-pass filtering to enhance slow components like the P3b might have spuriously reduced effects on faster-varying components such as the N400 (e.g., Rosenfeld et al., 2006). A third reason is suggested by the present finding of a CIT effect on the N400 only in the semantic condition. Episodic stimuli do not show a CIT effect because, in the present work and many previous studies, they do not have sufficiently rich semantic memory representations to produce a CIT effect on the N400.

From a memory perspective, it is necessary to consider the alternative that the frontal N2 and centroparietal N400 effects reflect instead episodic memory. Indeed, frontal negativity between 100 and 300 ms (during the N2) does show memory effects, being more negative for new than old items during recognition tasks (Tsivilis et al., 2001), but the interpretation of such repetition effects, and similar ones on the N400, is controversial (Rugg and Curran, 2007) and, if anything, points to knowledge, conceptual memory, and semantic memory (Paller et al., 2007; Voss et al., 2010; Voss and Federmeier, 2011). These issues have been discussed in detail in the debate about whether an N400-like component, which sometimes appears to have a more frontal (and so labeled "FN400") than centroparietal distribution, reflects episodic familiarity or conceptual implicit memory (due to activation of meaning representations) (Paller et al., 2007; Rugg and Curran, 2007; Voss and Federmeier, 2011). While this debate is beyond the scope of this paper, it is relevant to consider whether the frontal N2 or N400 pattern might reflect episodic familiarity. We suggest that familiarity can't simply or easily explain the N400. First, one might argue that the semantic

probe (one's birthday) has more lifetime familiarity than the episodic probe (an arbitrary date with no other meaning) because episodic memories set up prior to the EEG recording are numerous (albeit more remote) for one's birthday but only singular (albeit more recent) for the episodic probe. These pre-existing episodic memories for one's birthday therefore reduce the N400 (or N2) for the semantic more than the episodic probe. Second, both semantic and episodic probes are equally as familiar in terms of exposure during EEG recording (i.e., repeated the same number of times), which is the typical way that episodic familiarity is defined experimentally. This predicts no difference between semantic and episodic probes, in contrast to the clear memory effects observed. Third, connectivity between the hippocampus and neocortex is stronger for recent episodic memory, which is the kind primarily associated with the episodic probe, relative to remote episodic memory, which is only associated with the semantic probe (Soderlund et al., 2012). Such differential hippocampal-cortical linkages would predict a greater reduction in episodic memory-related cortical activity for the episodic than semantic probe, thereby resulting in a smaller N400 (or N2) for the episodic than semantic probe—the opposite of the observed pattern. Fourth, altogether, these episodic memory considerations would predict a larger N400 for the irrelevants than the episodic probe because the episodic probe was studied beforehand but the irrelevants were not (and so are less familiar), but no evidence was found for any difference between the episodic probe and irrelevants. The parsimonious explanation is that the consolidated semantic memory in the cortex for the semantic probe drives the N400 pattern, as argued here. Consistent with this, the FN400 has been argued to be identical to the N400 and to reflect semantic memory and conceptual implicit memory for repeated items (Paller et al., 2007; Voss et al., 2010; Voss and Federmeier, 2011). Note, as all items repeated here many times, conceptual priming (due to conceptual implicit memory) could explain the N400 pattern, not only semantic memory (Renoult and Debraille, 2011). The N400 shows robust modulation with conceptual priming, being smaller for repeated than new meaningful items (Paller et al., 2007). Conceptual priming would be greater for the meaningful semantic probe than the minimally meaningful episodic probe (Voss et al., 2010), consistent with the observed pattern.

Together, these findings suggest that the N2 and N400, as highly sensitive markers of knowledge and semantic memory, respectively, could potentially be used for detecting concealed information, but only if the type of memory is considered carefully and the concealed information that one is trying to detect is stored in the brain systems for knowledge and semantic memory. In contrast, if the goal is to detect episodic memory, then later brain potentials, like the P3b and LPC, may be more suitable markers. In most realistic cases, in which the probes are associated with both semantic and episodic memory, both types of markers should be considered.

P3b

The main prior study that addressed an issue similar to the one addressed here is the one by Rosenfeld and Collaborators (2006). The relevant finding from that study is that the P3b difference

between probes and irrelevants was much smaller for low-impact probes (the recently learned experimenter's name) than high-impact probes (a participant's name). In fact, the difference between probes and irrelevants in the low-impact condition was close to zero. Like that study, we found that semantic probes elicit a larger P3b than episodic ones: The CIT effect is larger in the semantic than episodic condition. Even so, at least in the analysis within a fixed P3b window, episodic probes show a sizeable CIT effect on the P3b. One possible explanation for this discrepancy is that, in the prior study, the low-impact probes were incidentally learned, even though they were encountered numerous times in the experiment. In the current study, participants were explicitly told that the (episodic) probe was a secret date that they had to lie about. This constitutes intentional encoding, which results in greater episodic memory than incidental encoding and likely also increases the saliency of such a date (Hyde and Jenkins, 1973; Craik and Tulving, 1975; Kellogg et al., 1982). Given the intentional nature of deception, intentional study would also be expected to transfer more appropriately to the intentional retrieval situation of the CIT paradigm than incidental study (Tulving and Thomson, 1973; Morris et al., 1977). Since the stimulus sequences used in the two memory conditions were identical, this finding confirms that the P3b is modulated by the type of memory triggered by the probe, not just by context updating taking place in working memory (Johnson, 1986, 1993). Importantly, the CIT effect on the P3b did not become smaller with repetition, but rather, tended to become larger. This indicates that the duration of the test is not a major issue in P3-based CITs, and the benefit of longer ERP sessions with more trials may not be cancelled by habituation effects, as usually seen with electrodermal measures (Ben-Shakhar et al., 1975). Future work will have to determine whether the CIT effect on the P3b is constant for even longer sessions that may be required in the field. It is noteworthy that our results may underestimate the size of the P3b in the episodic condition during its initial phase when it overlaps the N400 at central sites due to the opposing polarities of these ERPs (Figures 3–5), but not afterwards around the P3b peak and thereafter from 500 to 600 ms.

The P3b pattern bolsters the interpretation of the earlier frontal N2 and centroparietal N400 patterns in terms of knowledge and semantic memory. Both the present findings and the Rosenfeld et al. (2006) study indicate that the CIT effect on the P3b is larger for semantic items (e.g., your own birth date and name, respectively) relative to items that are less meaningful or about which one has less knowledge (e.g., a secret date, irrelevants). Likewise, in experiments on semantic memory using an object categorization task, a parietal P3b-like component, peaking around 600 ms, is more positive for objects categorized more than less successfully (Schendan and Kutas, 2002, 2003; Schendan and Maher, 2009). Consequently, in the present study, the P3b is larger for the semantic than episodic probe possibly because subjects more successfully identify the semantic probe as their birthdate relative to the episodic probe as the secret date and discriminate the semantic better than the episodic probe from the irrelevant dates in order to generate a deceptive response; after all, the episodic probe and the irrelevants have minimal to no meaning and so they may be less discriminable from each other

in terms of knowledge and semantic memory. Altogether, the N2, N400, and P3b all point to the importance of knowledge and semantic memory for demonstrating a CIT effect on these ERPs.

LPC

Even though previous CIT studies have considered episodic memory, no previous ERP study has examined specifically the LPC, which is well established as a marker of conscious recollection from episodic memory (Paller and Kutas, 1992; Rugg and Curran, 2007). In the present study, episodic probes elicit a larger LPC than semantic ones. This finding is consistent with studies of episodic memory in which people decide whether an item is new or old (e.g., Paller et al., 1995). In these studies, a larger LPC is typically found for old items recognized as such (e.g., items with associated episodic knowledge) relative to new items. The parietal distribution and time course of this old/new LPC resembles the LPC memory effect found here. The present LPC finding thus indicates that, when the concealed information is thought to be primarily or predominantly due to episodic memory, then the LPC may be the most robust ERP to examine in CIT paradigms. Intriguingly, the LPC pattern is the only ERP finding that parallels the RT pattern: The LPC is more positive and RTs are slower for episodic than semantic probes, which are slower than irrelevants. However, the LPC starts after the RTs in the episodic condition, on average, suggesting that the recollection process underlying the LPC cannot drive the RT effect.

Repetition

Repetition effects should be explored in future ERP work, especially given the sensitivity demonstrated here of RTs and accuracy to this factor. Overall, RTs and ERPs between 250 and 350 ms, 400 and 600 ms, and 750 and 900 ms show repetition effects, but focal analyses on the N2, N400, P3b, and LPC show no repetition effects, perhaps due to insufficient power. Intriguingly, RT repetition effects to all item types (i.e., faster responses in the second than first half) tend to be larger (albeit non-significantly) in the episodic than semantic condition. However, no ERP repetition effect is larger in the episodic than semantic condition, but given the weakness of the RT interaction, power may have been insufficient to detect this also in the ERPs. Nonetheless, it appears that the N2 and N400 are larger for episodic than semantic in the first (**Figure 6**) more than the second half of trials (**Figure 7**), whereas the P3b is larger for semantic than episodic, and the LPC is larger for episodic than semantic in the second more than first half. Future ERP studies should manipulate repetition with a greater number of trials and in more subtle ways and evaluate whether repetition modulates the CIT effect on these ERPs and if so, how.

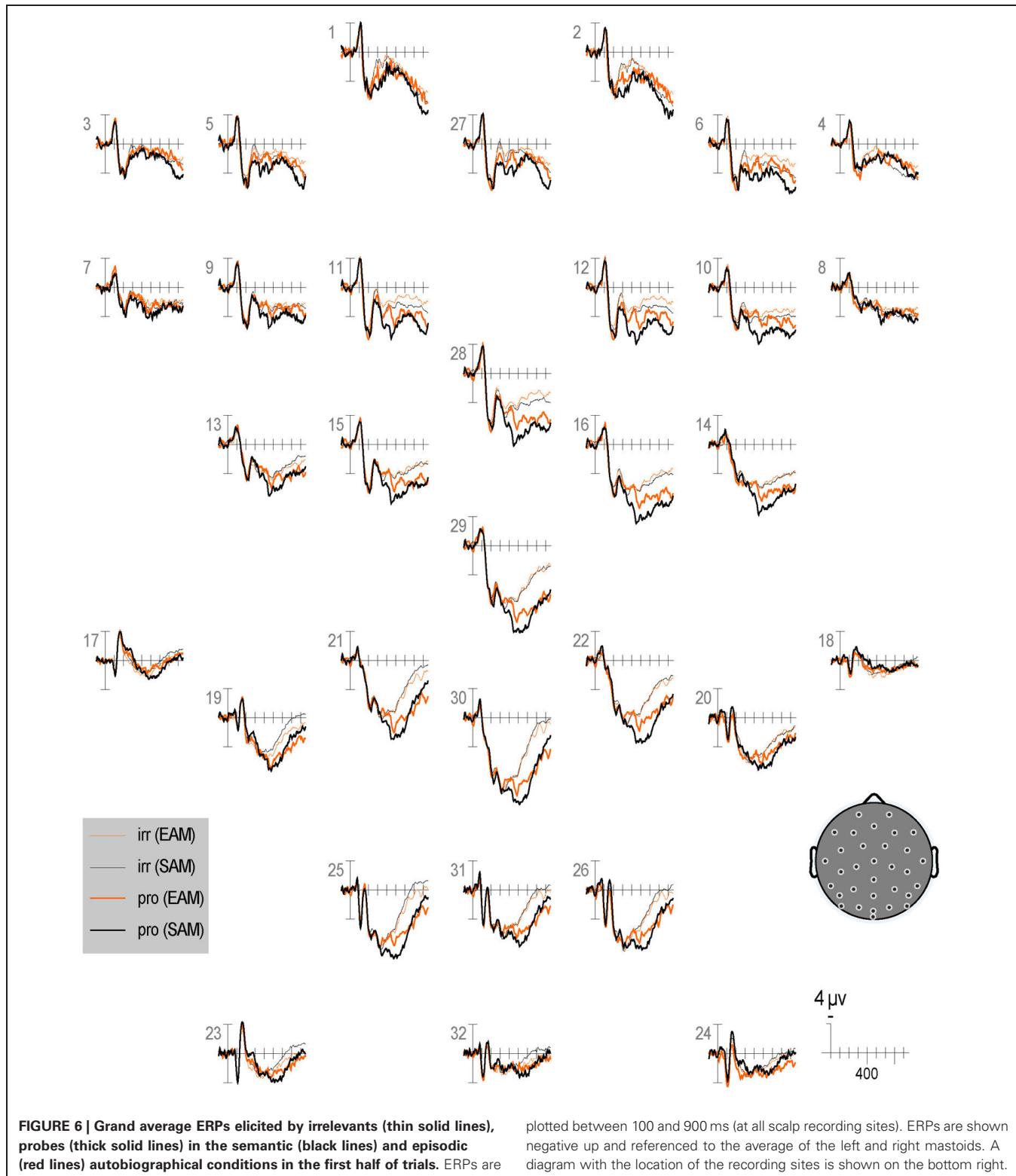
Performance and ERPs

RTs, accuracy, and ERP effects differed from each other so it is unclear which ERP effects drive the behavioral effects. Nonetheless, a few points can be made. The LPC starts too late (after 700 ms) to drive RT effects (all faster than 700 ms, on average) and corresponding accuracy of these responses. The P3b (400–600 ms) overlaps the earliest RTs, which are to irrelevants (500–600 ms), and so is also probably too late to influence RTs to

irrelevants and even too late to influence RTs to probes much if at all (600–700 ms). The N2 and N400 are thus the ERP markers that are most likely to be responsible for the RTs and corresponding accuracy. Consistent with this, the N2 has long been recognized as having a time course and relationship with RTs consistent with the underlying decision processes driving the RTs during discrimination tasks, such as the CIT, in part because the N2 is early enough to drive the RTs, whereas the P3 is often too late, and N2 latency is related to RTs (Ritter et al., 1979). Likewise, the N3 complex, which includes the (mis)match N2, is related to RTs during category decisions (Philiastides and Sajda, 2006, 2007; Philiastides et al., 2006). Given that both the N2 and N400 show a CIT effect in the semantic condition, the RT CIT effect in this condition could reflect both knowledge and semantic memory processes underlying these ERPs. Given that the N2 but not the N400 shows a CIT effect in both the semantic and episodic conditions, the knowledge processes underlying the N2 but not the N400 drives the CIT effect in the episodic condition. However, it is likely that at least the initial CIT effect on the P3b, which starts within 400 ms in the semantic condition and within 475 ms in the episodic condition, could further influence the RTs. However, the finding that P3b CIT effects end around 650 ms, which is after the response to all items except the episodic probe in the first half of trials suggests that the processes underlying the P3b are unlikely to be the only factor influencing behavior. This highlights the importance of considering the (mis)match frontal N2 and N400 and underlying knowledge and semantic memory processes, respectively, in future CIT studies and for deception detection, in general.

Saliency and related factors

Memories can differ in saliency, but how they differ depends upon many factors. Manipulating these factors was beyond the scope of this initial experiment but will be important for future research. We highlight here a few key issues regarding saliency and memory. First, saliency needs to be defined clearly but as yet a good definition is lacking, in general and in the memory field. Saliency has been most clearly defined in the context of selective attention to perceptual information. In particular, saliency is defined operationally by search performance: Items that differ along certain perceptual dimensions from the surrounding context are more salient and can be detected faster than items that do not differ as much from the surrounding context (e.g., a red dot against a background of green distractor dots). Saliency so defined orients selective attention, which can occur in parallel in early visual areas (Treisman, 2006) and, after the initial separation of stimulus information into features, binds these features together into an object representation and searches a scene serially (Wolfe et al., 1989; Treisman, 2006). Depending upon the context and task goals, these computations can inform the selective attention system to attend to salient features (e.g., the red dot) and filter out distractors and less salient features (Kastner and Pinsk, 2004). Selective attention can also be driven endogenously (e.g., by task goals and memory), and the top-down feedback inputs that perform these functions proceed from higher to lower order areas of information processing (Buffalo et al., 2010).



Second, in CIT paradigms usually the various items do not differ perceptually and so “salience” is driven entirely by stored memory and to its interaction with the details of the CIT paradigm. For example, probes and irrelevants are perceptually

identical between conditions (i.e., all strings of two numbers and three letters) and have the same motor demands and so perceptual differences cannot drive saliency here: Saliency is determined primarily by memory. Notably, this dictates that,

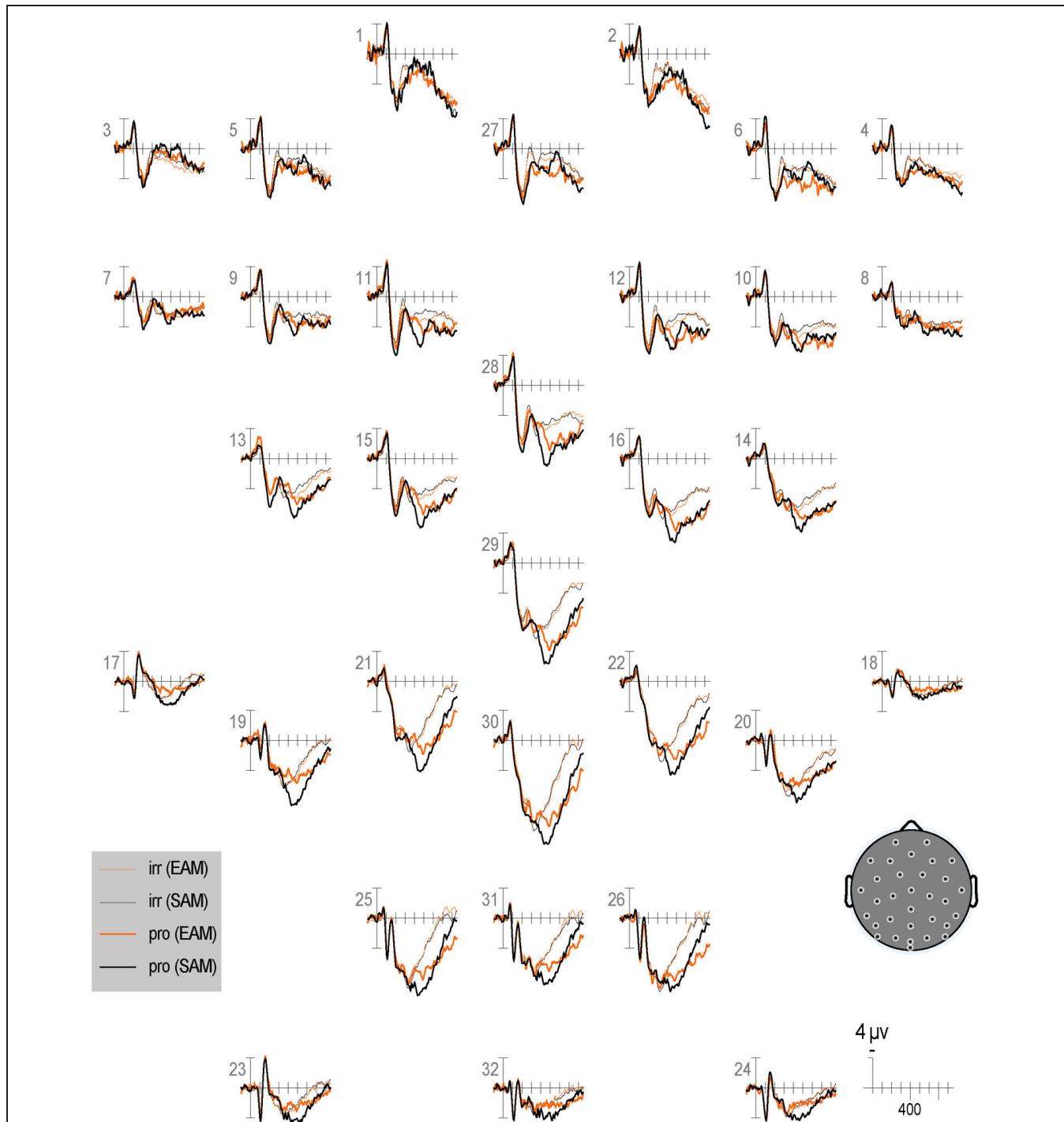


FIGURE 7 | Grand average ERPs elicited by irrelevants (thin solid lines), probes (thick solid lines) in the semantic (black lines) and episodic (red lines) autobiographical conditions in the second half of trials. ERPs are

plotted between 100 and 900 ms (at all scalp recording sites). ERPs are shown negative up and referenced to the average of the left and right mastoids. A diagram with the location of the recording sites is shown on the bottom right.

because memory is stored where it is processed in the cortex (Slotnick and Schacter, 2004; Schendan and Maher, 2009), saliency effects might be observed at the same time as memory differences are computed and/or afterwards when an earlier memory computation influences later cognitive and other

memory processes (Moses et al., 2005): Saliency effects can only be observed once the first memory effect has begun. Thus, it is necessary to consider how saliency has been defined in some of the few memory studies that have tried to address its role.

Third, in the memory field, definitions of saliency are based on memory representations, not perception, and differ between memory types: the information encoded in memory and its interactions with the task determine the salience of memory of a particular type (e.g., semantic or episodic), and such memory saliency computations can potentially influence another memory (of the same or a different type) activated simultaneously or later on in stimulus processing (e.g., semantic memory could influence episodic memory). For example, saliency of semantic memory has been defined based on (1) conceptual or perceptual distinctiveness in terms of dominance of meaning (e.g., for "bank," the dominant meaning is associated with "money," not "river") (Rajaram, 1998) or learned statistical regularities of the stimuli (e.g., orthographic frequency) (Rajaram, 1998), respectively, or (2) the representation strength of semantic features (e.g., high visual vs. high motor) (Kellenbach et al., 2002; Koriat and Pearlman-Avnion, 2003). (3) For episodic memory, saliency (or significance) of autobiographical information is based on the personal relevance of the learning episode and is closely related to emotional salience (Westmacott and Moscovitch, 2003), and this helps to preserve episodic memory (Levin et al., 1985) and semantic memory despite brain injury (Westmacott and Moscovitch, 2003). Note, by all these definitions, memory saliency is intrinsically entangled with the memory itself. On this basis, the birth date is higher in the saliency of both semantic and episodic memory than the secret date, predicting larger CIT effects across the entire ERP waveform. However, this was not the case because the LPC, consistent with this ERP as an index of episodic recollection, shows a larger CIT effect for the secret date. This raises the possibility that, for episodic memory, an additional definition of saliency is based on the role of recency (Soderlund et al., 2012). The present findings suggest that episodic memory can be more salient when recent than remote, as the secret date was associated with more recent episodic memory than the birth date. Finally, these considerations and multiple memory systems theory (Schacter and Tulving, 1994), more generally, highlight that saliency can only be defined within a particular type of memory; otherwise, one would be comparing apples and oranges. Saliency for semantic memory is not the same as saliency for episodic memory (e.g., meaning dominance determines salience for semantic memory vs. personal relevance determines salience for episodic memory). Thus, it would be difficult, if not theoretically impossible, to compare directly the saliency of items such as the birth and secret dates. For instance, on the one hand, it is uncommon to ask people explicitly to lie, and so the secret date has a very distinctive and salient episodic memory, and the recency of this memory may also enhance its saliency. On the other hand, the birth date usually has more personally relevant associated episodic memories than the secret date, and so it is very salient as well. In short, memory salience is greater for the birth date based on some definitions and memory types, whereas it is greater for the secret date based on others. Finally, we note that, by definition, semantic memory represents meaning, whereas episodic memory can store information regardless of meaning, as when people recognize non-sense visual patterns (Voss et al., 2011). Accordingly, the birth date is highly meaningful due to activating semantic memory, whereas

the secret date is less meaningful due to activating semantic memory less successfully but instead activates episodic memory, due to its recency, more successfully than the birth date. Thus, as we argued, the birth and secret dates differ as a function of their ability to activate semantic or episodic memory, and differ in meaningfulness (as one definition of saliency) only as a function of the extent to which they activate semantic memory.

Fourth, saliency is not a property of an item alone but rather a property of the item in a particular context. For example, the frequent word "table" may be low in saliency when embedded in a list of other frequent words but highly salient in the context of famous names. Furthermore, memory saliency depends on the task at hand and can be modulated by attentional manipulations (Rajaram, 1998), and so the word "table" can become highly salient in the context of other common words when the task requires detecting furniture words that appear infrequently. In the present CIT paradigm, the birth and secret dates occurred (in different blocks) infrequently within a stream of random dates, and they were the only items for which a lie had to be produced, making them highly salient in both conditions. The behavioral results support this and provide an operational definition of saliency for this task, as done in the attention field: faster RTs to probes are taken to reflect higher saliency in the task. Specifically, although RTs differ reliably between probes and irrelevants (documenting that the study had sufficient power to detect such differences), the memory conditions show no evidence of any difference that could be attributed to saliency. Thus, the CIT paradigm and procedures made both types of probes highly salient (being the only items for which a lie had to be produced) so that any residual saliency differences are very small; at most, probes show a non-significant trend to be faster in the semantic than episodic condition in the first block (596 vs. 640 ms). This could be due to the specific content of the memory (birthdate) or to the fact that the semantic probes were the only ones with a strong semantic content in the stream of irrelevants and targets (items with predominantly episodic memory associated with them and minimal semantic memory).

Finally, other reasons why we believe differential saliency was not an issue in the present study include the following. (1) If the birth date is more salient than the secret date, then a saliency-based account of the frontal N2 would predict a larger N2 to probes than irrelevants and a larger N2 for the birth than secret date. This is because, by definition, saliency engages attentional and other cognitive control processes (e.g., response monitoring, depending on the stimulus-response mapping) and these processes are typically associated with a larger N2 (Folstein and Van Petten, 2008). So, counterfactually, a smaller N2 for the date of birth than the secret date implies that date of birth was not more salient than the secret date. The same logic applies to the comparison between probes and irrelevants. Thus, the evidence is at odds with a saliency account and more in tune with a memory matching explanation. (2) Further, the LPC effect should be larger for the birth than the secret date, but the opposite was found, consistent with the idea that memory differences primarily drive the effects. (3) To our knowledge, there has been

no systematic P300 CIT work suggesting that a recent memory leads to smaller P300s than one's birth date, unless such information is acquired incidentally (Rosenfeld et al., 2006), which was not the case in our study, and no CIT studies have been conducted in which stimulus saliency was non-circularly defined and its systematic manipulation affected P300 amplitude (or any other ERP) in a way that could easily explain the current findings.

Memory task orientation

Memory is task-dependent; task instructions can influence the importance of a particular type of memory for performance and alter the pattern of effects (Richardson-Klavehn and Bjork, 1988). Thus differential activation of semantic and episodic memory by the birth and secret dates may also change the memory orientation of the CIT task to favor semantic vs. episodic memory, respectively. Consider that a seemingly subtle difference in task instructions from categorization (e.g., categorize the object) to recognition (e.g., recognize the item as old or new) alters the cortical networks involved in processing the same item (Schendan and Stern, 2008), consistent with multiple memory system theory (Squire and Zola-Morgan, 1991; Schacter and Tulving, 1994). In the SAM condition, the task requirement to lie about the birthdate likely focuses attention on associated rich semantic memory and remote episodic memories because these memories distinguish the birth date from the target date and irrelevants. Likewise, in the EAM condition, the task requirement to lie about the secret date should instead focus attention on the one recent episodic memory to cue the task response, while also minimizing attention to any associated semantic memory or knowledge, because recent episodic memory most clearly distinguishes the secret date from the target date and irrelevants. Moreover, the EAM condition probably focuses attention on episodic memory more than the SAM condition because strategic retrieval processes in inferior prefrontal cortex are required more for remote than recent episodic recollection (Soderlund et al., 2012), and lying recruits such prefrontal processes to inhibit the prepotent truthful response. Altogether, this would make these neural resources less available to recollect episodic autobiographical memories, which is more of a problem for the birth date in the SAM condition where in these memories are more remote, than the EAM condition where in the memory is recent. Likewise, such prefrontal processes are also implicated in selecting episodic memory from competing alternatives (Badre and Wagner, 2007), and the many remote episodic memories associated with the birthdate would require such selection processes more than the single recent episodic memory associated with the secret date. Consequently, the SAM condition orients the task predominantly toward using semantic memory as the primary cue to guide performance because semantic memory is retrieved more readily than is remote episodic memory. In contrast, the EAM condition orients the task predominantly toward recent episodic memory as the primary cue to guide performance. This predicts that the SAM condition should produce larger CIT effects on ERP components related to knowledge and semantic memory (i.e., N2, N400, and P3b), whereas the EAM

condition should produce larger CIT effects on ERP components related to episodic memory (i.e., LPC), consistent with these findings. Such task orientation effects would then interact with the memory retrieved to determine task performance (Schyns, 1998).

CONCLUSIONS

This study demonstrates that memory associated with a probe has multiple effects on the ERPs in CIT paradigms (N2, N400, P3b, and LPC), and the exact pattern of each effect depends upon the type of memory. Here, these effects are broadly consistent with the known properties of semantic and episodic memory systems, as assessed using ERPs. Documenting and examining these ERP effects is necessary to understand fully the neural basis of the processes engaged during CIT and related paradigms. For practical applications, analysis methods may be fine-tuned to detect concealed information depending on the associated memory type: For semantic probes, the focus should be the frontal N2, centroparietal N400, and parietal P3b, and for episodic probes the focus should be the P3b, and the LPC. Further, using longer sessions with more trials can efficiently improve signal-to-noise ratio because ERP CIT effects exhibited no signs of habituation or fatigue.

Altogether, the findings indicate that the frontal N2, centroparietal N400, and P3b are especially sensitive to information stored in knowledge and semantic memory systems of the neocortex, whereas the LPC is especially sensitive to information stored in the episodic memory system that depends on the hippocampus and adjacent cortical structures of the mediotemporal lobe. This conclusion highlights that clearly defining, manipulating, and considering the type of memory that may be concealed may be important for accurate detection of concealed information. Future CIT studies will need to consider carefully the semantic and episodic memory associated with each item, as well as how this interacts with task and experimental context. For example, if the goal is to reveal concealed episodic memory, then the semantic memory associated with each episodic item may need to be equated or specifically manipulated to separate out the semantic from episodic contributions. The present memory manipulation essentially orients subjects to focus on one memory system over another because the semantic and episodic probes that determined the memory system (i.e., semantic vs. episodic) for retrieval were presented in separate blocks of trials; participants did not have to discriminate between semantic and episodic probes directly. The strength of this approach is that it parallels the methods of memory research. Semantic memory experiments would involve asking subjects to report if the item exists in the real world, the meaning of a stimulus, or to categorize or name it, and items would differ in how well they are known (i.e., how well semantic memory is activated), akin to the semantic probe and irrelevants in this CIT paradigm. Episodic memory experiments would involve asking subjects to report whether the item is familiar from a prior study experience and/or to recollect associated information from that study experience, and familiar items would be mixed with unfamiliar items that had not been studied, akin

to the episodic probe and irrelevants in this CIT paradigm. The present experiment was a first attempt at teasing apart semantic and episodic memory contributions to the CIT. However, here, as in the real world, items will usually activate both semantic and episodic memory to some extent. This mix needs to be carefully documented in CIT paradigms and, more broadly, in deception research.

REFERENCES

- Allen, J. J., Iacono, W. G., and Danielson, K. D. (1992). The identification of concealed memories using the event-related potential and implicit behavioral measures: a methodology for prediction in the face of individual differences. *Psychophysiology* 29, 504–522.
- Badre, D., and Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia* 45, 2883–2901.
- Barrett, S. E., and Rugg, M. D. (1989). Event-related potentials and the semantic matching of faces. *Neuropsychologia* 27, 913–922.
- Barrett, S. E., and Rugg, M. D. (1990). Event-related potentials and the semantic matching of pictures. *Brain Cogn.* 14, 201–212.
- Bayley, P. J., O'Reilly, R. C., Curran, T., and Squire, L. R. (2008). New semantic learning in patients with large medial temporal lobe lesions. *Hippocampus* 18, 575–583.
- Ben-Shakhar, G., and Elaad, E. (2002). Effects of questions' repetition and variation on the efficiency of the guilty knowledge test: a reexamination. *J. Appl. Psychol.* 87, 972–977.
- Ben-Shakhar, G., and Elaad, E. (2003). The validity of psychophysiological detection of information with the Guilty Knowledge Test: a meta-analytic review. *J. Appl. Psychol.* 88, 131–151.
- Ben-Shakhar, G., Lieblich, I., and Kugelmass, S. (1975). Detection of information and GSR habituation: an attempt to derive detection efficiency from two habituation curves. *Psychophysiology* 12, 283–288.
- Buffalo, E. A., Fries, P., Landman, R., Liang, H., and Desimone, R. (2010). A backward progression of attentional effects in the ventral stream. *Proc. Natl. Acad. Sci. U.S.A.* 107, 361–365.
- Christ, S. E., Van Essen, D. C., Watson, J. M., Brubaker, L. E., and McDermott, K. B. (2009). The contributions of prefrontal cortex and executive control to deception: evidence from activation likelihood estimate meta-analyses. *Cereb. Cortex* 19, 1557–1566.
- Craik, F. I., and Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *J. Exp. Psychol. Gen.* 104, 268–294.
- Curran, T., Tanaka, J. W., and Weiskopf, D. M. (2002). An electrophysiological comparison of visual categorization and recognition memory. *Cogn. Affect. Behav. Neurosci.* 2, 1–18.
- Dien, J., Spencer, K. M., and Donchin, E. (2004). Parsing the late positive complex: mental chronometry and the ERP components that inhabit the neighborhood of the P300. *Psychophysiology* 41, 665–678.
- Donchin, E., and Coles, M. G. (1988). Is the P300 component a manifestation of context updating? *Behav. Brain Sci.* 11, 355–425.
- Eichenbaum, H., Yonelinas, A. P., and Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annu. Rev. Neurosci.* 30, 123–152.
- Farwell, L. A., and Donchin, E. (1991). The truth will out: interrogative polygraphy ("lie detection") with event-related brain potentials. *Psychophysiology* 28, 531–547.
- Folstein, J. R., and Van Petten, C. (2004). Multidimensional rule, unidimensional rule, and similarity strategies in categorization: event-related brain potential correlates. *J. Exp. Psychol. Learn. Mem. Cogn.* 30, 1026–1044.
- Folstein, J. R., and Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: a review. *Psychophysiology* 45, 152–170.
- Folstein, J. R., Van Petten, C., and Rose, S. A. (2008). Novelty and conflict in the categorization of complex stimuli. *Psychophysiology* 45, 467–479.
- Gamer, M., Bauermann, T., Stoeter, P., and Vossel, G. (2007). Covariations among fMRI, skin conductance, and behavioral data during processing of concealed information. *Hum. Brain Mapp.* 28, 1287–1301.
- Gamer, M., and Berti, S. (2010). Task relevance and recognition of concealed information have different influences on electrodermal activity and event-related brain potentials. *Psychophysiology* 47, 355–364.
- Gamer, M., and Berti, S. (2012). P300 amplitudes in the concealed information test are less affected by depth of processing than electrodermal responses. *Front. Hum. Neurosci.* 6:308. doi: 10.3389/fnhum.2012.00308
- Ganis, G., and Kutas, M. (2003). An electrophysiological study of scene effects on object identification. *Brain Res. Cogn. Brain Res.* 16, 123–144.
- Ganis, G., Rosenfeld, J. P., Meixner, J., Kievit, R. A., and Schendan, H. E. (2011). Lying in the scanner: covert countermeasures disrupt deception detection by functional magnetic resonance imaging. *Neuroimage* 55, 312–319.
- Gati, I., and Ben-Shakhar, G. (1990). Novelty and significance in orientation and habituation: a feature-matching approach. *J. Exp. Psychol. Gen.* 119, 251–263.
- Glisky, E. L., and Schacter, D. L. (1987). Acquisition of domain-specific knowledge in organic amnesia: training for computer-related work. *Neuropsychologia* 25, 893–906.
- Gratton, C., Evans, K. M., and Federmeier, K. D. (2009). See what I mean? An ERP study of the effect of background knowledge on novel object processing. *Mem. Cognit.* 37, 277–291.
- Harand, C., Bertran, F., La Joie, R., Landeau, B., Mezenge, F., Desgranges, B., et al. (2012). The hippocampus remains activated over the long term for the retrieval of truly episodic memories. *PLoS ONE* 7:e43495. doi: 10.1371/journal.pone.0043495
- Hassabis, D., and Maguire, E. A. (2009). The construction system of the brain. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 1263–1271.
- Hyde, T. S., and Jenkins, J. J. (1973). Recall of words as a function of semantic, graphic, and syntactic orienting tasks. *J. Verb. Learn. Verb. Behav.* 12, 471–480.
- Johnson, R. Jr. (1986). A triarchic model of P300 amplitude. *Psychophysiology* 23, 367–384.
- Johnson, R. Jr. (1993). On the neural generators of the P300 component of the event-related potential. *Psychophysiology* 30, 90–97.
- Kastner, S., and Pinsk, M. A. (2004). Visual attention as a multilevel selection process. *Cogn. Affect. Behav. Neurosci.* 4, 483–500.
- Kellenbach, M. L., Wijers, A. A., Hovius, M., Mulder, J., and Mulder, G. (2002). Neural differentiation of lexico-syntactic categories or semantic features? event-related potential evidence for both. *J. Cogn. Neurosci.* 14, 561–577.
- Kellogg, R. T., Cocklin, T., and Bourne, L. E. Jr. (1982). Conscious attentional demands of encoding and retrieval from long-term memory. *Am. J. Psychol.* 95, 183–198.
- Kesner, R. P., and Hunsaker, M. R. (2010). The temporal attributes of episodic memory. *Behav. Brain Res.* 215, 299–309.
- Koriat, A., and Pearlman-Avnion, S. (2003). Memory organization of action events and its relationship to memory performance. *J. Exp. Psychol. Gen.* 132, 435–454.
- Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647.
- Kutas, M., McCarthy, G., and Donchin, E. (1977). Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time. *Science* 197, 792–795.
- Langleben, D. D., Schroeder, L., Maldjian, J. A., Gur, R. C., McDonald, S., Ragland, J. D., et al. (2002). Brain Activity during simulated deception: an event-related functional magnetic resonance study. *Neuroimage* 15, 727–732.
- Levin, H. S., High, W. M., Meyers, C. A., Von Laufen, A., Hayden, M. E., and Eisenberg, H. M. (1985). Impairment of remote memory after closed head injury. *J. Neurol. Neurosurg. Psychiatry* 48, 556–563.
- Lykken, D. T. (1959). The GSR in the detection of guilt. *J. Appl. Psychol.* 43, 385–388.
- Maguire, E. A., Vargha-Khadem, F., and Hassabis, D. (2010). Imaging

ACKNOWLEDGMENTS

Giorgio Ganis and Haline Schendan contributed equally to this work. We would like to thank Leanne Kiff for assistance with data collection. This research was supported in part by a UoP International Research, Networking and Collaboration Award, and by a Marie Curie Career Integration Grant (CoND) to Giorgio Ganis.

- fictitious and future experiences: evidence from developmental amnesia. *Neuropsychologia* 48, 3187–3192.
- Mankin, E. A., Sparks, F. T., Slayyah, B., Sutherland, R. J., Leutgeb, S., and Leutgeb, J. K. (2012). Neuronal code for extended time in the hippocampus. *Proc. Natl. Acad. Sci. U.S.A.* 109, 19462–19467.
- Martinelli, P., Sperduti, M., and Piolino, P. (2012). Neural substrates of the self-memory system: new insights from a meta-analysis. *Hum. Brain Mapp.* doi: 10.1002/hbm.22008. [Epub ahead of print].
- Matsuda, I., Nittono, H., Hirota, A., Ogawa, T., and Takasawa, N. (2009). Event-related brain potentials during the standard autonomic-based concealed information test. *Int. J. Psychophysiol.* 74, 58–68.
- McPherson, W. B., and Holcomb, P. J. (1999). An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology* 36, 53–65.
- Morris, C. D., Bransford, J. D., and Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *J. Verb. Learn. Verb. Behav.* 16, 519–533.
- Moses, S. N., Cole, C., Driscoll, I., and Ryan, J. D. (2005). Differential contributions of hippocampus, amygdala and perirhinal cortex to recognition of novel objects, contextual stimuli and stimulus relationships. *Brain Res. Bull.* 67, 62–76.
- Muller, H. M., and Kutas, M. (1996). What's in a name? Electrophysiological differences between spoken nouns, proper names and one's own name. *Neuroreport* 8, 221–225.
- Nadel, L., Samsonovich, A., Ryan, L., and Moscovitch, M. (2000). Multiple trace theory of human memory: computational, neuroimaging, and neuropsychological results. *Hippocampus* 10, 352–368.
- Nose, I., Murai, J., and Taira, M. (2009). Disclosing concealed information on the basis of cortical activations. *Neuroimage* 44, 1380–1386.
- Paller, K. A., and Kutas, M. (1992). Brain potentials during memory retrieval provide neurophysiological support for the distinction between conscious recollection and priming. *J. Cogn. Neurosci.* 4, 375–391.
- Paller, K. A., Kutas, M., and McIsaac, H. K. (1995). Monitoring conscious recollection via the electrical activity of the brain. *Psychol. Sci.* 6, 107–111.
- Paller, K. A., Voss, J. L., and Boehm, S. G. (2007). Validating neural correlates of familiarity. *Trends Cogn. Sci.* 11, 243–250.
- Paller, K. A., and Wagner, A. D. (2002). Observing the transformation of experience into memory. *Trends Cogn. Sci.* 6, 93–102.
- Patterson, K., Nestor, P. J., and Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat. Rev.* 8, 976–987.
- Phan, K. L., Magalhaes, A., Ziemlewicz, T. J., Fitzgerald, D. A., Green, C., and Smith, W. (2005). Neural correlates of telling lies: a functional magnetic resonance imaging study at 4 Tesla. *Acad. Radiol.* 12, 164–172.
- Philastrides, M. G., Ratcliff, R., and Sajda, P. (2006). Neural representation of task difficulty and decision making during perceptual categorization: a timing diagram. *J. Neurosci.* 26, 8965–8975.
- Philastrides, M. G., and Sajda, P. (2006). Temporal characterization of the neural correlates of perceptual decision making in the human brain. *Cereb. Cortex* 16, 509–518.
- Philastrides, M. G., and Sajda, P. (2007). EEG-informed fMRI reveals spatiotemporal characteristics of perceptual decision making. *J. Neurosci.* 27, 13082–13091.
- Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* 118, 2128–2148.
- Rajaram, S. (1998). The effects of conceptual salience and perceptual distinctiveness on conscious recollection. *Psychon. Bull. Rev.* 5, 71–78.
- Renoult, L., and Debraille, J. B. (2011). N400-like potentials and reaction times index semantic relations between highly repeated individual words. *J. Cogn. Neurosci.* 23, 905–922.
- Richardson-Klavehn, A., and Bjork, R. A. (1988). Measures of memory. *Annu. Rev. Psychol.* 39, 475–543.
- Ritter, W., Simson, R., Vaughan, H. G. Jr., and Friedman, D. (1979). A brain event related to the making of a sensory discrimination. *Science* 203, 1358–1361.
- Rosenfeld, J. P., Angell, A., Johnson, M., and Qian, J. H. (1991). An ERP-based, control-question lie detector analog: algorithms for discriminating effects within individuals' average waveforms. *Psychophysiology* 28, 319–335.
- Rosenfeld, J. P., Biroschak, J. R., and Furey, J. J. (2006). P300-based detection of concealed autobiographical versus incidentally acquired information in target and non-target paradigms. *Int. J. Psychophysiol.* 60, 251–259.
- Rosenfeld, J. P., Cantwell, B., Nasman, V. T., Wojdag, V., Ivanov, S., and Mazzeri, L. (1988). A modified, event-related potential-based guilty knowledge test. *Int. J. Neurosci.* 42, 157–161.
- Rosenfeld, J. P., Labkovsky, E., Winograd, M., Lui, M. A., Vandenberg, C., and Chedid, E. (2008). The Complex Trial Protocol (CTP): a new, countermeasure-resistant, accurate, P300-based method for detection of concealed information. *Psychophysiology* 45, 906–919.
- Rosenfeld, J. P., Soskins, M., Bosh, G., and Ryan, A. (2004). Simple, effective countermeasures to P300-based tests of detection of concealed information. *Psychophysiology* 41, 205–219.
- Rugg, M. D., and Curran, T. (2007). Event-related potentials and recognition memory. *Trends Cogn. Sci.* 11, 251–257.
- Rugg, M. D., Mark, R. E., Walla, P., Schloerscheidt, A. M., Birch, C. S., and Allan, K. (1998). Dissociation of the neural correlates of implicit and explicit memory. *Nature* 392, 595–598.
- Schacter, D. L., and Tulving, E. (eds.). (1994). *Memory Systems* 1994. Cambridge, MA: MIT Press.
- Schank, R. C., and Abelson, R. (1977). *Scripts, Plans, Goals and Understanding*. Hillsdale, NJ: Lawrence Erlbaum and Associates.
- Schendan, H. E., and Kutas, M. (2002). Neurophysiological evidence for two processing times for visual object identification. *Neuropsychologia* 40, 931–945.
- Schendan, H. E., and Kutas, M. (2003). Time course of processes and representations supporting visual object identification and memory. *J. Cogn. Neurosci.* 15, 111–135.
- Schendan, H. E., and Maher, S. M. (2009). Object knowledge during entry-level categorization is activated and modified by implicit memory after 200 ms. *Neuroimage* 44, 1423–1438.
- Schendan, H. E., and Stern, C. E. (2008). Where vision meets memory: prefrontal-posterior networks for visual object constancy during categorization and recognition. *Cereb. Cortex* 18, 1695–1711.
- Schmolck, H., Kensinger, E. A., Corkin, S., and Squire, L. R. (2002). Semantic knowledge in patient H.M. and other patients with bilateral medial and lateral temporal lobe lesions. *Hippocampus* 12, 520–533.
- Schyns, P. G. (1998). Diagnostic recognition: task constraints, object information, and their interactions. *Cognition* 67, 147–179.
- Scott, L. S., Tanaka, J. W., Sheinberg, D. L., and Curran, T. (2008). The role of category learning in the acquisition and retention of perceptual expertise: a behavioral and neurophysiological study. *Brain Res.* 1210, 204–215.
- Slotnick, S. D., and Schacter, D. L. (2004). A sensory signature that distinguishes true from false memories. *Nat. Neurosci.* 7, 664–672.
- Soderlund, H., Moscovitch, M., Kumar, N., Mandic, M., and Levine, B. (2012). As time goes by: hippocampal connectivity changes with remoteness of autobiographical memory retrieval. *Hippocampus* 22, 670–679.
- Sokolov, E. N. (1963). Higher nervous functions; the orienting reflex. *Annu. Rev. Physiol.* 25, 545–580.
- Squire, L. R., and Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science* 253, 1380–1386.
- Treisman, A. M. (2006). How the deployment of attention determines what we see. *Vis. Cogn.* 14, 411–443.
- Tsivilis, D., Otten, L. J., and Rugg, M. D. (2001). Context effects on the neural correlates of recognition memory: an electrophysiological study. *Neuron* 31, 497–505.
- Tulving, E. (1972). "Episodic and semantic memory," in *Organization of Memory*, eds E. Tulving and W. Donaldson (New York, NY: Academic Press), 381–403.
- Tulving, E. (1993). "Self-knowledge of an amnesic individual is represented abstractly," in *The Mental Representation of Self and Autobiographical Knowledge Abata the Self: Advances in Social Cognition*, Vol. 5, eds T. K. Srull and R. S. Wyer (Hillsdale, NJ: Lawrence Erlbaum Associates), 147–156.
- Tulving, E. (2002). Episodic memory: from mind to brain. *Annu. Rev. Psychol.* 53, 1–25.
- Tulving, E., Schacter, D. L., McLachlan, D. R., and Moscovitch, M. (1988). Priming of semantic autobiographical knowledge: a case study of retrograde amnesia. *Brain Cogn.* 8, 3–20.
- Tulving, E., and Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychol. Rev.* 80, 359–380.
- Vargha-Khadem, F., Gadian, D. G., Watkins, K. E., Connelly, A., Van Paesschen, W., and Mishkin, M. (1997). Differential effects of early hippocampal pathology on episodic

- and semantic memory. *Science* 277, 376–380.
- Verfaellie, M., and Cermak, L. S. (1994). Acquisition of generic memory in amnesia. *Cortex* 30, 293–303.
- Verleger, R. (2008). P3b: towards some decision about memory. *Clin. Neurophysiol.* 119, 968–970.
- Voss, J. L., and Federmeier, K. D. (2011). FN400 potentials are functionally identical to N400 potentials and reflect semantic processing during recognition testing. *Psychophysiology* 48, 532–546.
- Voss, J. L., Federmeier, K. D., and Paller, K. A. (2011). The potato chip really does look like elvis! neural hallmarks of conceptual processing associated with finding novel shapes subjectively meaningful. *Cereb. Cortex* 22, 2354–2364.
- Voss, J. L., and Paller, K. A. (2006). Fluent conceptual processing and explicit memory for faces are electrophysiologically distinct. *J. Neurosci.* 26, 926–933.
- Voss, J. L., Schendan, H. E., and Paller, K. A. (2010). Finding meaning in novel geometric shapes influences electrophysiological correlates of repetition and dissociates perceptual and conceptual priming. *Neuroimage* 49, 2879–2889.
- Westmacott, R., and Moscovitch, M. (2003). The contribution of autobiographical significance to semantic memory. *Mem. Cognit.* 31, 761–774.
- Wolfe, J. M., Cave, K. R., and Franzel, S. L. (1989). Guided search: an alternative to the feature integration model for visual search. *J. Exp. Psychol. Hum. Percept. Perform.* 15, 419–433.

Received: 31 August 2012; accepted: 20 December 2012; published online: 24 January 2013.

Citation: Ganis G and Schendan HE (2013) Concealed semantic and episodic autobiographical memory electrified. *Front. Hum. Neurosci.* 6:354. doi: 10.3389/fnhum.2012.00354

Copyright © 2013 Ganis and Schendan. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.



P300 amplitudes in the concealed information test are less affected by depth of processing than electrodermal responses

Matthias Gamer^{1*} and Stefan Berti²

¹ Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

² Department of Psychology, Johannes Gutenberg-University Mainz, Mainz, Germany

Edited by:

Wolfgang Ambach, Institute for Frontier Areas of Psychology and Mental Health (IGPP), Germany

Reviewed by:

Izumi Matsuda, National Research Institute of Police Science, Japan
J. Peter Rosenfeld, Northwestern University, USA
Ali Baioui, Justus Liebig University Giessen, Germany

***Correspondence:**

Matthias Gamer, Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Martinistr. 52, Bldg W34, D-20246 Hamburg, Germany.
e-mail: m.gamer@uke.uni-hamburg.de

The Concealed Information Test (CIT) has been used in the laboratory as well as in field applications to detect concealed crime related memories. The presentation of crime relevant details to guilty suspects has been shown to elicit enhanced N200 and P300 amplitudes of the event-related brain potentials (ERPs) as well as greater skin conductance responses (SCRs) as compared to neutral test items. These electrophysiological and electrodermal responses were found to incrementally contribute to the validity of the test, thereby suggesting that these response systems are sensitive to different psychological processes. In the current study, we tested whether depth of processing differentially affects N200, P300, and SCR amplitudes in the CIT. Twenty participants carried out a mock crime and became familiar with central and peripheral crime details. A CIT that was conducted 1 week later revealed that SCR amplitudes were larger for central details although central and peripheral items were remembered equally well in a subsequent explicit memory test. By contrast, P300 amplitudes elicited by crime related details were larger but did not differ significantly between question types. N200 amplitudes did not allow for detecting concealed knowledge in this study. These results indicate that depth of processing might be one factor that differentially affects central and autonomic nervous system responses to concealed information. Such differentiation might be highly relevant for field applications of the CIT.

Keywords: concealed information test, memory, depth of processing, N200, P300, skin conductance, mock crime

INTRODUCTION

The valid differentiation of offenders and people who are innocent of a crime under investigation is one of the most important issues in forensic sciences. Correspondingly, there has been a huge interest in developing adequate techniques that allow for such detection (Vrij, 2008). One such line of research has focused on questioning techniques and different interrogation methods have been proposed that all have distinct advantages and drawbacks. A method that seems to be widely accepted in the scientific community is the so-called Guilty Knowledge or Concealed Information Test (GKT, CIT; Lykken, 1959; Ben-Shakhar et al., 2002). This method can be regarded as an indirect test for an involvement in a crime since the participant is not asked accusatory questions (e.g., “Did you rob the grocery store last night?”) but instead confronted with a series of questions presented in a multiple choice format that ask for specific details of the crime under investigation (e.g., the weapon that was used for the robbery, the amount of money that was stolen). Each question consists of the known critical detail (relevant or probe item) and a number of neutral alternatives (irrelevant items) that are equally plausible to an innocent person. It is assumed that the recognition of crime related details by a guilty examinee results in enhanced physiological responses as compared to the irrelevant items. Innocent examinees without

such knowledge should show a non-systematic response pattern. Indeed, this general response pattern has been found for a number of different behavioral and physiological measures (for a comprehensive overview see Verschuere et al., 2011).

Traditionally, CIT examinations were designed to measure autonomic responses, and the majority of studies relied on skin conductance as the only dependent measure (Ben-Shakhar and Elaad, 2003). Guilty examinees normally respond to the presentation of the crime related detail by showing enhanced skin conductance responses (SCRs). However, respiratory suppression and heart rate deceleration were also found to reliably differ between probes and irrelevant CIT items when the examinee is able to recognize crime related information (for a review see Gamer, 2011). With respect to measures of central nervous system activity, the majority of CIT studies focused on event-related brain potentials (ERPs). In most of these studies, a third item type (so-called targets) was introduced to the examination protocol (e.g., Farwell and Donchin, 1991; Allen et al., 1992; Rosenfeld et al., 2004; Meijer et al., 2007; Mertens and Allen, 2008). These target items require a different behavioral response than all other items and were used to maintain the subject’s attention during the testing procedure. ERP-studies on the CIT consistently reported larger P300 amplitudes following the presentation

of probes as compared to irrelevant items (for a review see Rosenfeld, 2011). Moreover, a few recent studies also reported larger N200 amplitudes when examinees were confronted with previously encoded information (Matsuda et al., 2009; Gamer and Berti, 2010). Taken together, autonomic responses as well as ERP measures allow for validly identifying concealed knowledge.

The majority of studies either focused on autonomic responses or ERPs but did not directly compare response patterns of these two types of measures. Only in recent years, a few studies tried to overcome this limitation and acquired autonomic and ERP measures within the same CIT setting (Matsuda et al., 2009, 2011; Ambach et al., 2010; Gamer and Berti, 2010). These studies mainly replicated what had already been established for separate measurements but also provided some evidence for an incremental validity of these measures. Thus, a combination of P300 amplitudes and autonomic responses enhanced CIT validity above each single measure (Ambach et al., 2010; Matsuda et al., 2011). A comparable result was obtained for the combination of N200 amplitudes and electrodermal measures (Gamer and Berti, 2010). What these studies did not address, however, is the underlying cause of this incremental validity.

At least three explanations seem plausible: First, the reliability of each single measure might be corrupted by an unknown amount of error. Assuming that this error is not substantially correlated between measures, a combination would always yield higher validity coefficients because of an increase in the signal-to-noise ratio. Second, physiological measures might be differentially responsive to concealed information or the pattern of responsiveness might vary between individuals (Matsuda et al., 2006). Such physiological effects might explain the previously observed enhanced validity for a combination of measures. Third, it seems possible that different physiological measures cover partially different psychological processes that are involved in the CIT (e.g., attentional orienting, memory retrieval, response selection, and monitoring). This would also explain the incremental validity of individual responses as well as the usually observed low to moderate correlations between measures (Matsuda et al., 2009; Gamer and Berti, 2010). Of course, these explanations are not mutually exclusive and incremental validity could rely on all these aspects. In the present study, however, we focus on the last explanation of incremental validity.

The current study was designed to shed further light on a psychological factor—namely depth of processing—that might have a differential influence on electrodermal and electrophysiological responses: It is a well established finding that items are better remembered when they have been processed more elaborately. In a seminal study on this issue, participants had to classify words according to specific criteria (e.g., typescript, rhyme, or meaning) that required a different depth of processing (Craik and Tulving, 1975). In a surprise memory test, participants were much better in remembering the initially presented words when they had been processed deeply during the encoding phase. From these data, it seems likely that memory for certain (mock) crime details also depends on depth of processing during crime execution. Since the CIT basically resembles a memory test, comparable differences should also be evident in the physiological data that is acquired during the CIT examination.

Two recent studies examined this hypothesis and tested whether autonomic responses during the CIT differ between central items that were encoded deeply during the mock crime as compared to more peripheral details which were only shallowly encoded (Nahari and Ben-Shakhar, 2011; Peth et al., 2012). Items were defined as central when they were directly related to the execution of the crime or actively handled during the course of the mock crime (e.g., the stolen item). By contrast, peripheral details were also present at the crime scene but unrelated to its execution (e.g., a picture hanging on the wall). Consistent with the hypotheses on the influence of depth of processing, participants remembered central details much better in these studies and also showed stronger autonomic responses to central items as compared to more peripheral details (c.f., Carmel et al., 2003). However, it is difficult to infer whether the latter effect is only an epiphenomenon of the reduced recognition rate or represents a specific effect of depth of processing on the autonomic level. Evidence for the latter interpretation comes from a recent study by Gamer et al. (2012). In this study, a dissociation was observed between autonomic responses and an explicit memory test: In detail, the results showed larger SCRs to deeply encoded details while no differences were obtained in an explicit memory test. Comparable results were also reported by Ambach et al. (2011) who found enhanced autonomic responses to stolen items as compared to details that were only seen during the mock crime. This result emerged even when excluding items that were not explicitly remembered but it only occurred when emphasizing the act of stealing in the CIT questions. Taken together, it seems that electrodermal responses might be sensitive to depth of processing. A study by Ferlazzo et al. (1993) suggests that this also applies to the P300 amplitude: In this study, P300 amplitudes in the test phase were enhanced for items which were deeply processed in the study phase compared to shallowly processed items. In the context of the CIT the effect of depth of processing on the P300 amplitude was not directly tested. However, previous studies consistently reported differences in P300 amplitudes between recognized items and irrelevant details for different domains including autobiographical or other personally relevant knowledge (e.g., Rosenfeld et al., 1995), mock crime details (e.g., Farwell and Donchin, 1991; Abootalebi et al., 2006), or explicitly learnt items (e.g., van Hooff et al., 1996). Moreover, it has been reported that P300 amplitudes were larger for highly salient autobiographical information (such as one's own name) as compared to incidentally acquired knowledge (such as the experimenter's name) (Rosenfeld et al., 2006). Similar results were obtained when comparing autobiographical information to explicitly learnt items (Ellwanger et al., 1996) or incidentally acquired mock crime details (Rosenfeld et al., 2007). Assuming some similarity between autobiographical information and deeply processed knowledge, it can thus be speculated that depth of processing also affects P300 amplitudes in a CIT protocol. However, the autobiographical information that was used in previous CIT studies (e.g., the own name, birthday, home town or school) was highly salient and could therefore be identified as personally relevant without difficulties. Such knowledge might qualitatively differ from other information that is only personally significant because it is linked to a specific episode in

life. To further examine how the strength of episodic memories relates to physiological responses in the CIT, the current study focused on information that was encoded in the same situation (a mock crime) but with varying depth of processing (central vs. peripheral details).

Taken together, the current study was designed to test whether depth of processing differentially affects electrodermal and ERP responses in a CIT. A mock crime procedure was constructed that involved the incidental encoding of two central and two peripheral items which were all relatively salient in order to heighten the probability of successful item recognition during the CIT. To overcome limitations of previous studies using a compromise between the requirements for a reliable quantification of ERPs (i.e., large number of stimulus repetitions) and SCRs (i.e., long interstimulus-intervals, ISIs), we implemented a new approach that has been originally developed for event-related functional magnetic resonance imaging (fMRI) studies. In this kind of studies it becomes highly problematic or even impossible to detect experimentally induced changes in slow fluctuations of the blood oxygen level dependent signal when using very short fixed ISIs. Dale (1999) demonstrated that this problem can be solved by properly jittering the ISIs and optimizing the sequence of events to ensure adequate randomization. Such an approach substantially enhances the statistical efficiency of rapid event-related fMRI designs. Because the problem experienced by such fMRI studies is very similar to the problem of reliably quantifying stimulus-related SCRs with short ISIs, we transferred the approach developed by Dale (1999) to the current study in order to simultaneously measure SCRs and ERPs within the same experiment.

MATERIALS AND METHODS

PARTICIPANTS

Twenty-two subjects volunteered to participate in the study and gave written informed consent according to the Declaration of Helsinki. Two participants were excluded. In one case, the electrodermal data was lost due to technical problems. For another subject, the EEG-data could not be analyzed due to excessive blinking. The final sample consisted of 20 persons (19 right-handed) of whom 13 were female and 7 male. Most participants were students and their mean age was 24.5 years ($SD = 6.3$ years). All subjects were paid 10,-EUR for their participation.

PROCEDURE

All participants were instructed to accomplish a realistic mock crime scenario where all critical details (printed in *italics* below) were only incidentally encoded during the mock crime itself and not mentioned in the instructions. Participants were given a room number and were asked to find this specific room in the psychology department. Upon entering the room (an *office*), they were instructed to search for a key to unlock a desk drawer. A *keyring pendant* was affixed to this key that was always placed on the desk and could thus be easily found. Participants were instructed to “steal” a data storage medium (a *CD*) from the desk drawer. They were told to have a look at the content of the CD at home and send a short message to an email address that was given to them in advance. In this email, participants should briefly describe what

they found on the CD and attach all files. There were six *pictures* on the CD that depicted a woman entering a store and leaving it with an envelope in her hands. The photos were taken from a perspective that looked like an observation of the respective woman. All participants correctly accomplished the mock crime and gave a valid description of the CD content in their email.

Following previous studies (Gamer et al., 2010; Nahari and Ben-Shakhar, 2011), we defined half of the relevant crime details as central and peripheral, respectively. Items that must have been perceived in order to successfully accomplish the mock crime were designated as central details (the CD and the content of the CD). The other details (the office where the mock crime took place and the keyring pendant) were not directly relevant for the mock crime itself and might or might not have been encoded during the course of the mock theft. Therefore, they were defined as peripheral. Thus, the question type (central vs. peripheral) was varied as a within-subjects factor.

Participants were instructed to return to the laboratory one week after execution of the mock crime to take a polygraph test. Before the examination, they were informed that a theft occurred in the psychology department a week ago. Because they were seen in the building at that time, they were told to be suspects of this theft but they would have an opportunity to demonstrate their innocence in a polygraph test examination. To increase motivation, all participants were encouraged to convince the polygraph examiner of their innocence and it was announced that all participants who successfully passed the CIT would have a chance to win 100,- EUR in a subsequent lottery.

For the polygraph examination, the participant was seated in a semi-reclining chair approximately 130 cm in front of a 19' color screen in an electrically shielded and sound attenuated chamber. The stimuli were presented as pictures with a size of 16.0×12.1 degrees of visual angle for duration of 1000 ms each (all stimuli are shown in **Figure 1**). An IBM-PC using the ERTS stimulation software (BeriSoft, Germany) controlled stimulus presentation. The examination was conducted according to the 3-item CIT protocol that includes probes (the crime related details), irrelevant items (equally plausible neutral details) and targets (Farwell and Donchin, 1991; Rosenfeld et al., 2004; Meijer et al., 2007). Target items are similar to irrelevant details with the exception that they require a different behavioral response to ensure that participants are paying attention to the stimulus presentation. Therefore, participants had to memorize four specific target items (i.e., one for each CIT question defined by the respective probe item) before the examination started. Participants were instructed to press the right key of a response pad with the middle finger of their right hand whenever a target item would be presented on a display screen. The left key had to be pressed with the right index finger following all other stimuli. That is, probe and irrelevant items shared the same response key. It was emphasized that key presses should be as accurate and fast as possible.

The two central and two peripheral probe items were presented in separate blocks whose order was counterbalanced across participants. Within each block, four irrelevant items and one target were additionally presented for each probe item and each item was shown in 15 trials. The blocks were divided into three sessions each with short breaks in between to prevent tiredness.

	Central details		Peripheral details	
Relevant items				
Target items				
Irrelevant items				

FIGURE 1 | List of all stimuli that were presented during the Concealed Information Test (CIT) in the current study. Note that all faces were blurred for this illustration to prevent identification of the depicted persons.

Each session started with the presentation of an additional irrelevant item that was used as a buffer and discarded in the analyses. Furthermore, the amount of trials of each item category (irrelevant, probe, target) was constant across sessions. Altogether, 30 probe (two relevant details \times 15 repetitions), 30 target (one target for each probe item \times 15 repetitions) and 120 irrelevant items (four irrelevant items for each probe item \times 15 repetitions) were presented for central and peripheral CIT questions, respectively.

Stimulus sequence and timing were optimized using the software optseq2 (<http://surfer.nmr.mgh.harvard.edu/optseq/>). This software was originally developed to optimize the statistical efficiency of event-related fMRI studies (Dale, 1999). However, since the hemodynamic response that is measured by fMRI has a similar response morphology and comparable timing characteristics, it seems plausible to transfer these calculations to the measurement of SCR in fast event-related paradigms. We used optseq2 to generate individual stimulus sequences for each experimental session. This software first randomizes the sequence of events (without restrictions) and then introduces random jitter between events such that no ISI is shorter than the previously defined minimum ISI (2400 ms in this study) and the whole sequence duration does not exceed the previously defined maximum length (5 min in this study). This process is iteratively repeated and the design efficiency is stored for each sequence. We used 10,000 iterations and saved the best (in terms of design efficiency) 120 sequences (six sessions for each of 20 subjects) for later use. The optimization procedure as implemented in optseq2 does not make strong predictions about the shape of event-related responses since it does not use a template of the hemodynamic

response function but instead a finite impulse response function spanning a predefined time window (0–8 s in this study). Thus, it is only assumed that the response morphology of stimulus-related responses is adequately described by the signal change within 8 s after stimulus presentation and stable across trials of one condition. Both these assumptions seem reasonable for stimulus-related SCRs that have a typical latency of 1–3 s, a rise time of 1–3 s (Dawson et al., 2007) and a response shape that is relatively stable within one examinee (Lim et al., 1997). With the currently chosen timing parameters, mean ISI amounted to 4952 ms (SD = 30 ms) across examinees. The average maximum ISI was 16230 ms (SD = 2323 ms) across examinees.

After completing the test, all participants were required to recall the probe items in a post-experimental memory test by means of a multiple-choice procedure. This test consisted of all items that were used in the CIT examination and participants were asked to identify the critical detail within each question. Finally, all participants were paid and fully debriefed about the nature of the study and the mock crime.

DATA ACQUISITION AND ANALYSIS

The electroencephalogram (EEG) was recorded continuously with a SynAmps amplifier (NeuroScan, Sterling, VA) from 19 cap-mounted Ag/AgCl electrodes (EasyCap, Germany) with positions according to the international 10–20 system; the reference electrode was placed at the right mastoid. To control for eye movements, vertical and horizontal electro-oculograms (EOG) were recorded. Data were digitized at 250 Hz and online filtered using a 0.05–40 Hz bandpass and a 50-Hz notch. Trials with eye blinks or eye movements (i.e., whenever the standard deviation within a 200-ms interval exceeded 30 μ V in the horizontal or vertical EOG) as well as erroneous trials were excluded from further analyses. For central details, the number of valid trials amounted to 112.0 irrelevant (SD = 7.8), 28.2 probe (SD = 2.0) and 26.0 (SD = 3.1) target trials. The respective trial numbers for peripheral details were 112.8 irrelevant (SD = 8.7), 28.3 probe (SD = 1.8) and 25.4 (SD = 3.3) target trials. The amount of valid trials did not differ significantly between central and peripheral CIT questions (all $p > 0.25$ in paired t -tests contrasting the number of irrelevant, probe and target trials, respectively). The ERPs were separately computed for the three item types of central and peripheral CIT questions, with a time window ranging from –200 to 1400 ms relative to the visual stimulus onset. The 200-ms pre-stimulus interval served as baseline.

Similar to our previous work (Gamer and Berti, 2010), we determined the N200 amplitude by computing the maximally negative segment average of 50 ms at Cz within a time window ranging from 200 to 350 ms after stimulus onset. P300 amplitudes were calculated using the peak-to-peak method as described by Rosenfeld (Rosenfeld et al., 1991; Soskins et al., 2001) and used by a number of previous studies (e.g., Rosenfeld et al., 2004; Meijer et al., 2007; Verschuere et al., 2009). In a first step, the maximal positive 100 ms segment average was determined in a time-window ranging from 300 to 800 ms. Subsequently, the maximal negative 100 ms segment between the latency of the positive peak and 1400 ms was determined. Peak-to-peak P300 amplitude was defined as the difference between these two

segments. This method was shown to be superior for the detection of concealed knowledge than the traditional base-to-peak measure (Soskins et al., 2001). As the P300 is most pronounced at Pz, amplitude calculations were limited to this site.

Skin conductance was measured by a constant voltage system (0.5 V) using a bipolar recording with two Ag/AgCl electrodes (0.8 cm diameter) filled with 0.05 M NaCl electrolyte. The electrodes were attached to the thenar and hypothenar eminences of the left hand. Skin conductance was digitized at 10 Hz and stored on an IBM PC for offline analysis.

To determine the amplitude of stimulus-related SCRs, we decomposed the skin conductance tracing into tonic and phasic components using an individually fitted template of a discrete SCR for each participant (Lim et al., 1997). In a first step, the algorithm that was implemented using the statistical programming language R (<http://www.r-project.org>) generated a template to match the individual SCR morphology. This template was optimized by minimizing the squared difference between the measured electrodermal data and the modeled response. In a second step, this SCR template was fitted to the whole skin conductance tracing of the respective participant. Additional SCRs were added when the model fit related to its complexity increased which was quantified using the Bayesian information criterion. The procedure resulted in a set of SCRs for each electrodermal recording that best resembled the measured data. Subsequently, SCRs that were elicited by the stimuli were identified by searching for responses with an onset between 1 and 3 s after stimulus onset. The amplitudes of these responses were finally log-transformed using the natural logarithm (Venables and Christie, 1980). To allow for a meaningful comparison of the ERPs with the electrodermal and behavioral data, we used the same trial selection as described above for these data channels.

To examine the effects of question (central vs. peripheral) and item type (probe vs. irrelevant) on the behavioral, electrodermal, and ERP data, we conducted a series of repeated measures analyses of variance (ANOVAs) on the corresponding dependent variables. Responses to target items were not included in the analyses since these items are typically not used to detect concealed knowledge (Meijer et al., 2007) and they were only included in the current study to ensure that participants are paying attention to the stimulus presentation. In addition to the factorial analyses, we also examined the interrelation between response systems. To this aim, we first computed differences between SCR, N200, and P300 amplitudes to probes and irrelevant items separately for central and peripheral details as well as for the whole test and subsequently calculated correlation coefficients between these measures.

In a second set of analyses, we tested whether the ISI affected SCR and ERP measures. To this aim, we separately averaged physiological responses for trials that were preceded by a short or a long ISI using a median split of the ISI distribution within each examinee. Ties were broken at random. The average ISIs amounted to 3236 ms ($SD = 64$ ms) for short ISIs and 6698 ms ($SD = 104$ ms) for long ones. After splitting the ISI distribution, we averaged the responses for irrelevant, probe and target details. Responses to central and peripheral questions were pooled in this analysis to have sufficient trials for ERP averaging. Finally, we

conducted a series of repeated measures ANOVAs on these values using ISI (short vs. long) and item type (probe vs. irrelevant) as factors.

A rejection region of $p < 0.05$ was used for all statistical tests but effects yielding p -values below 0.10 are mentioned as marginally significant effects. Cohen's f (Cohen, 1988) is reported as an effect size estimate for ANOVA results.

RESULTS

POST-EXPERIMENTAL MEMORY TEST

All participants remembered both central details (recognition rate $M = 100\%$) and all but two participants additionally recognized both peripheral details ($M = 95\%$, $SD = 15\%$). The participants who forgot crime-related information only recognized one of two peripheral details. No statistically significant difference was observed between the memory for central and peripheral details, $t_{(19)} = 1.45$, $p = 0.16$. The whole pattern of results that is mentioned in the following does not change when confining the analyses to the 18 participants with perfect recognition of central and peripheral crime details.

BEHAVIORAL RESPONSES

The ANOVA on the proportion of correct responses yielded a significant main effect of question type, $[F_{(1, 19)} = 5.81, p = 0.03, f = 0.18]$, indicating that responses within the set of peripheral CIT questions were slightly more accurate (see Table 1). The main effect of item type, $[F_{(1, 19)} = 2.78, p = 0.11, f = 0.23]$, as well as the interaction, $[F_{(1, 19)} = 2.10, p = 0.16, f = 0.13]$, failed to reach statistical significance.

With respect to the response times, we observed a statistically significant main effect of item type, $[F_{(1, 19)} = 29.53, p < 0.001, f = 0.37]$. Responses to probes were slower than responses to irrelevant items and this effect seemed to be more pronounced for central details as indicated by a marginally significant interaction of question and item type, $[F_{(1, 19)} = 4.33, p = 0.05, f = 0.09]$. Moreover, response times tended to be longer for peripheral questions on average, $[F_{(1, 19)} = 4.32, p = 0.05, f = 0.15]$ (see Table 1).

Table 1 | Average proportion of correct responses and reaction times for central and peripheral mock crime details as a function of item type.

Question type	Item type	Correct responses <i>M</i> (<i>SD</i>)	Reaction time (ms) <i>M</i> (<i>SD</i>)
Central	Irrelevant	99.6% (0.6%)	534.7 (45.7)
	Probe	98.7% (2.0%)	592.3 (57.9)
	Target	90.0% (7.4%)	636.2 (61.4)
Peripheral	Irrelevant	99.8% (0.5%)	566.7 (63.2)
	Probe	99.5% (1.6%)	600.7 (78.6)
	Target	91.0% (6.4%)	663.8 (78.7)

Note: Mean and standard deviation for the proportion of correct responses were calculated for all trials of the corresponding experiment. In contrast, for reaction times, these values were computed on the basis of all valid responses that were selected for each question and item type (see text).

PHYSIOLOGICAL MEASURES: EFFECTS OF QUESTION TYPE

As can be seen in the grand average ERPs that are depicted in **Figure 2**, all items were associated with a prominent P300 at Pz that was most pronounced for targets. The ANOVA on the P300 amplitudes revealed a significant main effect of item type, $[F_{(1, 19)} = 28.90, p < 0.001, f = 0.19]$, indicating that P300 amplitudes to probes were larger than to irrelevant items (see **Figure 3A**). The main effect of question type, $[F_{(1, 19)} = 0.18, p = 0.68, f = 0.03]$, and the interaction of question and item type were not statistically significant, $[F_{(1, 19)} = 1.09, p = 0.31, f = 0.04]$.

N200 effects were also evident in the grand average ERPs (**Figure 2**) but seemed to be comparable between item types. The statistical analysis of the N200 amplitudes only revealed a significant main effect of question type, $[F_{(1, 19)} = 8.22, p < 0.01, f = 0.19]$, indicating that central details were accompanied by an enhanced N200 (see **Figure 3A**). Neither the main effect of item type, $[F_{(1, 19)} = 0.12, p = 0.73, f = 0.02]$, nor the interaction of both factors reached statistical significance, $[F_{(1, 19)} = 0.21, p = 0.65, f = 0.02]$.

The response pattern of electrodermal responses was slightly different than that of the ERP measures. The ANOVA also yielded a significant main effect of item type, $[F_{(1, 19)} = 13.18, p < 0.01, f = 0.28]$, indicating that SCRs were larger for probes as compared to irrelevant items. However, we additionally obtained a significant interaction of question and item type, $[F_{(1, 19)} = 9.20, p < 0.01, f = 0.08]$, demonstrating that differential SCR amplitudes were more pronounced for central CIT questions (see **Figure 3A**). Overall, a significant main effect of question type, $[F_{(1, 19)} = 5.74, p = 0.03, f = 0.10]$, further indicates that SCR amplitudes were larger within the set of central CIT questions.

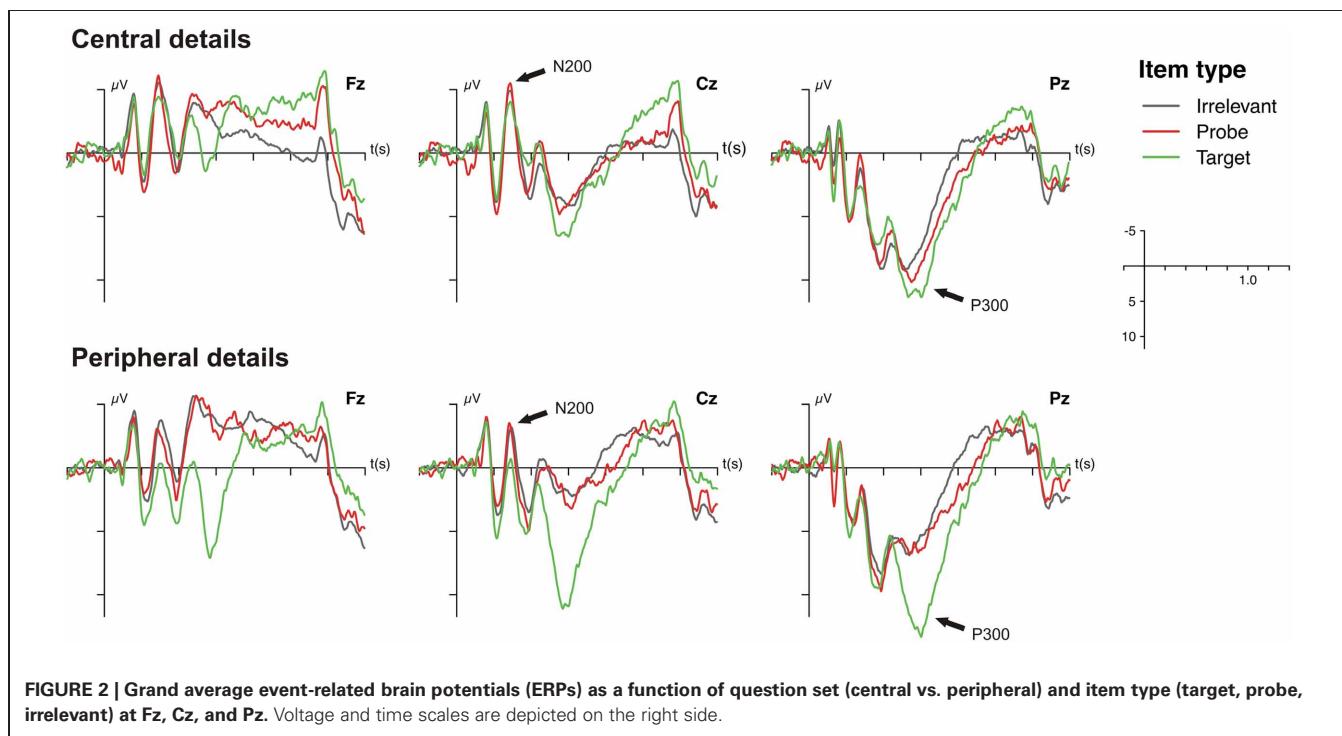
Correlations between SCR, N200, and P300 amplitude differences contrasting probes and irrelevant CIT items were not significant when splitting the question set into central and peripheral items (**Table 2**). Only when pooling responses across the whole test, moderate correlations between SCR amplitudes on the one hand and N200 as well as P300 amplitudes on the other hand emerged. Importantly, positive associations were observed which resemble the expected pattern for SCR and P300 amplitudes. However, since larger N200 amplitudes were thought to index probe recognition (Matsuda et al., 2009; Gamer and Berti, 2010), a positive correlation between SCR and N200 amplitudes was not expected.

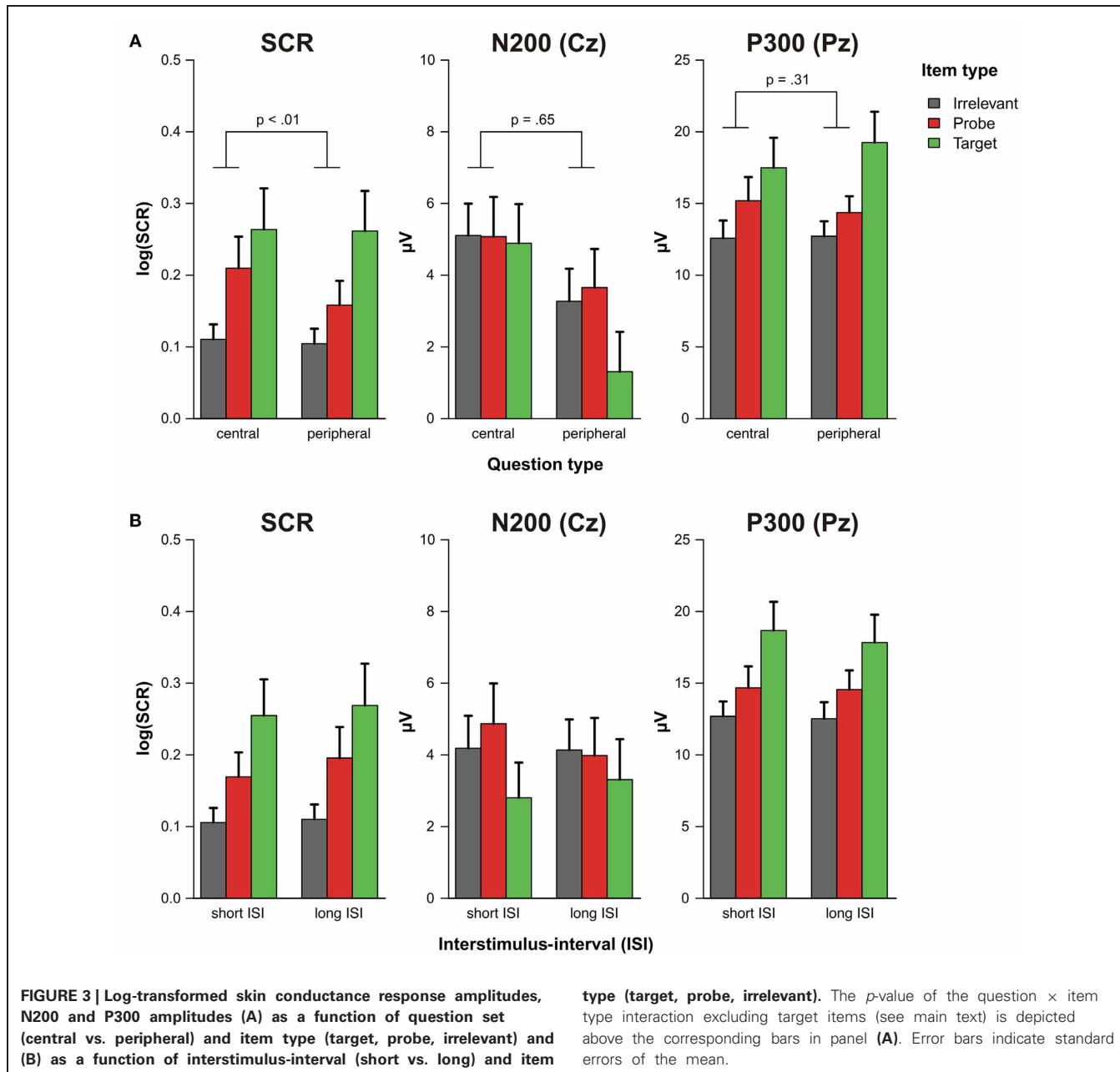
PHYSIOLOGICAL MEASURES: EFFECTS OF ISI

The pattern of the SCR and the ERP responses was highly similar irrespective of whether a trial was preceded by a relatively short or long ISI (**Figure 3B**). The ANOVA on the P300 amplitudes yielded a significant main effect of item type, $[F_{(1, 19)} = 21.42, p < 0.001, f = 0.18]$, indicating larger responses to probes as compared to irrelevant items. Neither the main effect of ISI, $[F_{(1, 19)} = 0.15, p = 0.71, f = 0.01]$, nor the interaction of both factors reached statistical significance, $[F_{(1, 19)} < 0.01, p = 0.96, f < 0.01]$.

In the ANOVA on the N200 amplitudes, we did not obtain any significant effect: Main effect of item type, $[F_{(1, 19)} = 0.33, p = 0.57, f = 0.03]$, main effect of ISI, $[F_{(1, 19)} = 0.64, p = 0.43, f = 0.05]$, interaction of both factors, $[F_{(1, 19)} = 0.76, p = 0.39, f = 0.05]$.

For the electrodermal responses, the ANOVA yielded a significant main effect of item type, $[F_{(1, 19)} = 13.66, p < 0.01, f = 0.28]$, demonstrating larger SCR amplitudes to probes than





to irrelevant items. The main effect of ISI, $[F_{(1, 19)} = 2.31, p = 0.14, f = 0.05]$, as well as the interaction of ISI and item type, $[F_{(1, 19)} = 0.97, p = 0.34, f = 0.04]$, failed to reach statistical significance.

DISCUSSION

The current study aimed at examining whether depth of processing differentially affects electrodermal and ERP measures in a CIT. Although crime related details were only incidentally encoded, participants showed very high recognition rates of central and peripheral crime details in a memory test that was conducted one week after the mock crime. Since recognition memory did not differ significantly between question types, any differences

in behavioral or physiological responses between central and peripheral crime details could not be attributed to differences in explicit memory.

Consistent with previous studies using a comparable CIT protocol, we observed longer response times for probes as compared to irrelevant items (Farwell and Donchin, 1991; Seymour et al., 2000; Gamer and Berti, 2010). This effect tended to be more pronounced for central details. Moreover, we obtained an overall trend for longer response times in the set of peripheral details, which might indicate that the response selection task was more difficult for this type of information. However, this effect could also result from a speed-accuracy tradeoff since response accuracy was also higher for peripheral crime details.

Table 2 | Intercorrelations between the response differences of all physiological measures for central and peripheral mock crime details as well as for the whole test.

	N200	P300
CENTRAL QUESTIONS		
SCR	0.35 (0.13)	0.33 (0.15)
N200	—	0.02 (0.92)
PERIPHERAL QUESTIONS		
SCR	0.13 (0.59)	0.15 (0.54)
N200	—	0.05 (0.82)
WHOLE TEST		
SCR	0.45 (0.04)	0.45 (0.04)
N200	—	0.16 (0.50)

Note: Values in brackets indicate *p*-values derived from tests for significant differences from $r = 0$ ($N = 20$). SCR, skin conductance response.

Electrodermal responses were significantly larger for probes than for irrelevant items but this difference was more pronounced for central mock crime details. This pattern of results replicates previous studies showing enhanced electrodermal responses to central as compared to peripheral crime details (Nahari and Ben-Shakhar, 2011; Peth et al., 2012) even in the absence of differences in explicit memory (Gamer et al., 2012). Thus, it seems that SCRs are sensitive to depth of processing.

In this study, we expanded current research by also testing whether an optimization of the experimental stimulation in terms of item sequence and ISIs (Dale, 1999) would improve the quantification of SCRs while keeping the whole duration of the experiment sufficiently short. It was recently shown that ISIs can be reduced to 10 s without reducing CIT validity (Breska et al., 2010). In the current study, the mean ISI was as short as 5 s but we still observed substantial differences in SCR amplitudes between probes and irrelevant items. Moreover, the ISI did not have a significant influence on this response pattern. Thus, SCR amplitudes were stable irrespective of whether a given stimulus was preceded by a relatively short (~3.2 s) or long (~6.7 s) ISI. Taken together, our procedure seems to be a viable method for future studies requiring short ISIs because of a simultaneous measurement of electrodermal responses and ERPs.

Consistent with previous studies, we observed larger P300 amplitudes for probes as compared to irrelevant items (Rosenfeld, 2011). Interestingly however, this effect was similar for central and peripheral crime details. Thus, in contrast to electrodermal responses, this measure might be less affected by depth of processing and seems to primarily reflect successful item recognition (Meijer et al., 2009). This result is at odds with previous studies showing enhanced P300 responses to deeply encoded episodic or autobiographical information as compared to incidentally acquired knowledge or shallowly encoded details (Ferlazzo et al., 1993; Ellwanger et al., 1996; Rosenfeld et al., 2006, 2007). However, the autobiographical information that was used in these previous studies was very salient and highly relevant to the participant. Such information might not be representative for episodic memories even when they concern deeply encoded details of high personal relevance (e.g., a weapon that was used

in a murder). Importantly, the current data does not suggest that a large number of peripheral details could be included in a CIT examination without affecting the validity of P300 amplitudes for detecting concealed knowledge. Since peripheral details are usually remembered less well especially when the CIT is conducted weeks or even months after the crime, it is likely that CIT validity will drop when including such details in the examination (Gamer et al., 2010; Nahari and Ben-Shakhar, 2011; Peth et al., 2012). In the current study, there was no difference in explicit memory and only under these circumstances, P300 amplitudes seem to reflect successful recognition instead of encoding depth.

Unexpectedly, ISI did not influence P300 amplitudes in the current study, and response differences between probes and irrelevant items were stable irrespective of the preceding ISI. Previous studies showed that P300 amplitudes depend on stimulus probability, stimulus sequence structure and ISI (Polich and Bondurant, 1997; Sambeth et al., 2004). All these factors affect the target-to-target interval (TTI) and it has been demonstrated that the TTI is indeed the major determinant of P300 amplitudes (Gonsalvez et al., 1999; Gonsalvez and Polich, 2002). Moreover, it has been described that the temporal structure of events affects P300 amplitudes even when TTI is kept constant (Schwartz et al., 2011). Thus, a random ISI as used in the current study should reduce P300 amplitudes. However, we did neither observe an effect of ISI on P300 amplitudes, nor did we obtain reduced P300 responses. By contrast, the overall pattern of P300 amplitudes in the current study as well as their size were very similar to previous CIT studies using a short, fixed ISI (e.g., Rosenfeld et al., 2006, 2007; Verschueren et al., 2009). It seems possible that effects of the preceding ISI were reduced in the current study because of the use of random ISIs. Thus, even though the timing between successive stimuli was highly variable in the current study, the interval between stimuli of the same category (i.e., probe, target, irrelevant item) was relatively stable due to the random nature of the stimulus sequence. Therefore, the overall influence of our session structure and timing on P300 amplitudes might have been less pronounced as compared to previous studies using different sets of fixed ISIs to examine predictors of P300 amplitudes (e.g., Polich, 1990). Moreover, the mean ISI in the current study (4952 ms) was much longer as compared to a previous study reporting a reduction of P300 amplitudes for random as compared to isochronous sequences (900 ms, Schwartz et al., 2011). Thus, it seems possible that effects of the temporal structure of stimulation are less pronounced when using larger ISIs. Nevertheless, it would be interesting and important for future studies to examine the influence of ISI structure (random vs. fixed) and length on P300 amplitudes in the CIT in more detail.

In contrast to other studies, we did not observe differences in N200 amplitudes between crime details and irrelevant CIT items (Matsuda et al., 2009; Gamer and Berti, 2010). The N200 has previously been linked to response monitoring demands as well as to the orienting of attentional resources (for a review see Folstein and Van Petten, 2008). Thus, it was reasoned that enhanced N200 amplitudes to crime related information in the CIT might index the automatic orienting of attention toward probe items in order to facilitate a more extensive processing of such personally relevant information (Matsuda et al., 2009).

Additionally, it was suggested that the N200 reflects enhanced response monitoring as a pre-requisite for correctly responding to probes that pop out of the stimulus stream (as targets do) but usually require a different behavioral response (Gamer and Berti, 2010). Both these circumstances also apply to the current study but we did not obtain differences in N200 amplitudes between crime related and irrelevant details. However, the present study differs from previous experiments with respect to the stimulus timing (i.e., the ISI) as well as to the stimuli that were used in the CIT: In line with recommendations for the field use of the CIT, we constructed our item set in such a way that all items were clearly separable from each other but equally plausible for an innocent examinee (Nakayama, 2002; Meijer et al., 2011). By contrast, in the study by Gamer and Berti (2010), the visual stimuli presented as probes and targets were perceptually less distinct (for instance, the jack of spades vs. the king of spades from a set of playing cards). Matsuda et al. (2009) used spoken digits that also form a more homogeneous stimulus set than the natural visual objects used in the present study. Indeed, it has been reported that N200 effects are modulated by the perceptual overlap of stimuli that require different behavioral responses (Nieuwenhuis et al., 2004): When stimuli were clearly separable, the N200 was substantially reduced. Therefore, a lack of an effect of item type in our results might be attributed to the fact that crime details and irrelevant CIT items were easily separable in the current study and presumably did not require additional processing demands. This is in line with the functional interpretation by Folstein and Van Petten (2008), supposing that the fronto-central N200 is a correlate of cognitive control, reflecting for instance attentional allocation to one (Gramann et al., 2007) or different behaviorally relevant visual dimensions (Berti and Wühr, 2012). Since we did not observe a differential N200 between crime related and irrelevant items but instead a general difference between central and

peripheral CIT questions, it seems that the N200 does not mirror processing of crime related information per se but is more dependent on stimulus characteristics. In line with previous studies, ISI did not affect N200 amplitudes (Polich, 1990), but it seems that certain features of the stimulus set modulated N200 amplitudes. Since we have no further interpretation of why N200 responses were generally enhanced for central details, we propose that further investigations have to determine when enhanced N200 amplitudes to specific items can be expected in the CIT.

To sum up, the present study revealed a differential sensitivity of ERP measures and electrodermal responses to depth of processing. We observed larger SCRs for central items along with stable P300 responses across question types. This differential sensitivity of response systems might be one reason for the small correlations between measures (Matsuda et al., 2009; Gamer and Berti, 2010) and the incremental validity that has been reported previously (Ambach et al., 2010; Matsuda et al., 2011). Thus, also from an applied perspective, it seems useful to combine different physiological measures that cover partly different psychological processes that are all involved in CIT examinations (e.g., item recognition, attentional orienting, response selection and monitoring). An important question for future research is the identification and characterization of these processes that might differentially affect autonomic (Ambach et al., 2008; Gamer et al., 2008) and central nervous system responses in the CIT (Gamer and Berti, 2010).

ACKNOWLEDGMENTS

We thank Matthias Wagner for his help and support during data collection and analysis and Nicole Waller for comments on the written English form. Part of this work was supported by grants from the European Science Foundation (09-ECRP-025) and the German Research Foundation (GA 1621/1-1).

REFERENCES

- Abootalebi, V., Moradi, M. H., and Khalilzadeh, M. A. (2006). A comparison of methods for ERP assessment in a P300-based GKT. *Int. J. Psychophysiol.* 62, 309–320.
- Allen, J. J., Iacono, W. G., and Danielson, K. D. (1992). The identification of concealed memories using the event-related potential and implicit behavioral measures: a methodology for prediction in the face of individual differences. *Psychophysiology* 29, 504–522.
- Ambach, W., Bursch, S., and Stark, R. (2010). A Concealed Information Test with multimodal measurement. *Int. J. Psychophysiol.* 75, 258–267.
- Ambach, W., Dummel, S., Lüer, T., and Vaitl, D. (2011). Physiological responses in a Concealed Information Test are determined interactively by encoding procedure and questioning format. *Int. J. Psychophysiol.* 81, 275–282.
- Ambach, W., Stark, R., Peper, M., and Vaitl, D. (2008). Separating deceptive and orienting components in a Concealed Information Test. *Int. J. Psychophysiol.* 70, 95–104.
- Ben-Shakhar, G., Bar-Hillel, M., and Kremnitzer, M. (2002). Trial by polygraph: reconsidering the use of the guilty knowledge technique in court. *Law Hum. Behav.* 26, 527–541.
- Ben-Shakhar, G., and Elaad, E. (2003). The validity of psychophysiological detection of information with the Guilty Knowledge Test: a meta-analytic review. *J. Appl. Psychol.* 88, 131–151.
- Berti, S., and Wühr, P. (2012). Using redundant visual information from different dimensions for attentional selection. *J. Psychophysiol.* 26, 99–104.
- Breska, A., Maoz, K., and Ben-Shakhar, G. (2010). Interstimulus intervals for skin conductance response measurement. *Psychophysiology* 48, 437–440.
- Carmel, D., Dayan, E., Naveh, A., Raveh, O., and Ben-Shakhar, G. (2003). Estimating the validity of the Guilty Knowledge Test from simulated experiments: the external validity of mock crime studies. *J. Exp. Psychol. Appl.* 9, 261–269.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd Edn. Hillsdale, NJ: Erlbaum.
- Craik, F. I. M., and Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *J. Exp. Psychol.* 104, 268–294.
- Dale, A. M. (1999). Optimal experimental design for event-related fMRI. *Hum. Brain Mapp.* 8, 109–114.
- Dawson, M. E., Schell, A. M., and Filion, D. L. (2007). “The electrodermal system,” in *Handbook of Psychophysiology*, eds J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson (Cambridge, MA: University Press), 159–181.
- Ellwanger, J., Rosenfeld, J. P., Sweet, J. J., and Bhatt, M. (1996). Detecting simulated amnesia for autobiographical and recently learned information using the P300 event-related potential. *Int. J. Psychophysiol.* 23, 9–23.
- Farwell, L. A., and Donchin, E. (1991). The truth will out: interrogative polygraphy (“lie detection”) with event-related brain potentials. *Psychophysiology* 28, 531–547.
- Ferlazzo, F., Conte, S., and Gentilomo, A. (1993). Event-related potentials and recognition memory within the “levels of processing” framework. *Neuroreport* 4, 667–670.
- Folstein, J. R., and Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: a review. *Psychophysiology* 45, 152–170.
- Gamer, M. (2011). “Detecting concealed information using

- autonomic measures," in *Memory Detection: Theory and Application of the Concealed Information Test*, eds B. Verschueren, G. Ben-Shakhar, and E. H. Meijer (Cambridge, MA: University Press), 27–45.
- Gamer, M., and Berti, S. (2010). Task relevance and recognition of concealed information have different influences on electrodermal activity and event-related brain potentials. *Psychophysiology* 47, 355–364.
- Gamer, M., Gödert, H. W., Keth, A., Rill, H.-G., and Vossel, G. (2008). Electrodermal and phasic heart rate responses in the Guilty Actions Test: comparing guilty examinees to informed and uninformed innocents. *Int. J. Psychophysiol.* 69, 61–68.
- Gamer, M., Klimecki, O., Bauermann, T., Stoeter, P., and Vossel, G. (2012). fMRI-activation patterns in the detection of concealed information rely on memory-related effects. *Soc. Cogn. Affect. Neurosci.* 7, 506–515.
- Gamer, M., Kosiol, D., and Vossel, G. (2010). Strength of memory encoding affects physiological responses in the Guilty Actions Test. *Biol. Psychol.* 83, 101–107.
- Gonsalvez, C. J., Gordon, E., Grayson, S., Barry, R. J., Lazzaro, I., and Bahramali, H. (1999). Is the target-to-target interval a critical determinant of P3 amplitude? *Psychophysiology* 36, 643–654.
- Gonsalvez, C. L., and Polich, J. (2002). P300 amplitude is determined by target-to-target interval. *Psychophysiology* 39, 388–396.
- Gramann, K., Toellner, T., Krummacher, J., Eimer, M., and Müller, H. J. (2007). Brain electrical correlates of dimensional weighting: an ERP study. *Psychophysiology* 44, 277–292.
- Lim, C. L., Rennie, C., Barry, R. J., Bahramali, H., Lazzaro, I., Manor, B., et al. (1997). Decomposing skin conductance into tonic and phasic components. *Int. J. Psychophysiol.* 25, 97–109.
- Lykken, D. T. (1959). The GSR in the detection of guilt. *J. Appl. Psychol.* 43, 385–388.
- Matsuda, I., Hirota, A., Ogawa, T., Takasawa, N., and Shigematsu, K. (2006). A new discrimination method for the Concealed Information Test using pretest data and within-individual comparisons. *Biol. Psychol.* 73, 157–164.
- Matsuda, I., Nittono, H., Hirota, A., Ogawa, T., and Takasawa, N. (2009). Event-related brain potentials during the standard autonomic-based concealed information test. *Int. J. Psychophysiol.* 74, 58–68.
- Matsuda, I., Nittono, H., and Ogawa, T. (2011). Event-related potentials increase the discrimination performance of the autonomic-based concealed information test. *Appl. Psychophysiol. Biofeedback* 34, 221–226.
- Meijer, E. H., Smulders, F. T., and Wolf, A. (2009). The contribution of mere recognition to the p300 effect in a concealed information test. *Appl. Psychophysiol. Biofeedback* 34, 221–226.
- Meijer, E. H., Verschueren, B., and Ben-Shakhar, G. (2011). "Practical guidelines for developing a CIT," in *Memory Detection: Theory and Application of the Concealed Information Test*, eds B. Verschueren, G. Ben-Shakhar, and E. H. Meijer (Cambridge, MA: University Press), 63–89.
- Mertens, R., and Allen, J. J. (2008). The role of psychophysiology in forensic assessments: deception detection, ERPs, and virtual reality mock crime scenarios. *Psychophysiology* 45, 286–298.
- Nahari, G., and Ben-Shakhar, G. (2011). Psychophysiological and behavioral measures for detecting concealed information: the role of memory for crime details. *Psychophysiology* 48, 733–744.
- Nakayama, M. (2002). "Practical use of the concealed information test for criminal investigation in Japan," in *Handbook of Polygraph Testing*, ed M. Kleiner (San Diego, CA: Academic Press), 49–86.
- Nieuwenhuis, S., Yeung, N., and Cohen, J. D. (2004). Stimulus modality, perceptual overlap, and the go/no-go N2. *Psychophysiology* 41, 157–160.
- Peth, J., Vossel, G., and Gamer, M. (2012). Emotional arousal modulates the encoding of crime-related details and corresponding physiological responses in the Concealed Information Test. *Biol. Psychol.* 87, 146–151.
- Concealed Information Test. *Psychophysiology* 49, 381–390.
- Polich, J. (1990). Probability and inter-stimulus interval effects on the P300 from auditory stimuli. *Int. J. Psychophysiol.* 10, 163–170.
- Polich, J., and Bondurant, T. (1997). P300 sequence effects, probability, and interstimulus interval. *Physiol. Behav.* 61, 843–849.
- Rosenfeld, J., Angell, A., and Johnson, M. (1991). An ERP-based, control-question lie detector analog: algorithms for discriminating effects within individuals' average waveforms. *Psychophysiology* 28, 319–335.
- Rosenfeld, J. P. (2011). "P300 in detecting deception," in *Memory Detection: Theory and Application of the Concealed Information Test*, eds B. Verschueren, G. Ben-Shakhar, and E. H. Meijer (Cambridge, MA: University Press), 63–89.
- Rosenfeld, J. P., Biroshak, J. R., and Furedy, J. J. (2006). P300-based detection of concealed autobiographical versus incidentally acquired information in target and non-target paradigms. *Int. J. Psychophysiol.* 60, 251–259.
- Rosenfeld, J. P., Ellwanger, J., and Sweet, J. (1995). Detecting simulated amnesia with event-related brain potentials. *Int. J. Psychophysiol.* 19, 1–11.
- Rosenfeld, J. P., Shue, E., and Singer, E. (2007). Single versus multiple probe blocks of P300-based concealed information tests for self-referring versus incidentally obtained information. *Biol. Psychol.* 74, 396–404.
- Rosenfeld, J. P., Soskins, M., Bosh, G., and Ryan, A. (2004). Simple, effective countermeasures to P300-based tests of detection of concealed information. *Psychophysiology* 41, 205–219.
- Sambeth, A., Maes, J. H., and Brankack, J. (2004). With long intervals, inter-stimulus interval is the critical determinant of the human P300 amplitude. *Neurosci. Lett.* 359, 143–146.
- Schwartz, M., Rothermich, K., Schmidt-Kassow, M., and Kotz, S. A. (2011). Temporal regularity effects on pre-attentive and attentive processing of deviance. *Biol. Psychol.* 87, 146–151.
- Seymour, T. L., Seifert, C. M., Shafto, M. G., and Mosmann, A. L. (2000). Using response time measures to assess "guilty knowledge." *J. Appl. Psychol.* 85, 30–37.
- Soskins, M., Rosenfeld, J. P., and Niendam, T. (2001). Peak-to-peak measurement of P300 recorded at 0.3 Hz high pass filter settings in intraindividual diagnosis: complex vs. simple paradigms. *Int. J. Psychophysiol.* 40, 173–180.
- van Hooff, J. C., Brunia, C. H., and Allen, J. J. (1996). Event-related potentials as indirect measures of recognition memory. *Int. J. Psychophysiol.* 21, 15–31.
- Venables, P. H., and Christie, M. J. (1980). "Electrodermal activity," in *Techniques in Psychophysiology*, eds I. Martin and P. H. Venables (Chichester: Wiley), 3–67.
- Verschueren, B., Ben-Shakhar, G., and Meijer, E. H. (2011). *Memory Detection: Theory and Application of the Concealed Information Test*. Cambridge, MA: Cambridge University Press.
- Verschueren, B., Rosenfeld, J. P., Winograd, M. R., Labkovsky, E., and Wiersema, R. (2009). The role of deception in P300 memory detection. *Leg. Crim. Psychol.* 14, 253–262.
- Vrij, A. (2008). *Detecting Lies and Deceit. Pitfalls and Opportunities*. Chichester: Wiley.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 16 July 2012; accepted: 24 October 2012; published online: 15 November 2012.

Citation: Gamer M and Berti S (2012) P300 amplitudes in the concealed information test are less affected by depth of processing than electrodermal responses. Front. Hum. Neurosci. 6:308. doi: 10.3389/fnhum.2012.00308

Copyright © 2012 Gamer and Berti. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.



"Have you ever seen this face?" – individual differences and event-related potentials during deception

Anja Leue^{1,2*}, Sebastian Lange² and André Beauducel²

¹ Clinic of Epileptology, University of Bonn, Bonn, Germany

² Institute of Psychology, University of Bonn, Bonn, Germany

Edited by:

Matthias Gamer, University Medical Center Hamburg-Eppendorf, Germany

Reviewed by:

Jun'ichi Katayama, Kwansei Gakuin University, Japan

Stefan Berti, Johannes Gutenberg University Mainz, Germany

***Correspondence:**

Anja Leue, Clinic of Epileptology and Institute of Psychology, University of Bonn, Dietkirchenstrasse 28, D-53111 Bonn, Germany.

e-mail: anja.leue@uni-bonn.de

Deception studies emphasize on the importance of event-related potentials (ERP) for a reliable differentiation of the underlying neuro-cognitive processes. The stimulus-locked parietal P3 amplitude has been shown to reflect stimulus salience but also attentional control available for stimulus processing. Known stimuli requiring truthful responses (targets) and known stimuli requiring deceptive responses (probes) were hypothesized to be more salient than unknown stimuli. Thus, a larger P3 was predicted for known truthful and deceptive stimuli than for unknown stimuli. The Medial Frontal Negativity (MFN) represents the amount of required cognitive control and was expected to be more negative to known truthful and deceptive stimuli than to unknown stimuli. Moreover, we expected higher sensitivity to injustice (SI-perpetrator) and aversiveness (Trait-BIS) to result in more intense neural processes during deception. $N = 102$ participants performed a deception task with three picture types: probes requiring deceptive responses, targets requiring truthful responses to known stimuli, and irrelevants being associated with truthful responses to unknown stimuli. Repeated-measures ANOVA and fixed-links modeling suggested a more positive parietal P3 and a more negative frontal MFN to deceptive vs. irrelevant stimuli. Trait-BIS and SI-perpetrator predicted an increase of the P3 and a decrease of the MFN from irrelevants to probes. This suggested an intensification of stimulus salience and cognitive control across picture types in individuals scoring either higher on Trait-BIS or higher on SI-perpetrator. In contrast, individuals with both higher Trait-BIS and higher SI-perpetrator scores showed a less negative probe-MFN suggesting that this subgroup invests less cognitive control to probes. By extending prior research we demonstrate that personality modulates stimulus salience and control processes during deception.

Keywords: deception, P3, MFN, individual differences, fixed-links modeling

INTRODUCTION

One of the main interests in forensic psychophysiology refers to the differentiation of truthful and deceptive responses. Referring to different cognitive models on deception (e.g., Zuckerman et al., 1981; Walczyk et al., 2003), there is a considerable number of studies investigating the underlying processes of deception by means of verbal and non-verbal behavior (DePaulo et al., 2003) or behavioral parameters (Zuckerman et al., 1981). Moreover, the relevance of the P3 amplitude of the event-related potential (ERP) has been originally demonstrated for persons who recognize items on that information should be concealed (sometimes named as guilty group) vs. persons who do not recognize those items (sometimes named as innocent group) because they are unfamiliar with (Rosenfeld et al., 1988; Farwell and Donchin, 1991). These studies encouraged a growing research interest in the modulation of ERPs in deception settings and in elucidating the underlying neuro-cognitive processes of deceptive vs. truthful responses. The relevance of ERPs for the differentiation of deceptive vs. truthful responses has been successfully illustrated in guilty knowledge tasks (GKT, also named as concealed information test, CIT) and other deception tasks (Farwell and Donchin, 1991; Allen et al.,

1992; Fang et al., 2003; Johnson et al., 2008; Ambach et al., 2010; Gamer and Berti, 2010).

A considerable number of studies investigated variations of the P3 component for deceptive vs. truthful stimuli by means of CITs (e.g., Rosenfeld et al., 1988, 1991; Farwell and Donchin, 1991; Mertens and Allen, 2008; Ambach et al., 2010; Gamer and Berti, 2010; Meixner and Rosenfeld, 2011) or visual recognition tasks (Fang et al., 2003; Meijer et al., 2007; Dong et al., 2010a). In deception tasks, participants learn subsets of stimuli that require deceptive vs. truthful responses prior to task performance. Most P300-based CITs comprise three types of stimuli: probe, target, and irrelevant stimuli. Probe stimuli are deception-relevant stimuli that are known by participants who are requested to deceive the knowledge of these stimuli in their responses. Target stimuli are known by participants and they require truthful responses. Target stimuli are useful to ensure that participants attend to the presented stimuli and do not ignore them (cf. Farwell and Donchin, 1991; Fang et al., 2003; Mertens and Allen, 2008; Gamer and Berti, 2010). Irrelevant stimuli incorporate stimuli that participants have not seen before task performance (e.g., Meijer et al., 2009; Meixner and Rosenfeld, 2011). All deception tasks have in common that

individuals “need to consciously select and execute a response that is incompatible with the truth. . .” (Johnson et al., 2008, p. 469). Accordingly, executive processes like attentional control and cognitive control should play an important role during deception (e.g., Gombos, 2006; Carrión et al., 2010).

Although different conceptual meanings have been discussed for the P3 amplitude depending on tasks and context (Mulder, 1986; Mecklinger et al., 1992; Kok, 2001; Beauducel et al., 2006; Polich, 2007), the parietal P3 amplitude (mainly occurring between 300 and 800 ms post-stimulus) is one of the most frequently investigated ERPs in deception tasks. In some studies on P3 and deception, the P3 is regarded as an indicator of task relevance or stimulus salience leading to larger P3 amplitudes for target and probe stimuli compared to irrelevant stimuli (e.g., Ambach et al., 2010; Gamer and Berti, 2010). Moreover, deceptive responses might involve additional processes related to justification or ethical discomfort that are not relevant for target stimuli. Moreover, effects of personality on P3 amplitudes have been demonstrated in different quasi-experimental settings (Beauducel et al., 2006; Leue et al., 2009; Wacker et al., 2010). The P3 amplitude captures individual differences that have been related to stimulus salience or stimulus complexity (Stenberg, 1994; Fink and Neubauer, 2004) and cognitive resources (Beauducel et al., 2006).

In addition to P3-related processes of salience and attentional control, response-related cognitive control has been discussed as a neuro-cognitive process during deception because individuals either have to adapt their responses to deceptive information in comparison to truthful information or they have to inhibit responses in order to successfully conceal knowledge. In this respect, the response-locked Medial Frontal Negativity (MFN) has been investigated as an indicator of cognitive control (e.g., Johnson et al., 2004, 2008). The MFN has a fronto-central topography and occurs 0–70 ms post-response in deception settings (Johnson et al., 2008). In non-deception settings, the MFN (or Error Related Negativity) is more negative when actions fail to meet motivational goals (Potts et al., 2006) and following erroneous responses compared to correct responses (e.g., Luu et al., 2000). Presuming that individuals interpret deceptive responses as erroneous responses (i.e., violating social norms), deceptive responses should be more aversive than truthful responses. Therefore, the MFN should be more negative following deceptive responses than following truthful responses (cf. Dong et al., 2010a, 2011).

To the best of our knowledge, there is no P3-study relating personality traits and neural responses to deceptive vs. truthful stimuli, whereas trait-related differences of the P3 have been intensely studied in other contexts (Stenberg, 1994; Fink and Neubauer, 2004; Beauducel et al., 2006; Leue et al., 2009; Wacker et al., 2010). However, studies on cognitive control reported that personality dimensions like fairness concerns modulate the feedback-locked MFN amplitude (Boksem and De Cremer, 2010). More precisely, Boksem and De Cremer (2010) reported a more negative MFN in an ultimatum game for individuals with high compared to low scores in moral identity. A conflict based on an individuals’ moral and social standards could also be induced by the instruction to conceal information. Therefore, individuals who are highly sensitive to moral and social norms should demonstrate a more negative MFN to probe compared to target and irrelevant items in

a deception task. A personality dimension that might correspond to an individuals’ sensitivity to moral and social norms is the trait-dimension sensitivity to injustice (SI, Schmitt et al., 2005). Another trait-dimension that reflects personality differences of cognitive-motivational conflict processing is Trait-BIS (Carver and White, 1994). Trait-BIS refers to the activation of the behavioral inhibition system (BIS) that serves as a device for conflict detection and resolution (Gray and McNaughton, 2000; Corr, 2008). Individuals with higher Trait-BIS scores show a more pronounced BIS activation compared to lower Trait-BIS individuals. Trait-BIS has been shown to reflect aversiveness sensitivity in conflict and non-conflict situations (Leue and Beauducel, 2008). Because deception might induce conflict with one’s social and moral standards (to respond honestly) deception could be aversive especially to higher vs. lower Trait-BIS individuals. This enhanced sensitivity to aversiveness might increase the salience of probes for higher vs. lower Trait-BIS individuals. Accordingly, we investigated whether probe stimuli compared to target and irrelevant stimuli are more salient for higher vs. lower Trait-BIS individuals. If this prediction would be true, a larger P3 amplitude should be observed in higher vs. lower Trait-BIS individuals for probe stimuli compared to target and irrelevant stimuli. Probe stimuli compared to target and irrelevant stimuli might be of special salience to higher Trait-BIS individuals because lying is an aversive event, and higher Trait-BIS individuals should be more sensitive to aversive events (cf. Corr, 2008; Leue and Beauducel, 2008). Moreover, according to the above-cited literature we hypothesized a more intense cognitive control of higher Trait-BIS individuals for probe stimuli relative to target and irrelevant stimuli. This should be indicated by a more negative probe-MFN in higher Trait-BIS individuals compared to lower Trait-BIS individuals. We also aimed at probing whether individuals with a higher sensitivity to social norms (i.e., higher SI scores) and a higher sensitivity to aversiveness (i.e., Trait-BIS) show more pronounced P3 and MFN-amplitudes on deceptive stimuli compared to truthful stimuli. Altogether, the present study aimed at investigating neuro-cognitive processes of deception – namely stimulus salience or attentional control (by means of stimulus-locked P3) and cognitive control (by means of response-locked MFN) – as well as the modulation of these processes by individual differences of SI and Trait-BIS.

From a more general point of view the investigation of inter-individual differences within deception processes brings together the correlative personality research tradition with the tradition of experimental deception research. The crossbreeding of correlative and quasi-experimental research has been promoted intensely since Cronbach (1975) so that the methodological approaches for modeling individual differences within repeated-measures designs have also been improved (Raudenbush and Bryk, 2002; Muthén, 2004). For example, Raudenbush and Bryk (2002) described individual differences as random effects in the context of hierarchical linear models whereas Muthén (2004) proposed a modeling of the individual differences together with treatment effects as latent variables in the context of structural equation modeling (SEM). Both approaches have their merits, but in the present context the modeling of individual differences and treatment effects as latent variables was considered as an advantage, because latent variables are regarded as more appropriate indicators of psychological

constructs than measured variables. Thus, measurement models can be specified in SEM that allow for a separation of construct relevant common variance represented by latent variables from the specific error variance. Since this separation is only possible with SEM, this framework was regarded as most appropriate for the present study (see Materials and Methods for further specifications). Nevertheless, conventional analyses of variance (ANOVA) were also reported in order to facilitate comparisons with previous research. Altogether, the complex aim of investigating individual differences (i.e., Trait-BIS, SI) in conjunction with treatment effects in a deception task fits well to the flexibility of the SEM approach.

MATERIALS AND METHODS

PARTICIPANTS

A total of $N = 114$ students from a German University participated individually in the present study. Written informed consent has been given by all participants. Artifacts that could not be corrected by means of Independent Component Analysis (ICA; see below) resulted in an in-sufficient number of trials per picture type (i.e., less than 20 trials per picture type) in 12 participants so that a sample of $N = 102$ (48 male, age: $M = 23.80$ years, $SD = 3.75$, range: 19–37 years) participants remained for the analysis of the P3 components. Due to an increased number of muscle artifacts during response preparation, a sample of $N = 91$ participants (42 male) was available for the analysis of the MFN. Based on the Edinburgh Handedness Inventory (Oldfield, 1971) all included participants were right-handed. The ethical standards of this study were approved by the ethical commission of the German Research Foundation.

MEASURES

Participants filled in the German version of the BIS/BAS scales (Strobel et al., 2001). The BIS/BAS scales measure an individual's sensitivity to aversiveness (Trait-BIS) and an individuals' sensitivity to appetitive reinforcement (Trait-BAS) with 24-items using a five-point Likert-type answer format. The Trait-BIS scale is an established personality scale in studies investigating individual differences of cognitive control (e.g., Boksem et al., 2006; Amadio et al., 2008; Lange et al., 2012; Leue et al., 2012a,b). Therefore, the Trait-BIS scale was applied to investigate individual differences of the P3 and the MFN in our deception study (Cronbach's $\alpha: 0.80$). The Sensitivity to Injustice questionnaire (Schmitt et al., 2005) measures individual differences of SI for different perspectives (perpetrator, victim, observer, and one's favor) and consists of 40 items with a seven-point answer format (0 = not at all, 6 = strong agreement). To elucidate those individual differences of deception that might be related to justice or fairness concerns, we focused on the SI-perpetrator subscale (10 items) in our ERP analyses (Cronbach's $\alpha: 0.87$) because this subscale is related to an individual's moral standards of feeling guilty when he/she treats others unfairly. Trait-BIS and SI-perpetrator correlated significantly, $r(102) = 0.24$, $p < 0.05$ (two-tailed).

DECEPTION TASK

The present task incorporated three types of pictures that were taken from the International Affective Picture System (IAPS,

Bradley and Lang, 2007) and the task was designed in accordance with the study of Fang et al. (2003). All selected pictures showed either a face of a woman or a face of a man. Regarding "probe" pictures (number of IAPS pictures: 2190, 2516, 2214), participants were asked to conceal their knowledge by pressing on the left cursor button as required for the irrelevant pictures (see below). On "target" pictures (number of IAPS pictures: 2500, 2305, 2215), participants should indicate truthfully by a button press on the right cursor button that they knew the pictures. Finally, there was a total of 20 "irrelevant" pictures that were completely unknown to the participants. Participants were asked to indicate truthfully by pressing on the left cursor button that they did not know them (number of IAPS pictures: 2372, 2383, 2512, 2200, 2210, 2221, 2630, 2104, 2102, 2495, 2510, 2230, 2005, 2020, 2493, 2000, 2010, 2385, 2499, 2513). We chose a large number of different pictures that participants have not seen before performing the task to ensure that these irrelevant pictures would remain rather strange and, thus, of low relevance throughout the experimental task. Averaged valence and arousal values have been calculated based on the IAPS manual for probe pictures, target pictures, and irrelevant pictures. Means of the valence dimension were widely comparable for the three picture types (probe: $M = 4.91$, $SD = 0.09$; target: $M = 5.40$, $SD = 0.77$, irrelevant: $M = 5.34$, $SD = 0.76$). The same was true for means of the arousal dimension (probe: $M = 3.12$, $SD = 0.62$; target: $M = 3.54$, $SD = 0.14$; irrelevant: $M = 3.48$, $SD = 0.37$).

All task-related instructions were presented on the screen. To make the requirement of concealing knowledge more salient, participants were encouraged to give their best so that the computer program could not detect based on EEG and response data when participants concealed knowledge to the probe pictures. In accordance with Fang et al. (2003), participants received the information that if the computer program recognized deception, they would lose 15 Cent even if they pressed the correct cursor button. Otherwise they would win 5 Cent. Altogether participants performed 150 trials (50 probe, 50 irrelevant, and 50 target items) presented in a pseudo-random order with a 2-min break after 75 trials. In order to realize the traditionally applied ratio with less frequently occurring probe and target stimuli relative to irrelevant stimuli, three different probe and three different target pictures were selected, whereas 20 different irrelevant pictures were chosen. Thus, the number of three different probe, three different irrelevant pictures, and three different target pictures followed a 3:20:3 ratio, which is comparable to other studies (e.g., Meijer et al., 2007). Thus, per task block (including 75 trials) each of the three probe pictures and each of the three target pictures was presented about eight times and most of the 20 irrelevant pictures were applied once.

Each trial consisted of a fixation point that was presented in the center of the TFT screen (20'') for 1000 ms followed by a picture presented for 700 ms (picture size: 6 cm × 4 cm). Participants were instructed to indicate the picture type (probe, target, or irrelevant) by pressing the left-hand site cursor for a probe or an irrelevant picture and by pressing the right-hand site cursor for a target picture as soon as they were sure of the picture type. When a picture disappeared after 700 ms, participants could respond up to a maximum of 2000 ms. During this time interval the screen remained black.

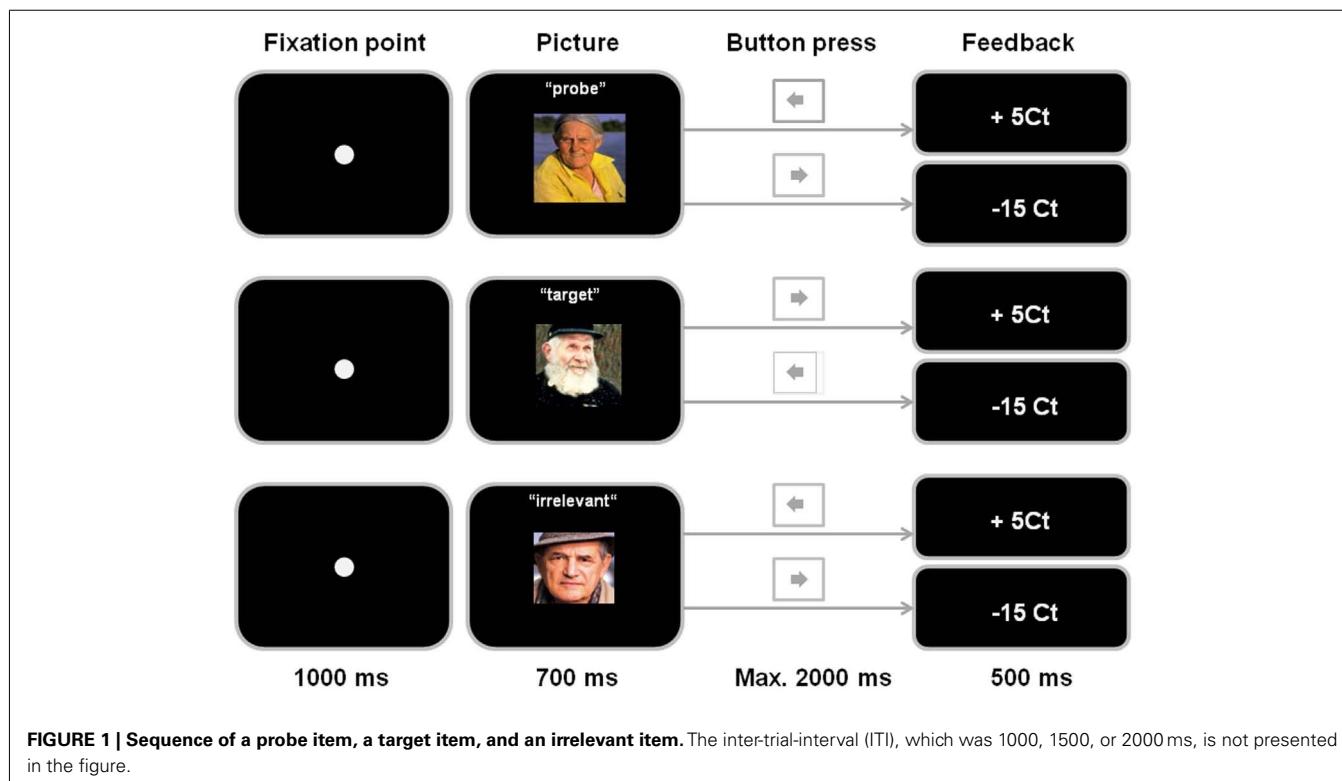
Correct responses to target pictures (i.e., pressing the right cursor button) and to irrelevant pictures (i.e., pressing the left cursor button) as well as successfully concealed knowledge to probe pictures (i.e., pressing the left cursor button) resulted in a win feedback (+5 Ct). Loss feedback (-15 Ct) occurred following each incorrect response (i.e., pressing the left cursor following a target picture and pressing the right cursor button to an irrelevant or probe picture). Moreover, following five out of 20 correct responses on probe items per block participants received a loss feedback (-15 Ct) even when they had pressed the left cursor button as required per instruction (cf. Fang et al., 2003). Participants always received a feedback that corresponded to the correctness of their responses in case of target and irrelevant pictures. Feedback to probe pictures corresponded to the correctness of their responses for 20 probe trials per block, whereas in five probe trials per block loss feedback (-15 Ct) occurred even when participants had correctly pressed the left cursor button (i.e., loss feedback was pre-defined). The sequence of a trial with a pre-defined loss feedback was as follows: participants saw a probe picture and responded to the left cursor button as they should for probes according to the instruction. Subsequently, they received a loss feedback of -15 Ct indicating that the computer program had detected that participants had concealed knowledge to the presented picture. This pre-defined loss feedback was realized in order to enhance the motivation of the participants to give their best in successfully concealing knowledge (Fang et al., 2003). The feedback was displayed for 500 ms on the screen (**Figure 1**). The inter-trial-interval (ITI) varied in a pseudo-random order between 1000, 1500, and 2000 ms. During ITI the screen remained black.

PROCEDURE

After arriving, participants gave written informed consent and were prepared for physiological recording. Participants were seated in a comfortable chair approximately 95 cm from the 20" computer TFT screen. The room was sound-attenuated and well-lit without dazzling the participants. Presentation V12.1 (Neurobehavioral Systems, Albany, NY, USA) was used to present the deception task. At the beginning of the task, participants learned the three pictures of the target category and the three pictures of the probe category for 5 min, whereas irrelevant pictures were not learned. Afterward participants performed 15 practice trials (including five probe, five irrelevant, and five target pictures). When responses to at least 12 pictures were correct, the main part of the deception task started. Otherwise the practice trials were repeated to make sure that participants were sufficiently familiar with the task. The deception task took on average 30 min (including learning and practice trials). The EEG was recorded during task performance. Each examination lasted about 1.5 h. At the end of the examination participants were thanked and paid depending on their performance (max. 15 EUR, about 20 USD).

EEG RECORDING

EEG recording, quantification, and analysis were conducted with reference to the guidelines for the study of human ERPs (Picton et al., 2000). The EEG was recorded using the ActiveTwo EEG system (BioSemi, Amsterdam, Netherlands) with 64 scalp active electrodes based on the extended 10/20 system (Jasper, 1958). The electrooculogram (EOG) was recorded from two horizontal electrodes placed beyond the epi canthi of both eyes and one



vertical electrode located approximately 1 cm below the right eye. As per BioSemi's design, the ground electrode during acquisition was formed by the Common Mode Sense active electrode and the Driven Right Leg passive electrode. All bioelectric signals were digitized on a laboratory computer using ActiView software (BioSemi). The impedances were below $30\text{ k}\Omega$ during EEG recording. The EEG was sampled at 512 Hz. Off-line analysis was performed by using EEGLab v9.0.0.2 (Delorme and Makeig, 2004) based on MATLAB 7.10.0 (The MathWorks). All data were band-pass filtered (0.3–30 Hz) and were re-referenced to averaged mastoids (cf. Soskins et al., 2004 for filter settings in P300 studies). ICA (an automated infomax decomposition) was applied to correct for ocular artifacts. Further technical and muscle artifacts were rejected when the EEG signal exceeded $\pm 85\text{ }\mu\text{V}$. Artifact-free epochs with instruction-conform responses were separately segmented for the three picture types (probe, target, and irrelevant). Participants included into statistical analysis of the P3 components had at least 20 artifact-free epochs of each picture type (irrelevant: $M = 40.65$, $SD = 9.99$, target: $M = 39.90$, $SD = 9.52$, probe: $M = 39.39$, $SD = 9.97$) and for the MFN (irrelevant: $M = 40.75$, $SD = 9.89$, target: $M = 37.73$, $SD = 7.37$, probe: $M = 39.16$, $SD = 9.55$). Grand averages of the picture-related ERPs (0–1000 ms, with a 100 ms pre-stimulus baseline) indicate an early P3 amplitude between 280 and 350 ms post-stimulus and a late P3 amplitude between 440 and 610 ms post-stimulus (Figure 2A) both with a parietal topography (Figures 3A,B). The MFN (with 0 ms indicating the occurrence of the response) was identified between 0 and 40 ms post-response in a time window –1100 ms pre-response to 500 ms post-response with –1100 to –1000 ms serving as an ERP-neutral baseline (Figure 2B) and demonstrated a frontal topography (Figure 3C). The ERP components of interest were quantified as baseline-to-peak amplitudes (i.e., using the most positive peak for the P3 and the most negative peak for the MFN in the respective time interval). To correct for the influence of the positive ERP that occurred prior to the MFN we subtracted the positive peak of this preceding pre-response ERP from the MFN peak for each picture type and each electrode position included into statistical analysis.

STATISTICAL ANALYSIS

Using SPSS 18.0, repeated-measures ANOVAs were performed for behavioral and ERP data (i.e., stimulus-locked P3 amplitude and response-locked MFN amplitude). Picture type (probe, target, and irrelevant) was applied as a repeated-measures factor in ANOVA for behavioral and ERP data. In addition, Region (i.e., frontal sites collapsed across F3, Fz, F4; central sites collapsed across C3, Cz, C4; parietal sites collapsed across P3, Pz, P4) was applied as a repeated-measures factor in the ANOVA of ERP data. Repeated-measures ANOVAs were conducted with Gender, SI-perpetrator, and Trait-BIS as between-subjects factors. Participants were split into three personality subgroups based on percentiles. Individuals with personality scores below and equal to the 33rd percentile were classified as individuals with low personality scores (Trait-BIS ≤ 2.6 : $N = 38$, SI-perpetrator ≤ 3.2 : $N = 41$). Individuals with personality scores above the 33rd percentile and below or equal to the 66th percentile were classified as individuals with medium personality scores (Trait-BIS > 2.6 and ≤ 3.1 : $N = 30$, SI-perpetrator > 3.2 and

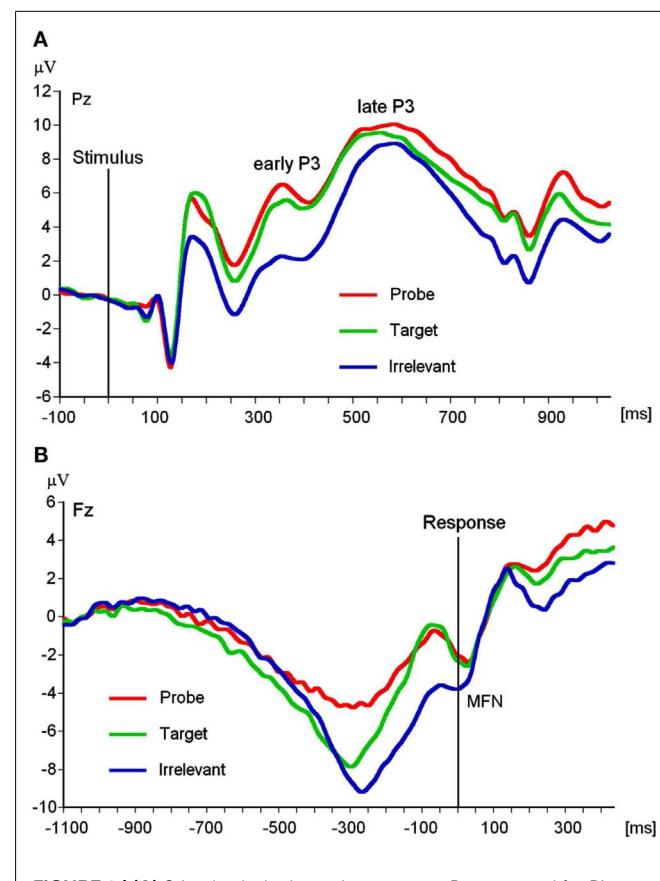


FIGURE 2 | (A) Stimulus-locked grand averages at Pz separated for Picture type ($N = 102$). **(B)** Response-locked grand averages at Fz separated for Picture type ($N = 91$).

≤ 3.9 : $N = 31$). Individuals with personality scores above the 66th percentile (Trait-BIS > 3.1 : $N = 34$, SI-perpetrator > 3.9 : $N = 30$) were classified as individuals with high scores.

Mean response times (RT) for the three picture categories were not normally distributed according to Kolmogorov-Smirnov test ($p < 0.10$). Therefore ln-transformed RT were applied to repeated-measures ANOVA (Wilkowski et al., 2010). The early and late P3 amplitudes (Kolmogorov-Smirnov test: $p = 0.32$ – 0.99) and the MFN-amplitudes (Kolmogorov-Smirnov test: $p = 0.60$ – 0.99) were normally distributed. For repeated-measures ANOVA, we report the uncorrected degrees of freedom along with Greenhouse-Geisser ϵ s that indicate the violation of the sphericity assumption in the repeated-measures design. In addition to the significance level we report effect size eta square (η^2). According to Cohen (1988) a small effect size is represented by an η^2 of about 0.010, a medium effect size is given for an η^2 of about 0.059, and a large effect size is represented by an η^2 of about 0.138. In the Section “Discussion” we focus on those results that are of a large effect size.

In the present study the effect of the within-subjects factor Picture type on ERP-amplitudes was analyzed together with the between-subjects factors Trait-BIS and SI-perpetrator. In repeated-measures ANOVA the interactions of the within-subjects and between-subjects factors can only be calculated

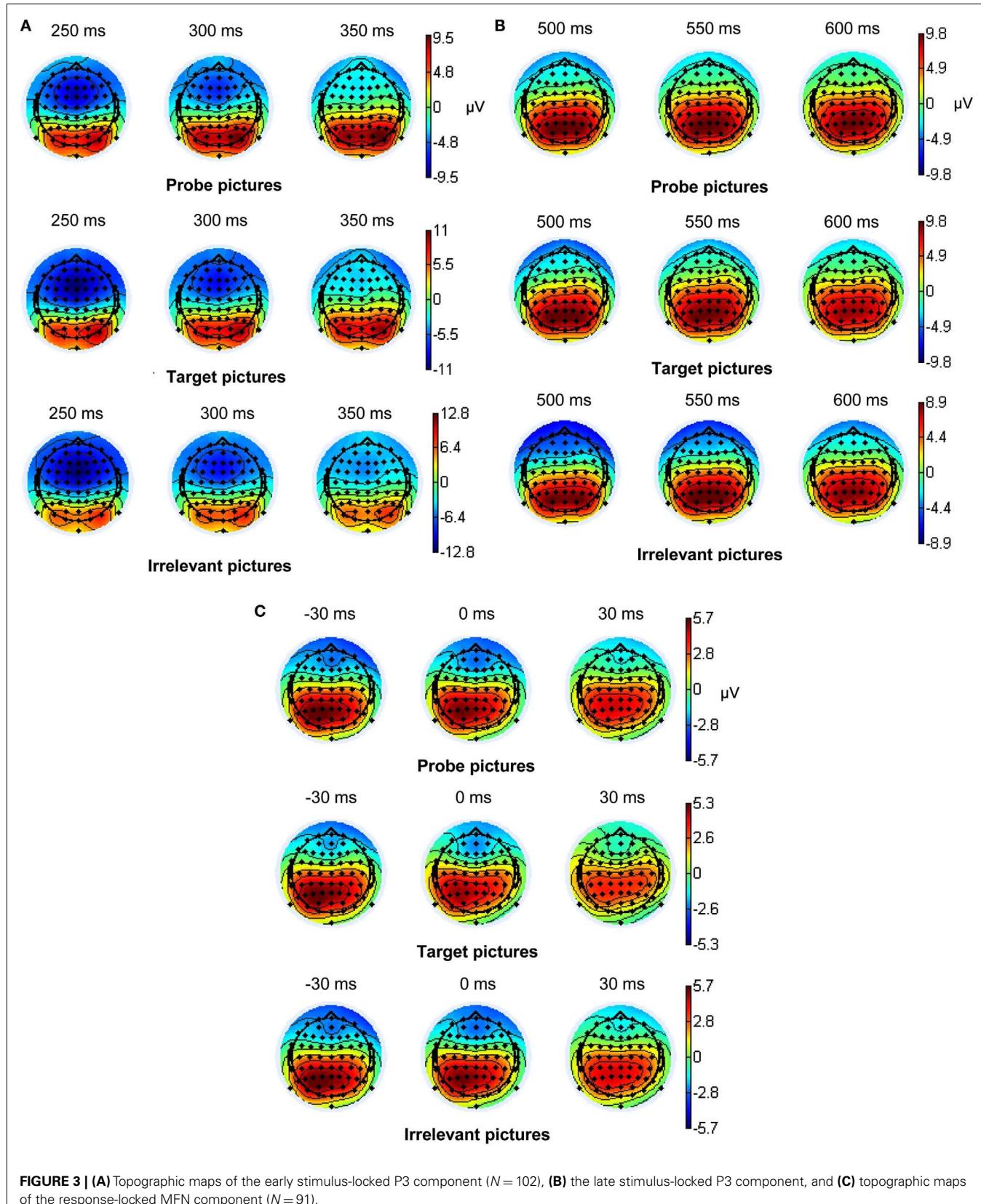


FIGURE 3 | (A) Topographic maps of the early stimulus-locked P3 component ($N = 102$), **(B)** the late stimulus-locked P3 component, and **(C)** topographic maps of the response-locked MFN component ($N = 91$).

and traced back in further analyses when Trait-BIS- and SI-perpetrator-groups are formed in order to represent the between-subjects factors. Thus, the individual differences are reduced to those aspects that can be represented by the group variables. Even when we already formed three groups for each trait, this does not account for the complete variability of individual differences. In order to overcome this limitation of the repeated-measures ANOVA different methods have been proposed. For example, mixed-model ANOVA allows for a more complete representation of individual differences. However, only relative fit indices (Akaike Information Criterion, Bayesian Information Criterion) are available for mixed-model ANOVA (Liu et al., 2012), which might be regarded as a limitation of this approach. Another approach that allows for a complete representation of individual differences together with the interesting experimental effects, are ‘fixed-links’ models, which have been introduced by Schweizer (2006, 2008) on the basis of latent-growth models (Chan, 1998; Muthén and Muthén, 2010) in the context of SEM. The major characteristic that the fixed-links model shares with conventional growth models in the context of SEM is that the loadings of the latent variables are fixed according to specific hypotheses and that the variances of the latent variables are estimated. In contrast to conventional growth models based on SEM, fixed-links models allow for modeling of treatment effects that do not necessarily represent a temporal order (Schweizer, 2008). The first advantage of these models is that both the absolute fit of the models (e.g., χ^2 -test) and the relative fit of the models can be determined. The second advantage of the fixed-links models is that estimation methods are available that allow for parameter estimation even when there is a violation of the multivariate normal distribution in the data (Satorra and Bentler, 1994). The third advantage of fixed-links models is that, besides the modeling of experimental effects, they allow for an evaluation of the measurement models for the dependent variables, because the dependent variables can be represented by latent variables. Here, the dependent variables were the ERP-amplitudes that were represented by latent variables so that the measurement models for ERP-amplitudes were also evaluated. The fixed-links model has been successfully applied in different analyses of cognitive tasks (e.g., Miller et al., 2010). Because of the above-mentioned advantages, fixed-links models were calculated with Mplus 6.1 (Muthén and Muthén, 2010) in order to represent the complete variability of individual differences together with the treatment effects. In addition to the χ^2 -test, the Root Mean Square Error of Approximation (RMSEA), the Comparative Fit Index (CFI), and the standardized root mean square residual (SRMR) were reported in order to evaluate model fit.

RESULTS

BEHAVIORAL DATA

A Picture type main effect was observed for the percentage of correct responses, $F(2,176) = 35.79, p < 0.01, \epsilon = 0.77, \eta^2 = 0.29$. Simple contrasts revealed that the percentage of correct responses was significantly lower to probe compared to irrelevant pictures, $F(1,88) = 10.92, p < 0.01, \eta^2 = 0.11$, and to target compared to irrelevant pictures, $F(1,88) = 48.52, p < 0.01, \eta^2 = 0.36$. The percentage of correct responses was significantly higher to

Table 1 | Descriptive statistics of the number of correct responses and mean response times depending on picture type.

Picture type	Percentage correct responses		Response times (ms)	
	M	SE	M	SE
Probe	97.65	0.31	730.35	24.85
Target	94.72	0.55	745.29	22.85
Irrelevant	99.50	0.13	690.23	23.01

N = 102. In purpose of simplicity response times are presented without In-transformation in the table.

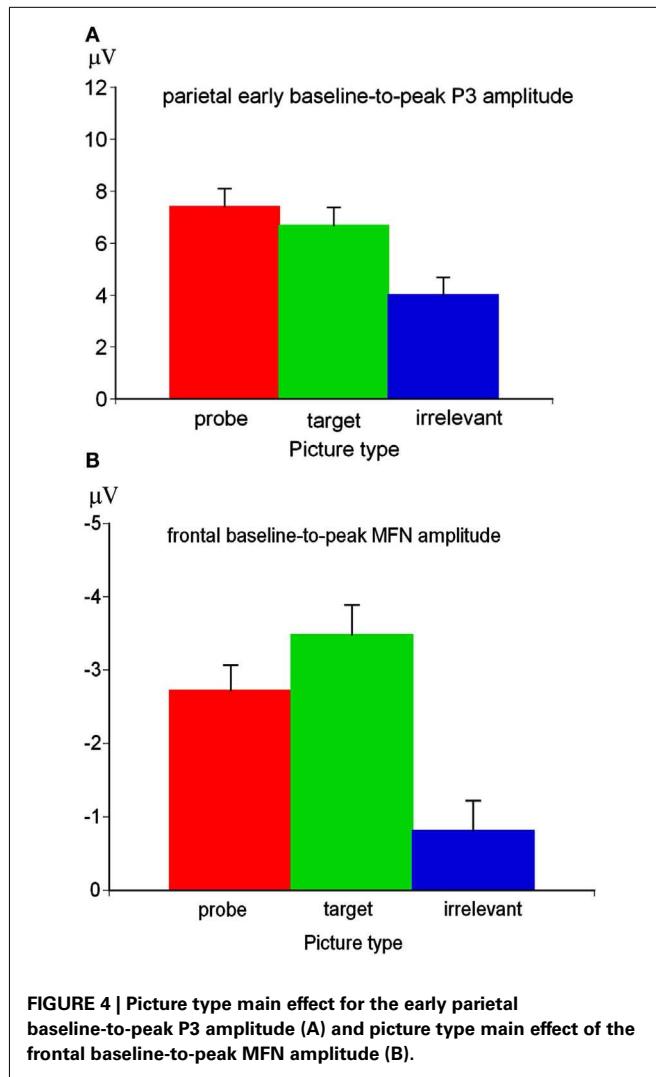
probe than to target pictures, $F(1,88) = 30.69, p < 0.01, \eta^2 = 0.26$ (**Table 1**).

Correct mean RT differed among Picture types, $F(2,176) = 26.30, p < 0.01, \epsilon = 0.93, \eta^2 = 0.23$. Simple contrasts revealed that RT were significantly longer for probe compared to irrelevant pictures, $F(1,88) = 19.79, p < 0.01, \eta^2 = 0.15$, and for target compared to irrelevant pictures, $F(1,88) = 58.69, p < 0.01, \eta^2 = 0.40$. RT to probe pictures were shorter than RTs to target pictures, $F(1,88) = 4.65, p < 0.05, \eta^2 = 0.05$ (**Table 1**). There was no main effect of Trait-BIS or SI-perpetrator and no interaction of Picture type \times SI-perpetrator or Picture type \times Trait-BIS for number of correct responses and RT.

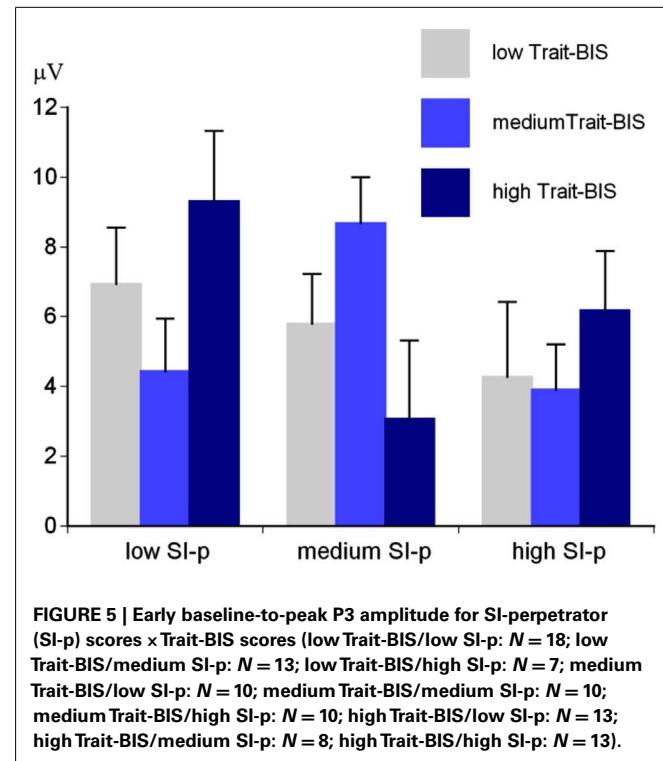
P3 AMPLITUDE

The Region main effect of the early P3 amplitude was significant, $F(2,176) = 127.57, p < 0.01, \epsilon = 0.60, \eta^2 = 0.59$. Simple contrasts revealed a more positive P3 amplitude at parietal sites ($M = 6.03 \mu\text{V}, SE = 0.65$) compared to central electrode sites ($M = -0.95 \mu\text{V}, SE = 0.82$), $F(1,88) = 142.34, p < 0.01, \eta^2 = 0.62$, and compared to frontal sites ($M = -3.60 \mu\text{V}, SE = 0.90$), $F(1,88) = 135.96, p < 0.01, \eta^2 = 0.61$. Since the Region main effect indicated the typical parietal P3 topography, further analyses have been conducted for the early parietal P3. At parietal sites, the Picture type main effect was significant for the P3 amplitude, $F(2,176) = 47.83, p < 0.01, \epsilon = 0.90, \eta^2 = 0.35$. Simple contrasts indicated that the P3 amplitude was more positive for probe compared to irrelevant pictures, $F(1,88) = 66.55, p < 0.01, \eta^2 = 0.43$, and for target compared to irrelevant pictures, $F(1,88) = 54.10, p < 0.01, \eta^2 = 0.38$. The early P3 amplitude was also more positive for probe compared to target pictures, $F(1,88) = 5.56, p < 0.05, \eta^2 = 0.06$ (**Figure 4A**).

Regarding personality, there was a significant SI-perpetrator \times Trait-BIS interaction for the early parietal P3 amplitude, $F(4,88) = 2.71, p < 0.05, \eta^2 = 0.11$. This interaction could be traced back to a significant SI-perpetrator main effect for individuals with medium Trait-BIS scores, $F(2,24) = 3.90, p < 0.05, \eta^2 = 0.25$ (**Figure 5**). Individuals with medium SI-perpetrator and medium Trait-BIS scores showed the more positive early parietal P3 amplitude ($M = 8.68 \mu\text{V}, SE = 1.31$) compared to individuals with low SI-perpetrator and medium Trait-BIS scores ($M = 4.44 \mu\text{V}, SE = 1.50$) and individuals with high SI-perpetrator and medium Trait-BIS scores ($M = 3.90 \mu\text{V}, SE = 1.31$).



For the late P3 amplitude, a significant Region main effect was observed, $F(2,176) = 122.62, p < 0.01, \epsilon = 0.60, \eta^2 = 0.58$. As for the early P3 amplitude, simple contrasts indicated a more positive late parietal P3 amplitude ($M = 9.80 \mu\text{V}, SE = 0.67$) compared to the central P3 amplitude ($M = 6.15 \mu\text{V}, SE = 0.71$), $F(1,88) = 71.22, p < 0.01, \eta^2 = 0.45$. Again, because of the parietal P3 topography, the Picture type main effect for the late P3 amplitude was analyzed at parietal sites, $F(2,176) = 4.31, p < 0.05, \epsilon = 0.93, \eta^2 = 0.05$. Simple contrasts suggested a more positive late P3 amplitude for probe pictures ($M = 10.24 \mu\text{V}, SD = 0.69$) compared to irrelevant pictures ($M = 9.17 \mu\text{V}, SD = 0.69$), $F(1,88) = 7.17, p < 0.01, \eta^2 = 0.08$, and for target pictures ($M = 10.00 \mu\text{V}, SD = 0.72$) compared to irrelevant pictures, $F(1,88) = 3.96, p = 0.05, \eta^2 = 0.04$. In contrast to the early parietal P3 amplitude, the late P3 amplitude of probe compared to target pictures did not substantially differ, $F(1,88) < 1, \text{ns}$. Also in contrast to the early P3, the SI-perpetrator \times Trait-BIS interaction was only marginally significant for the late parietal P3 amplitude, $F(4,88) = 2.32, p = 0.06, \eta^2 = 0.10$. As for the early parietal P3, individuals with medium SI-perpetrator and with medium



Trait-BIS scores showed the most positive late parietal P3 amplitude, $F(2,24) = 4.24, p < 0.05, \eta^2 = 0.26$. The Pearson correlations between the early P3 amplitude and the late P3 amplitude were 0.63 at Pz, 0.69 at P3, and 0.64 at P4 ($N = 102$, all $p < 0.01$, two-tailed). Thus, the parietal early and late P3 amplitudes were significantly correlated. It should also be noted that both the early P3 and the late P3 have a parietal topography so that the early P3 amplitude should probably not be regarded as a P3a or novelty P3, which is known to have a frontal topography (Kok, 2001).

MFN AMPLITUDE

The Region main effect of the MFN amplitude was significant, $F(2,154) = 12.94, p < 0.01, \epsilon = 0.67, \eta^2 = 0.14$. Simple contrasts indicated a more negative MFN amplitude at frontal sites ($M = -2.52 \mu\text{V}, SE = 0.35$) compared to central sites ($M = -1.04 \mu\text{V}, SE = 0.34$), $F(1,77) = 44.45, p < 0.01, \eta^2 = 0.37$, and at parietal sites ($M = -2.00 \mu\text{V}, SE = 0.31$) compared to central sites, $F(1,77) = 14.32, p < 0.01, \eta^2 = 0.16$, but not at frontal compared to parietal sites, $F(1,77) = 1.83, p = 0.18$. In order to investigate variations of cognitive control, further analyses focused on the frontal MFN. The Picture type main effect of the frontal MFN amplitude was significant, $F(2,148) = 27.14, p < 0.01, \epsilon = 0.94, \eta^2 = 0.27$. Simple contrasts indicated that the target-MFN was more negative than the probe-MFN, $F(1,74) = 6.02, p < 0.05, \eta^2 = 0.08$. The MFN amplitude was more negative for target pictures compared to irrelevant pictures, $F(1,74) = 41.83, p < 0.01, \eta^2 = 0.36$, and more importantly for probe compared to irrelevant pictures, $F(1,74) = 28.37, p < 0.01, \eta^2 = 0.28$ (Figure 4B). The Picture type \times SI-perpetrator interaction, $F(4,148) < 1, \text{ns}$, the Picture type \times Trait-BIS interaction,

$F(4,148) < 1$, ns, and the Picture type \times SI-perpetrator \times Trait-BIS interaction, $F(8,148) = 1.39$, ns, were not significant.

FIXED-LINKS MODELING

The first fixed-links model comprised latent variables representing the early P3 amplitudes for each Picture Type (irrelevant, target, and probe) at three relevant electrode sites (P3, Pz, and P4; see **Figure 6**). Residuals were allowed to correlate for electrode sites P3 and P4 indicating common variance of these electrode positions. Measurement invariance was specified by holding the means and factor loadings of the factor indicators equal across picture types. The intercept and the linear slope for the increase of the latent variables representing P3 amplitudes from irrelevant, to target, and probe pictures was calculated. The slope represents the effects of Picture type on P3 amplitudes and was predicted by Trait-BIS, SI-perpetrator, and the Trait-BIS \times SI-perpetrator interaction. Since the multivariate normal distribution was not given for the variables included into the model $\chi^2_{(2)} = 181.42$; $p < 0.01$ the robust maximum-likelihood estimation was performed and the Satorra–Bentler scaled χ^2_{SB} statistic (Satorra and Bentler, 1994) was reported. The model fits

quite well to the data $\chi^2_{\text{SB}(54)} = 66.30$; $p = 0.12$; RMSEA = 0.052; CFI = 0.99; SRMR = 0.060. Trait-BIS and SI-perpetrator were significant positive predictors of the slope of the early P3 amplitudes (see **Figure 6**). The positive predictions of the slope indicate that the increase of the early P3 amplitude from irrelevants over targets to probes is more substantial for individuals with higher Trait-BIS scores as well as for individuals with higher SI-perpetrator scores.

The second fixed-links model comprised latent variables representing the late P3 amplitudes for each Picture type (irrelevant, target, and probe) at three electrode sites (P3, Pz, and P4). The model was specified like the previous model so that an additional figure would have been redundant. Since the variables included deviate from the multivariate normal distribution $\chi^2_{(2)} = 275.75$; $p < 0.01$ robust maximum-likelihood estimation was performed and the Satorra–Bentler scaled χ^2_{SB} statistic was reported. The model fits very well to the data $\chi^2_{\text{SB}(53)} = 61.99$; $p = 0.98$; RMSEA = 0.045; CFI = 0.99; SRMR = 0.038. However, there were no effects for personality on the late P3 amplitudes.

The third fixed-links model comprised latent variables representing the MFN-amplitudes for each Picture Type (irrelevant,

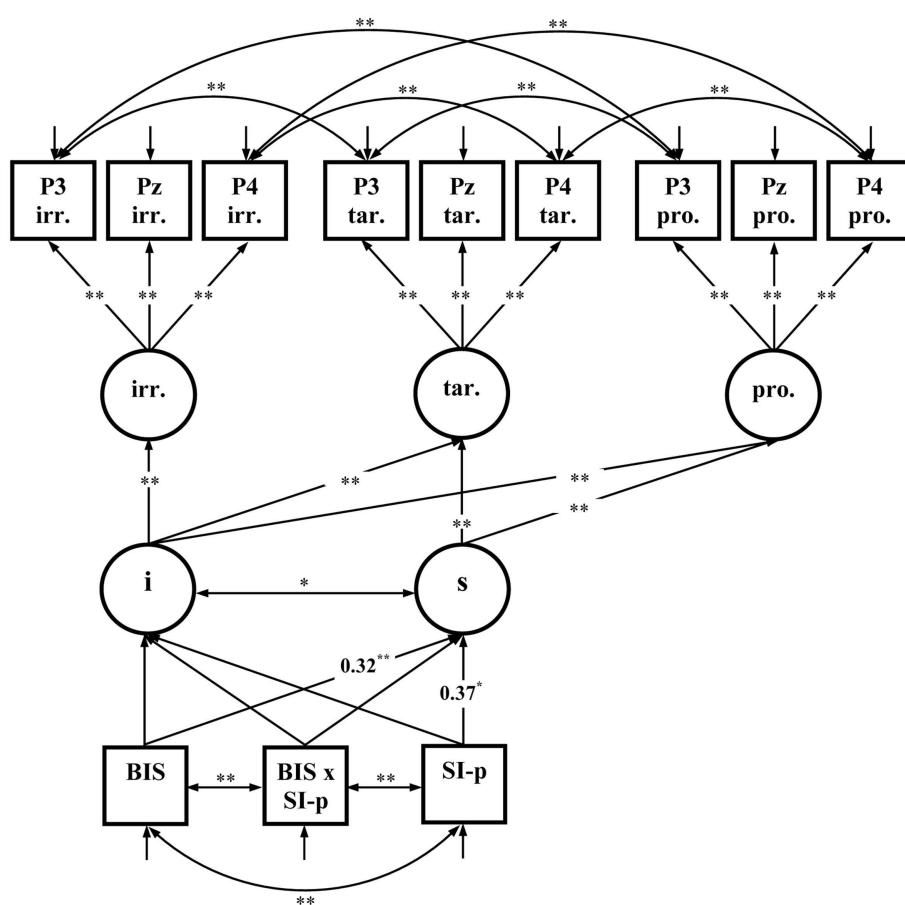


FIGURE 6 | Fixed-links model for early P3 amplitudes across picture types ($N = 85$); i, intercept; s, slope; irr., irrelevant pictures; tar., target pictures; pro., probe pictures; significant coefficients are marked with ** ($p \leq 0.05$, two-tailed) and *** ($p \leq 0.01$, two-tailed). For

convenience, numerical values of the completely standardized solution are only given for significant coefficients related to Trait-BIS (BIS), SI-perpetrator (SI-p), and the Trait-BIS \times SI-perpetrator interaction (BIS \times SI-p).

target, and probe) at three relevant electrode sites (F3, Fz, F4). Residuals were allowed to correlate for electrode sites indicating common variance due to electrode positions. Measurement invariance was specified by holding the means and factor loadings of the factor indicators equal across picture types. The intercept and the linear slope for the decrease of the latent variables representing MFN-amplitudes from irrelevant, to target, and probe pictures was calculated. The slope represents the effects of Picture type on MFN-amplitudes. Again, the multivariate normal distribution was not given for the variables included into the model $\chi^2_{(2)} = 218.02$; $p < 0.01$ so that robust maximum-likelihood estimation was performed and the Satorra–Bentler scaled χ^2_{SB} statistic was reported. The model fits well to the data $\chi^2_{\text{SB}(52)} = 61.39$; $p = 0.17$; RMSEA = 0.046; CFI = 0.99; SRMR = 0.073. There were significant negative path coefficients from Trait-BIS and SI-perpetrator to the MFN-slope indicating that individuals with higher Trait-BIS and SI-perpetrator scores have a more pronounced decrease of MFN-amplitudes from irrelevant, to target, and probe pictures (see Figure 7). Moreover, there is a significant positive path coefficient from the Trait-BIS \times SI-perpetrator interaction to the MFN-slope indicating that individuals with both higher Trait-BIS

and higher SI-perpetrator scores had a less pronounced decrease of MFN-amplitudes.

DISCUSSION

The present study investigated individual differences of Trait-BIS and SI-perpetrator with regard to stimulus salience, attentional control, and cognitive control in a visual deception task. Stimulus salience and attentional control were investigated by means of the stimulus-locked P3 amplitude and cognitive control was investigated by means of the response-locked MFN. The main ERP findings with strong effect sizes are: (a) According to ANOVA and fixed-links modeling, the parietal P3 amplitudes were more positive to probe and target pictures compared to irrelevant pictures. (b) Fixed-links modeling results indicated that higher Trait-BIS as well as higher SI-perpetrator scores were related to a more pronounced early P3 increase from irrelevant to target and to probe pictures. (c) The response-locked frontal MFN amplitude was more negative for probe and target compared to irrelevant pictures. (d) Fixed-links modeling demonstrated that higher Trait-BIS scores as well as higher SI-perpetrator scores predicted a more pronounced MFN decrease from irrelevant to probe pictures.

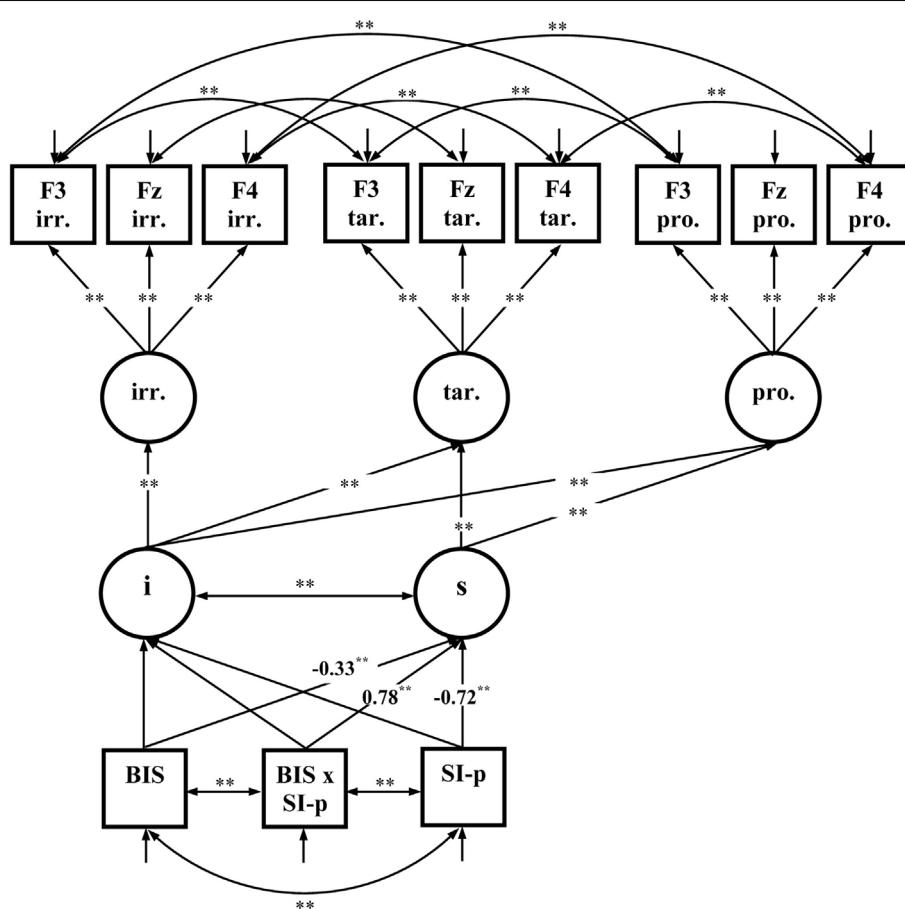


FIGURE 7 | Fixed-links model for MFN-amplitudes across picture types ($N = 85$). i , intercept; s , slope; irr , irrelevant pictures; tar , target pictures; pro , probe pictures; significant coefficients are marked with ** ($p \leq 0.05$, two-tailed) and *** ($p \leq 0.01$, two-tailed). For

convenience, numerical values of the completely standardized solution are only given for significant coefficients related to Trait-BIS (BIS), SI-perpetrator (SI-p), and the Trait-BIS \times SI-perpetrator interaction (BIS \times SI-p).

However, the Trait-BIS \times SI-perpetrator interaction in the fixed-links model suggested a smaller MFN decrease from irrelevant to probe pictures for individuals with both higher Trait-BIS and higher SI-perpetrator scores. We discuss the implications of these main findings subsequently.

VARIATIONS OF STIMULUS SALIENCE AND ATTENTIONAL CONTROL

Our P3 results in a reinforcement-related deception task support findings of prior deception studies showing more pronounced parietal P3 amplitudes to probe compared to irrelevant pictures (e.g., Mertens and Allen, 2008; Ambach et al., 2010). From the perspective of the salience hypothesis (Kok, 2001), the present findings suggest that irrelevant pictures are less salient (resulting in smaller P3 amplitudes) than probe and target pictures. In this line and in accordance with prior studies longer RT were observed for probe stimuli compared to irrelevant stimuli (Walczak et al., 2003; Dong et al., 2010b). This suggests that participants are more sensitive and subsequently more cautious in responding to probe pictures compared to irrelevant pictures supporting the salience hypothesis. Moreover, based on the attentional control approach it cannot be excluded that RT to probes were slower because more attentional and/or processing resources were needed to inhibit the primary task of responding truthfully (e.g., Johnson et al., 2003). As a new finding we could demonstrate by means of fixed-links modeling that deceiving knowledge is more salient for higher vs. lower Trait-BIS individuals and also for higher vs. lower SI-perpetrator individuals because the early P3 amplitude increased from irrelevant to probe pictures for both personality dimensions (**Figure 6**). Our results suggest that deceiving knowledge is more salient (resulting in a larger probe-P3) for those individuals who show an increased sensitivity to aversiveness (higher Trait-BIS) and those individuals who are more sensitive toward situations in that they treat others unfairly (higher SI-perpetrator).

VARIATIONS OF COGNITIVE CONTROL

In our study the variations of the MFN illustrate that probe compared to target and irrelevant pictures require more cognitive control. Our MFN findings correspond to prior studies in that the probe-MFN was more negative than the irrelevant-MFN (Dong et al., 2010a). Because this finding of a more negative probe-MFN compared to irrelevant-MFN parallels to MFN findings in non-deception studies illustrating a more negative MFN to erroneous compared to correct responses (Luu et al., 2000; Potts et al., 2006), one might conclude that erroneous as well as deceptive responses are more aversive and this might also contribute to an increase in cognitive control. Moreover, the decrease of the MFN from irrelevant to target and to probe pictures was more pronounced in higher vs. lower Trait-BIS individuals and in higher vs. lower SI-perpetrator individuals. This finding illustrates that individuals who have either higher Trait-BIS scores or higher SI-perpetrator scores invest more cognitive control in their responses to probe items. Since both trait-dimensions were positively correlated we presume that they share variance in aversiveness sensitivity. Therefore, we conclude with regard to the revised reinforcement sensitivity theory (Corr, 2008) that deceiving knowledge is not only more salient (see P3 findings) but

also evokes a more pronounced investment of cognitive control (see MFN findings). The more pronounced MFN of higher vs. lower Trait-BIS individuals corresponds to ERN findings in non-deception studies (e.g., Boksem et al., 2006) with higher Trait-BIS individuals showing more negative ERN amplitudes. This indicates that erroneous and deceptive responses share cognitive processes that are activated in the Anterior Cingulate Cortex (cf. Johnson et al., 2004, 2008). Moreover, our MFN findings suggest that a combination of higher SI-perpetrator and higher Trait-BIS scores reduces the amount of cognitive control invested to probes. This could be due to the fact that resources for response-related control might be still occupied by moral justification in these individuals. Overall, the results indicate that salience processes (P3) as well as cognitive control processes (MFN) co-occur in a deception task.

LIMITATIONS AND FUTURE DIRECTIONS

Dipole modeling in a deception setting (cf. Johnson et al., 2008) might be promising to further investigate the functioning of the fronto-parietal network during executive control that has been described in imaging studies (e.g., Christ et al., 2009). Moreover, our data suggest that both stimulus- and response-locked ERPs are promising in order to differentiate deceptive vs. truthful knowledge. Therefore, future research could clarify whether the combination of different ERPs contributes to more correct classifications of truthful vs. deceptive knowledge in guilty compared to innocent persons. Recent findings demonstrate that enhanced emotional arousal assessed by heart rate changes from baseline to experimental task was observed after committing a mock crime in the context of a CIT. Moreover, enhanced emotional arousal reduced memory of peripheral information in the CIT (Peth et al., 2012). Since individual differences like trait-anxiety or trait-BIS have been associated with an increased arousal (e.g., Gray and McNaughton, 2000) it might be interesting to investigate individual differences of trait-anxiety or trait-BIS with our deception task under different arousal conditions. Moreover, individual differences of Trait-BIS and SI-perpetrator predicted variations of the P3 and the MFN so that both trait-dimensions appear to be promising moderators for the classification of guilty vs. innocent individuals in CIT. By using 3 different probes, 20 different irrelevants, and 3 different target pictures we realized the traditional stimulus ratio applied in prior deception studies (e.g., Meijer et al., 2007). However, since each picture type occurred with the same total frequency, it remains for further clarification whether this has an effect on the P3 and MFN findings. It remains also for replication whether aspects of stimulus salience and attentional control can be related to different P3 components.

It should be noted that the effects of personality were found for the early P3 amplitude but not for the later P3 amplitude. At this point we can only speculate on the reasons for this result. One possibility could be that the more early P3 amplitude reflects a more spontaneous and therefore a more affective aspect of stimulus processing, whereas the later P3 amplitude is related to subsequent, more cognitive processes. Finally, despite applying a 0.3 Hz high-pass filter the late P3 component does not entirely return to the baseline level 1 s after stimulus-onset (for similar observations see

Fang et al., 2003; Ambach et al., 2010; Gamer and Berti, 2010). At this point of research we cannot exclude whether variable ITIs or variations in sampling rate might account for this phenomenon. According to Soskins et al. (2004) we can also not completely exclude that the negative waveform following the late P3 could be a distorted post-peak recovery of P3.

Based on the present findings we draw the following conclusions: First, parietal P3 and frontal MFN are ERPs that are related to an intensification of stimulus salience and cognitive control in a deception task. P3 and MFN became more pronounced from irrelevant to target and to probe stimuli. Second, Trait-BIS and SI-perpetrator modulate the intensity of stimulus

salience (early P3) and cognitive control (MFN) in a deception task, whereas behavioral parameters were not sensitive to personality differences. Third, our data encourage the simultaneous investigation of stimulus-locked and response-locked ERPs to further elucidate patterns of neuro-cognitive processes during deception.

ACKNOWLEDGMENTS

This research was supported by the Deutsche Forschungsgemeinschaft (DFG) to the first and the third authors (grant no. LE 2240/2-1). The authors wish to thank Anja Ritter, Leon Sautier, and Anne-Kristin Wissel for their assistance during data collection.

REFERENCES

- Allen, J. J., Iacono, W. G., and Danielson, K. D. (1992). The identification of concealed memories using the event-related potential and implicit behavioral measures: a methodologic for prediction in the face of individual differences. *Psychophysiology* 29, 504–522.
- Ambach, W., Stark, R., Peper, M., and Vaitl, D. (2010). An interfering go/no-go task does not affect accuracy in a concealed information test. *Int. J. Psychophysiol.* 75, 258–267.
- Amadio, D. M., Master, S. L., Yee, C. M., and Taylor, S. E. (2008). Neurocognitive components of the behavioral inhibition and activation systems: implications for theories of self-regulation. *Psychophysiology* 45, 11–19.
- Beauducel, A., Brocke, B., and Leue, A. (2006). Energetical bases of extraversion: effort, arousal, and vigilance performance. *Int. J. Psychophysiol.* 62, 212–223.
- Boksem, M. A. S., and De Cremer, D. (2010). Fairness concerns predict medial frontal negativity amplitude in ultimatum bargaining. *Soc. Neurosci.* 5, 118–128.
- Boksem, M. A. S., Tops, M., Wester, A. E., Meijam, T. F., and Lorist, M. M. (2006). Error-related ERP components and individual differences in punishment and reward sensitivity. *Brain Res.* 1101, 92–101.
- Bradley, M. M., and Lang, P. J. (2007). “The international affective pictures system (IAPS) in the study of emotion and attention,” in *Handbook of Emotion Elicitation and Assessment*, eds J. A. Coan and J. B. Allen (New York: Oxford University Press), 29–46.
- Carrión, R. E., Keenan, J. P., and Sebanz, N. (2010). A truth that's told with bad intent: an ERP study of deception. *Cognition* 114, 105–110.
- Carver, C. S., and White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to incentives. *Neurosci. Lett.* 479, 1–5.
- Chan, D. (1998). The conceptualization and analysis of change over time: an integrative approach incorporating longitudinal mean and covariance structures analysis (LMAC) and multiple indicator latent growth modeling (MLGM). *Organ. Res. Methods* 1, 421–483.
- Christ, S. E., Van Essen, D. C., Watson, J. M., Brubaker, L. E., and McDermott, K. B. (2009). The contributions of prefrontal cortex and executive control to deception: evidence from activation likelihood estimate metaanalyses. *Cereb. Cortex* 19, 1557–1566.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Corr, P. J. (2008). *The Reinforcement Sensitivity Theory*. Cambridge: Cambridge University Press.
- Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *Am. Psychol.* 30, 116–127.
- Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis (sccn.uscd.edu/eeglab/). *J. Neurosci. Methods* 134, 9–21.
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., and Cooper, H. (2003). Cues to deception. *Psychol. Bull.* 129, 74–118.
- Dong, G., Hu, Y., Lu, Q., and Wu, H. (2010a). The presentation order of cue and target matters in deception study. *Behav. Brain Funct.* 6, 1–9.
- Dong, G., Wu, H., and Lu, Q. (2010b). Attempting to hide our real thoughts: electrophysiological evidence from truthful and deceptive responses during evaluation. *Neurosci. Lett.* 479, 1–5.
- Dong, G., Hu, Y., and Wu, H. (2011). The presentation order of cue and target matters in deception study. *Behav. Brain Funct.* 7, 36.
- Fang, F., Liu, Y., and Shen, Z. (2003). Lie detection with contingent negative variation. *Int. J. Psychophysiol.* 50, 247–255.
- Farwell, L. A., and Donchin, E. (1991). The truth will out: interrogative polygraphy (“Lie detection”) with event-related brain potentials. *Psychophysiology* 28, 531–547.
- Fink, A., and Neubauer, A. C. (2004). Extraversion and cortical activation: effects of task complexity. *Pers. Individ. Dif.* 36, 333–347.
- Gamer, M., and Berti, S. (2010). Task relevance and recognition of concealed information have different influences on electrodermal activity and event-related brain potentials. *Psychophysiology* 47, 355–364.
- Gombos, V. A. (2006). The cognition of deception: the role of executive processes in producing lies. *Genet. Soc. Gen. Psychol. Monogr.* 132, 197–214.
- Gray, J. A., and McNaughton, N. (2000). *The Neuropsychology of Anxiety*. Oxford: University Press.
- Jasper, H. H. (1958). The twenty electrode system of the International Federation. *Electroencephalogr. Clin. Neurophysiol.* 10, 371–375.
- Johnson, R. Jr., Barnhardt, J., and Zhu, J. (2003). The deceptive response: effects of response conflict and strategic monitoring on the late positive component and episodic memory-related brain activity. *Biol. Psychol.* 64, 217–253.
- Johnson, R. Jr., Barnhardt, J., and Zhu, J. (2004). The contribution of executive processes to deceptive responding. *Neuropsychologia* 42, 878–901.
- Johnson, R. Jr., Henkell, H., Simon, E., and Zhu, J. (2008). The self in conflict: the role of executive processes during truthful and deceptive responses about attitudes. *Neuroimage* 39, 469–482.
- Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology* 38, 557–577.
- Lange, S., Leue, A., and Beauducel, A. (2012). Behavioral approach and reward processing: results on feedback-related negativity and P3 component. *Biol. Psychol.* 89, 416–425.
- Leue, A., and Beauducel, A. (2008). A meta-analysis of reinforcement sensitivity theory: on performance parameters in reinforcement tasks. *Pers. Soc. Psychol. Rev.* 12, 353–369.
- Leue, A., Chavanon, M.-L., Wacker, J., and Stemmler, G. (2009). On the differentiation of N2 components in an appetitive choice task: evidence for the revised Reinforcement Sensitivity Theory. *Psychophysiology* 46, 1244–1257.
- Leue, A., Lange, S., and Beauducel, A. (2012a). Modulation of the conflict monitoring intensity: the role of aversive reinforcement, cognitive demand, and trait-BIS. *Cogn. Affect. Behav. Neurosci.* 12, 287–307.
- Leue, A., Lange, S., and Beauducel, A. (2012b). Reinforcement sensitivity and conflict processing: a study of principal components in the N2 time domain. *J. Individ. Dif.* 33, 160–168.
- Liu, S., Rovine, M. J., and Molenaar, P. C. M. (2012). Selecting a linear mixed model for longitudinal data: repeated measures analysis of variance, covariance pattern model, and growth curve approaches. *Psychol. Methods* 17, 15–30.
- Luu, P., Flaisch, T., and Tucker, D. M. (2000). Medial frontal cortex in action monitoring. *J. Neurosci.* 20, 464–469.
- Mecklinger, A., Kramer, A. F., and Strayer, D. L. (1992). Event related potentials and EEG components in a semantic memory search task. *Psychophysiology* 29, 104–119.

- Meijer, E. H., Smulders, F. T. Y., Merckelbach, H. L. G. J., and Wolf, A. G. (2007). The P300 is sensitive to concealed face recognition. *Int. J. Psychophysiol.* 66, 231–237.
- Meijer, E. H., Smulders, F. T. Y., and Wolf, A. G. (2009). The contribution of mere recognition to the P300 effect in a concealed information test. *Appl. Psychophysiol. Biofeedback* 34, 221–226.
- Meixner, J. B., and Rosenfeld, J. P. (2011). A mock terrorism application of the P300-based concealed information test. *Psychophysiology* 48, 149–154.
- Mertens, R., and Allen, J. B. (2008). The role of psychophysiology in forensic assessments: deception detection, ERPs, and virtual reality mock crime scenarios. *Psychophysiology* 45, 286–298.
- Miller, R., Rammsayer, T. H., Schweizer, K., and Troche, S. J. (2010). Decay of iconic memory traces is related to psychometric intelligence: a fixed-links modeling approach. *Learn. Individ. Differ.* 20, 699–704.
- Mulder, G. (1986). “The concept and measurement of mental effort,” in *Ergonomics and Human Information Processing*, eds R. J. Hockey, A. W. K. Gaillard, and M. G. H. Coles (Dordrecht: Martinus Nijhoff), 175–198.
- Muthén, B. O. (2004). “Latent variable analysis: growth mixture modeling and related techniques for longitudinal data,” in *Handbook of Quantitative Methodology for the Social Sciences*, ed. D. Kaplan (Newbury Park, CA: Sage Publications), 345–368.
- Muthén, L. K., and Muthén, B. O. (2010). *MPlus User’s Guide*. Los Angeles, CA: Muthén & Muthén.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113.
- Peth, J., Vossel, G., and Gamer, M. (2012). Emotional arousal modulates the encoding of crime-related details and corresponding physiological responses in the concealed information test. *Psychophysiology* 49, 381–390.
- Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., et al. (2000). Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology* 37, 127–152.
- Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* 118, 2128–2148.
- Potts, G., Martin, L., Burton, P., and Montague, P. (2006). When things are better or worse than expected: medial frontal cortex and the allocation of processing resources. *J. Cogn. Neurosci.* 18, 1–8.
- Raudenbush, S. W., and Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks: CA: Sage.
- Rosenfeld, J. P., Cantwell, G., Nasman, V. T., Wojdac, V., Ivanov, S., and Mazzeri, L. (1988). A modified event-related potential-based guilty knowledge test. *Int. J. Neurosci.* 42, 157–161.
- Rosenfeld, J. P., Angell, A., Johnson, M. M., and Qian, J. H. (1991). An ERP-based, control-question lie detector analog: algorithms for discriminating effects within individuals’ average waveforms. *Psychophysiology* 28, 319–335.
- Satorra, A., and Bentler, P. M. (1994). “Corrections to test statistics and standard errors in covariance structure analysis,” in *Latent Variables Analysis: Applications for Developmental Research*, eds A. Von Eye and C. C. Clogg (Thousand Oaks, CA: Sage), 399–419.
- Schmitt, M., Gollwitzer, M., Maes, J., and Arbach, D. (2005). Justice sensitivity: assessment and location in the personality space. *Eur. J. Psychol. Assess.* 21, 202–211.
- Schweizer, K. (2006). The fixed-links model in combination with the polynomial function as a tool for investigating choice reaction time data. *Struct. Equ. Modeling* 13, 403–419.
- Schweizer, K. (2008). Investigating experimental effects within the framework of structural equation modeling: an example with effects on both error scores and reaction times. *Struct. Equ. Modeling* 15, 327–345.
- Soskins, M., Rosenfeld, J. P., and Niedam, T. (2004). Peak-to-peak measurement of P300 recorded at 0.3 Hz high pass filter settings in intraindividual diagnosis: complex vs. simple paradigms. *Int. J. Psychophysiol.* 40, 173–180.
- Stenberg, G. (1994). Extraversion and the P300 in a visual classification task. *Pers. Individ. Dif.* 16, 543–560.
- Strobel, A., Beauducel, A., Debener, S., and Brocke, B. (2001). Psychometrische und strukturelle Merkmale einer deutschsprachigen Version des BIS/BAS-Fragebogens von Carver und White. *Z Diff. Diag. Psychol.* 22, 216–227.
- Wacker, J., Chavanon, M.-L., Leue, A., and Stemmler, G. (2010). Trait BIS predicts alpha asymmetry and P300 in a go/no-go task. *Eur. J. Pers.* 24, 85–105.
- Walczak, J. J., Roper, K. S., Seemann, E., and Humphrey, A. M. (2003). Cognitive mechanisms underlying lying to questions: response time as a cue to deception. *Appl. Cogn. Psychol.* 17, 755–774.
- Wilkowski, B. M., Robinson, M. D., and Troop-Gordon, W. (2010). How does cognitive control reduce anger and aggression? The role of conflict monitoring and forgiveness processes. *J. Pers. Soc. Psychol.* 98, 830–840.
- Zuckerman, M., DePaulo, B. M., and Rosenthal, R. (1981). “Verbal and nonverbal communication of deception,” in *Advances in Experimental Social Psychology*, ed. L. Berkowitz (New York: Academic Press), 1–59.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 08 July 2012; **accepted:** 02 December 2012; **published online:** 20 December 2012.

Citation: Leue A, Lange S and Beauducel A (2012) “Have you ever seen this face?” – individual differences and event-related potentials during deception. *Front. Psychology* 3:570. doi: 10.3389/fpsyg.2012.00570

This article was submitted to Frontiers in Cognitive Science, a specialty of Frontiers in Psychology.

Copyright © 2012 Leue, Lange and Beauducel. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.



Face and voice as social stimuli enhance differential physiological responding in a Concealed Information Test

Wolfgang Ambach^{1*}, Birthe Assmann², Bennet Krieg¹ and Dieter Vaitl^{1,3}

¹ Institute for Frontier Areas of Psychology and Mental Health, Freiburg, Germany

² Institute of Biology, Freie Universität Berlin, Berlin, Germany

³ Bender Institute of Neuroimaging, University of Giessen, Giessen, Germany

Edited by:

Matthias Gamer, University Medical Center Hamburg-Eppendorf, Germany

Reviewed by:

Eitan Elaad, Ariel University Center, Israel

Nurit Gronau, The Open University of Israel, Israel

***Correspondence:**

Wolfgang Ambach, Institute for Frontier Areas of Psychology and Mental Health, Wilhelmstr. 3A, Freiburg D-79098, Germany.
e-mail: ambach@igpp.de

Attentional, intentional, and motivational factors are known to influence the physiological responses in a Concealed Information Test (CIT). Although concealing information is essentially a social action closely related to motivation, CIT studies typically rely on testing participants in an environment lacking of social stimuli: subjects interact with a computer while sitting alone in an experimental room. To address this gap, we examined the influence of social stimuli on the physiological responses in a CIT. Seventy-one participants underwent a mock-crime experiment with a modified CIT. In a between-subjects design, subjects were either questioned acoustically by a pre-recorded male voice presented together with a virtual male experimenter's uniform face or by a text field on the screen, which displayed the question devoid of face and voice. Electrodermal activity (EDA), respiration line length (RLL), phasic heart rate (pHR), and finger pulse waveform length (FPWL) were registered. The Psychopathic Personality Inventory – Revised (PPI-R) was administered in addition. The differential responses of RLL, pHR, and FPWL to *probe* vs. *irrelevant* items were greater in the condition with social stimuli than in the text condition; interestingly, the differential responses of EDA did not differ between conditions. No modulatory influence of the PPI-R sum or subscale scores was found. The results emphasize the relevance of social aspects in the process of concealing information and in its detection. Attentional demands as well as the participants' motivation to avoid detection might be the important links between social stimuli and physiological responses in the CIT.

Keywords: Concealed Information Test, deception, mock-crime, social stimuli

INTRODUCTION

THE CONCEALED INFORMATION TEST

Concealing information from an interrogator is a specific social behavior commonly performed by a culprit in order to hide his or her involvement in a criminal act. A scientific psychophysiological method to detect intentionally hidden information is the Concealed Information Test (CIT), which combines a systematic interrogation with a simultaneous measurement of several physiological data channels. The core assumption of the CIT is that a guilty subject's physiological responses are different for crime-related information compared to crime-irrelevant information (Lykken, 1959). The CIT consists of several multiple-choice questions each referring to another detail of the crime under investigation. Typically, there are four to five answer alternatives to each question but only one alternative, the "probe," refers to the critical detail. For example, if an envelope was stolen out of an office, a typical CIT question could be "An office requisite has been stolen. Is this the stolen object?"; this question is combined with a sequence of five pictures representing the respective answer alternatives, e.g., a picture of (a) a pencil sharpener, (b) an envelope, (c) a highlighter, (d) a stapler, and (e) a Scotch®Tape. In this example, the picture of the envelope (b) is the "probe" item; the other items are referred to as "irrelevant." It is assumed that only subjects possessing crime-related knowledge ("guilty" subjects) will recognize

the correct item and show a different physiological response to it. Subjects without such knowledge ("innocents") cannot discriminate between the *probe* and *irrelevant* alternatives and therefore will not show a systematic response pattern. Numerous laboratory studies have shown that the CIT is a highly valid test for differentiating between guilty and innocent subjects (for a review see Ben-Shakhar and Elaad, 2003).

Concealed Information Test theory is heavily based on cognitive approaches, particularly the orienting response (Sokolov, 1963; Lykken, 1974). While motivational and emotional influences are thought to play a minor, only mediating role in laboratory CIT experiments, their importance might well be enhanced in field examinations (Verschueren and Ben-Shakhar, 2011). So far, the qualitative and quantitative differences in attentional, intentional, motivational, emotional, and social factors influencing the CIT in laboratory and field situations are only barely understood. CIT mechanisms that go beyond the orienting reflex merit more attention; the relation between social situation and physiological responding in the CIT still has to be elaborated.

SOCIAL ASPECTS AND THE CIT

Within the last decades, the social aspects of concealing information have played only a minor role in CIT research. As a predominant trend occurring in parallel, the participants in laboratory CIT

experiments were mostly seated alone in an experimental chamber and the former “interrogator” was replaced by an interrogative computer interface to interact with. The availability of computerized experimental methods supported this change in CIT research. By minimizing uncontrolled social influences, particularly by the experimenter (Iacono, 2000), it became possible to standardize CIT experiments to a certain degree. Yet, in the course of this trend, the social aspects of withholding information have faded into the background, although information concealment is essentially a social action.

Earlier studies focused on the social influence on physiological responding in the CIT questioning situation (Orne, 1975; Waid and Orne, 1981; see also Iacono, 2000). Yet, neither social interactions, nor social roles, nor the presence of social stimuli were systematically varied in these studies.

The differential responses to *probe vs. irrelevant* items in a CIT are known to be influenced by attentional, intentional, and motivational factors: a greater motivation to remain undetected is related to greater differential responding (Gustafson and Orne, 1963; Elaad and Ben-Shakhar, 1989; Furedy and Ben-Shakhar, 1991; Ben-Shakhar and Elaad, 2003). Likewise, a demonstration of the effectiveness of the apparatus detection procedure enhances the physiological response differences (Stern et al., 1981; Saxe, 1991), as does a lack of perceived success in deceiving (Gustafson and Orne, 1965). The same holds for a stronger intention to deceive (Furedy and Ben-Shakhar, 1991), a greater response conflict between the predominant truthful and the required deceptive answer (Furedy and Ben-Shakhar, 1991; Bradley et al., 1996), and a greater attentiveness throughout the test (countered by countermeasures; see, e.g., Elaad and Ben-Shakhar, 1991). In addition, an “active” questioning format (e.g., “Did you steal this object?”) has been suggested to be more effective than a more “passive” questioning format (e.g., “Was this object in the deed room?”, “Did you see this object?”; Bradley et al., 1996; Ambach et al., 2011a; but see Gamer, 2010).

It is conceivable that several of these factors that influence differential responding in a CIT depend on the social situation in which the CIT takes place. For example, the physical presence of an interrogator might enhance the motivation to remain undetected or the fear of being detected; on the contrary, facilitating the motivation to confess is also conceivable; a combination of both might enhance response conflict. Participants might also perceive the interrogator as controlling their behavior in the CIT, which would help to focus attention on the test; on the other hand, a present interrogator might divert attention from the test. The presence of a person might also lead to a stronger emotional involvement in the situation and to a more intense conflict between disclosure and withholding information, i.e., between truthful and deceptive responding; a tendency toward withdrawal and alienation, i.e., lower emotional involvement, is thinkable as the opposite. While both directions of influence are principally conceivable, more general studies on the social influences on physiology predominantly suggest an increased involvement and enhanced physiological responding in a more “social” condition:

A general dependence of physiological responses on social aspects, particularly the presence of another person, is assumed due to the findings of earlier sociophysiological studies. Zajonc

(1965) derived his *social facilitation theory* from studies investigating the influence of the sheer presence of another person (“audience”) and the “co-action” with another person on a subject’s behavior; increased arousal, “stress,” and induced emotions (e.g., fear) were assumed to be important moderators of behavior and physiological correlates. Martens (1969) found palmar sweating increased when subjects learned a motor task in the presence of an audience as compared to learning the same task alone. Glass et al. (1970) found greater skin conductance levels (SCL) in participants watching an aversive film if they were accompanied by a second spectator. Apprehension about evaluation, i.e., the presence of an evaluative second person, has been shown to increase muscle tension (Chapman, 1973) as well as heart rate (HR; Hrycaiko and Hrycaiko, 1980) as indicators of arousal. Referring to the CIT, the social situation, under which the test is applied, is supposed to comprise aspects of (negative) social evaluation and enhanced negative emotions (e.g., guilt, fear), which increase stress and arousal in an individual.

If the social conditions, under which a participant is investigated in the CIT, are influencing the various physiological responses, another question immediately arises: which components of the social situation are crucial for influencing attention, intention, motivation, emotion, and the accompanying physiological responses in a CIT? Beyond the evidence that the sheer presence of a second person can influence behavior and physiology, the type of social interaction, and specific social elements in a given situation have proven important: negative social evaluation specifically increases salivary cortisol levels (Dickerson et al., 2008). Specific interaction with virtual others has been observed to lead to brain activity different to that induced by the mere presence of virtual others (Schilbach et al., 2006). Considering observable behavior, Haley and Fessler (2005) found that a picture with a pair of eyes increased generosity in an anonymous game. In a study by Sproull et al. (1996), a virtual “talking face,” in contrast to a “text display,” made participants more aroused and led them to present themselves in a more positive light.

In sum, specific situational components of a social interrogation (which the original CIT is) influence emotions, arousal, and motivation of a participant. Visual (i.e., seeing a face or parts of it) and auditory elements (i.e., hearing a voice) make a computer interface more human-like and can, thus, be assumed to induce behavior and physiology more similar to a real-life interpersonal interrogation. While some studies used an auditory presentation of the CIT questions, others used a text display; to our knowledge, a comparison of both has not yet been undertaken. In addition, so far no CIT studies exist employing other social stimuli like a virtual investigator’s face within a virtual interrogation situation.

PERSONALITY ASPECTS IN THE CIT

Differential psychology in the context of the CIT has been studied since the very origin of the test; yet, various questions still remain open. First, physiological responding strongly differs between individuals; differences in electrodermal lability or HR variability have been shown to be associated with personality traits such as neuroticism, extraversion, and impulsivity (Coles et al., 1971; Crider and Lunn, 1971; O’Gorman, 1990). Lykken (1957) found lower

overall electrodermal response amplitudes in sociopathic individuals. Later studies found personality traits such as the “level of socialization” (Waid, 1976; Waid et al., 1979; Waid and Orne, 1980, 1981) to be correlated with differential physiological responding in the CIT and the detection of deception in general.

Over the last decades, psychopathy has been a prominent personality concept in this line of research. An established assessment instrument for psychopathy, even in a standard population sample, is the Psychopathic Personality Inventory – Revised (*PPI-R*; Lilienfeld and Andrews, 1996; Lilienfeld and Widows, 2005; German version: Alpers and Eisenbarth, 2008). Its relation to individual differences in physiological responding to CIT items has repeatedly been studied, particularly from a forensic perspective. Accordingly, most of the research exclusively used male participants and followed the standard computer-based interrogation procedure. As summarized by Verschuere (2011), the so far reported studies investigating CIT accuracy in samples differing with respect to delinquency and psychopathy have yielded inconsistent results. While some studies (e.g., Verschuere et al., 2007) report reduced overall electrodermal responding in prison samples, others (e.g., Verschuere et al., 2005) do not; a solid correlation between psychopathy score and differential responding in the CIT cannot be regarded as confirmed. In sum, personality influences on physiological responding in the CIT need to be elucidated by further research.

In connection with the main focus of the study, we were particularly interested in a psychopathy measure, because psychopathy, repeatedly described as including an “affective, interpersonal facet” (summarized by Verschuere, 2011), can be assumed to influence social interaction and possibly its physiological correlates. Social stimuli might exert different impact in individuals with different psychopathy scores. With respect to the CIT, it is speculatively questioned whether the influence of a present person or other social stimuli might be modulated by specific personality traits such as psychopathy.

AIM OF THE PRESENT STUDY

1. We examined the influence of face and voice as social stimuli on the physiological responses in a CIT. In this first attempt to directly examine this influence, we did not aim at differentiating between the modality of question presentation (visual vs. auditory) and its social content. Rather, in order to maximize the effect of “text” vs. “face and voice” presentation, we employed a uniform male face in combination with a neutral but serious male voice to simulate the “virtual investigator.” In a between-subjects manipulation, two CIT variants, a “text” and a “social” condition, were compared with respect to differential physiological responding. We expected greater differential responses and higher correct-classification rates in the condition with social stimuli (as compared to the “text” condition) for all physiological measures.
2. We included the *PPI-R* questionnaire in order to investigate whether differential physiological responding in a CIT is mediated by psychopathic traits, i.e., whether participants with higher *PPI-R* sum scores show smaller differential physiological responses. An analysis of correlation coefficients between physiological response differences and the *PPI-R* scores was planned

for this purpose. Additionally, we were interested in possible interactions between the influence of social stimuli and the psychopathy score: we expected the influence of social stimuli on differential physiological responding to decrease with heightened *PPI-R* scores.

MATERIALS AND METHODS

SUBJECTS

Seventy-one healthy students (33 males, 38 females; mean age 23.4 ± 3.7 years) voluntarily participated in the study. They were paid 12 Euros, with an additional incentive of 3 Euros. Data from two subjects were discarded from evaluation because of technical problems or insufficient compliance with the instructions. An ethics committee confirmed that the study met all ethical requirements.

DESIGN AND PROCEDURE

The experiment was divided into two parts (mock-crime in an “office room” and detection procedure in the “laboratory”), each guided by a different experimenter.

To begin with, the first experimenter explained the procedure to the subjects in the reception room of the department; informed consent was obtained from all participants. A cover story and the use of two rolled-up documents were used to make participants believe that they randomly drew one of two different instructions to perform a “special task” in the first part of the experiment, while, in fact they all received an equivalent mock-crime instruction. The second experimenter, who (in accordance with the information given to the participants) was blind with respect to the mock-crime objects a particular participant had handled in the first part, was introduced as the person responsible for “detecting whether the subjects had stolen something in the office room or not.” Subjects were randomly assigned to either of two groups: half of the subjects (i.e., the *text* group; 34 valid data sets) underwent a CIT using questions presented within a text field on the screen. The other half (i.e., the *social* group; 35 valid data sets) underwent a CIT with questions being asked by a pre-recorded male voice presented via loudspeakers, while a male face was presented as a picture on the screen. Written CIT instructions for the *text* group stated that the experimenter’s aim was to find out the truth by means of “a computer program and physiological measurement,” whereas the corresponding instructions for the *social* group stated that the experimenter’s aim was to find out the truth by means of “a virtual investigator and physiological measurement.” After completing the CIT and a subsequent memory test, subjects filled in the *Psychopathic PPI-R* before they were debriefed and released. Payment included the incentive of 3 Euros, regardless of a participant’s responding in the CIT.

MOCK-CRIME SCENARIO

Alone and unwatched in an office room of the institute, subjects unrolled the “task instruction” obtained from the first experimenter. They had to remove (“steal”) nine objects from this room after having extensively viewed each of them. The choice of the nine objects, one from each category, was randomized and balanced across subjects. The object categories, each comprising five objects, were: key pendants, kitchen objects, boxes, office materials,

cosmetics, wooden toy fruits, drink packages, playing cards, and plastic flowers.

Subjects were advised to collect all nine items in a suitcase, which they should keep closely to themselves throughout the remaining experiment. An amount of 3 Euros was hidden in one of the stolen objects (a box); later, this served as an incentive to "remain undetected."

CONCEALED INFORMATION TEST

The "physiological investigation" took place in the laboratory with the second experimenter; recording devices were attached. The CIT consisted of nine blocks referring to the nine item categories (e.g., key pendants, cosmetics). Each block comprised one question with five answer alternatives: the *probe* ("stolen") item of each category and four corresponding *irrelevant* items, which were all unknown to the subjects.

For the *text* group, the text of each question appeared on the screen five times in sequence, each time followed by a different picture of one of the five answer alternatives, which appeared below the question with a delay of 3.5 s. For the *social* group, the question was presented acoustically with a pre-recorded male voice via speakers; instead of a written question, the picture of "the investigator," a uniform male face, appeared on the screen. This picture was derived from an Ekman picture in black and white; the man's facial expression was serious and he was about 40 years old. **Figure 1** shows pictures of the screen for both groups in the phase after a specific item was presented, but before the answer was given.

The first item presented for each question served as buffer item; the according trials were discarded from analysis. Preceding each block, two *neutral* items were presented as distractors. The according questions referred to everyday objects that had to be identified (e.g., "Is this a slide projector?"). The two questions had to be answered correctly, one with "yes" and the other with "no" (in a pseudorandomized sequence), to prevent subjects from answering

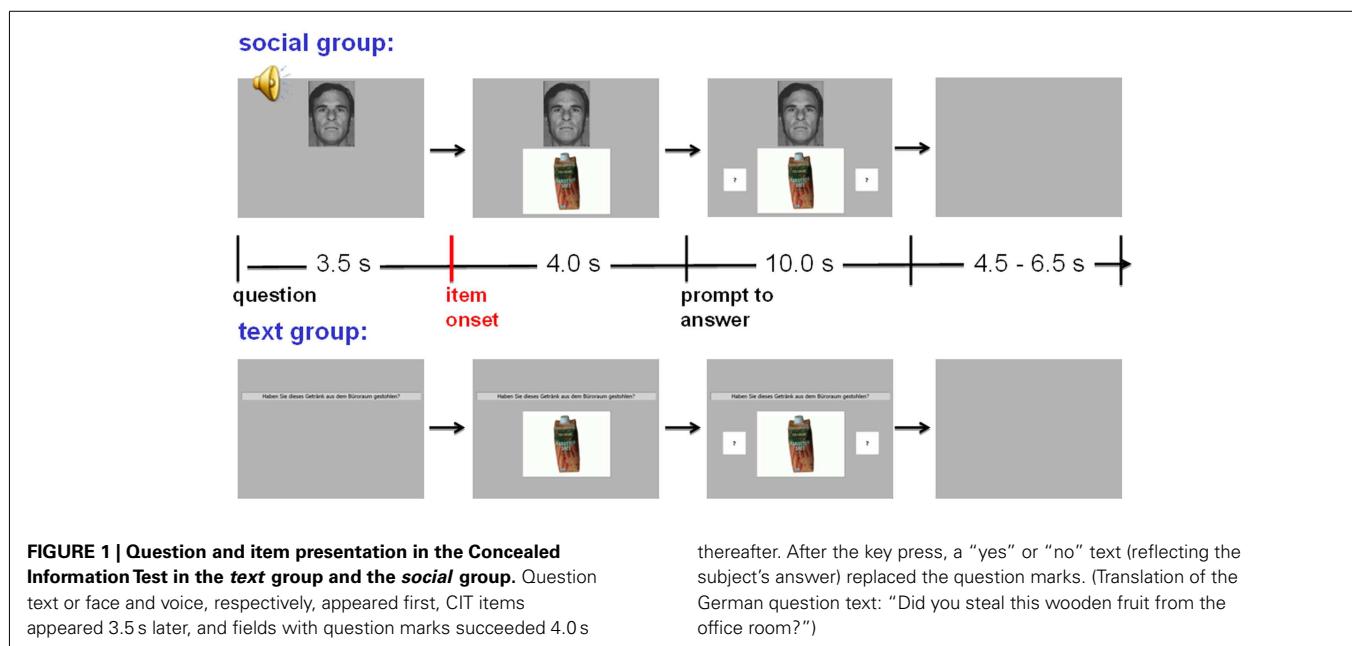
automatically with "no." Responses to these *neutral* questions were not evaluated. Together with the two *neutral* questions preceding each category, the entire procedure resulted in a total of 63 item presentations. The main run was preceded by a training run consisting of two blocks, each with five *neutral* items. Questions and item pictures were presented for 10 s foveally on a 19" monitor at a distance of 90 cm, followed by a blank screen for equally distributed 4.5–6.5 s intervals. Picture size was 10.6° by 8.0° of visual angle for the CIT items; the "investigator" presented in the *social* group was 5.6° by 8.0° in size. Four seconds after a question was asked, two indication fields containing question marks appeared on either side of the item picture; this prompted the subjects to answer. Then, answers had to be given as quickly as possible by pressing one of the two response keys and by vocally responding with "yes" or "no." Key assignment was balanced across subjects. Following the answer, the given "yes" or "no" replaced the question marks and remained visible on the screen as long as the item question was presented.

Subjects were told to hide their knowledge about the objects that had been stolen from the administration room, i.e., to deny all knowledge about *probe* items. Different from the typical CIT wording, an active questioning format was chosen: questions were, e.g., "Did you steal this cosmetic product from the administration room?"

After subjects were disconnected from the leads, they underwent a memory test: all five pictures of each category were presented on the screen simultaneously, one item category after the other; subjects were asked to identify the item they had stolen within each category.

PHYSIOLOGICAL MEASURES

The physiological recordings took place in a dimly lit, electrically and acoustically shielded experimental chamber (*Industrial Acoustics GmbH, Niederkrüchten, Germany*). Subjects sat in an upright position so that they could comfortably see the monitor



and reach the keyboard. Temperature in the cabin was set to 21°C at the beginning of the first run, with an increase of maximum 2°C throughout the course of the experiment.

Skin conductance, respiratory activity, electrocardiogram (ECG), and finger plethysmogram were registered. Physiological measures were A/D-converted and logged by the *Physiological Data System I 410-BCS* manufactured by *J&J engineering* (Poulsbo, WA, USA). The A/D-converting resolution was 14 bit, allowing skin conductance to be measured with a resolution of 0.01 µS. All data were sampled with 510 Hz. Triggers indicating question onsets were registered with the same sampling frequency.

For skin conductance recordings, standard Ag/AgCl electrodes (*Hellige*; diameter 0.8 cm), electrode paste of 0.5% saline in a neutral base (*TD 246 Skin Resistance, Mansfield R&D, St. Albans, Vermont, UK*), and a constant voltage of 0.5 V were used. The electrodes were fixed at thenar and hypothenar sites of the non-dominant hand. For registration of respiratory activity, two PS-2 biofeedback respiration sensor belts (*KarmaMatters, Berkeley, CA, USA*) with a built-in length-dependent electrical resistance were used. They were fixed at the upper thorax and the abdomen. ECG was measured with *Hellige* electrodes (diameter 1.3 cm) according to Einthoven II. Finger pulse signal was transmitted by an infrared system in a cuff around the middle finger of the non-dominant hand.

BEHAVIORAL MEASURES

Subjects responded verbally as well as by pressing a key. Key presses indicating “yes” or “no” answers were time-logged, synchronized with the physiological measures, and stored on the stimulus-presenting computer. Importantly, answers were delayed by 4 s in this study; after this delay, most stimulus processing and answer preparation can be assumed to be completed; in addition, it is rather easy to perform strategic manipulations by voluntarily controlling reaction speed after the delay. Therefore, behavioral data were not analyzed. CIT questions with at least one item answered incorrectly were discarded from analysis, but no such case occurred.

QUESTIONNAIRE

As the last part of the experiment, participants filled in the *PPI-R*. It comprises 154 items to assess the individual degree of psychopathic traits; the sum scale and (for exploratory purposes) the nine subscales were calculated from the raw data and then, in order to account for gender differences, transformed into *T* values (Alpers and Eisenbarth, 2008). To investigate the relationship between the psychopathy measure and physiological responsiveness in the CIT, correlation coefficients were calculated for the individual *T* values on each subscale and the individual standardized *probe-minus-irrelevant* response differences for each physiological data channel.

DATA PROCESSING

Skin conductance data from two subjects (one from the *text* group, one from the *social* group) had to be discarded from analysis because of electrodermal non-responding. Skin conductance reactions were assessed by a computerized method (see Ambach et al., 2008; Ambach et al., 2010) based on the decomposition of overlapping reactions as proposed by Lim et al. (1997). This method was

chosen because, two subsequent physiological reactions occurred with a short delay (due to the delay of 4 s between question and prompt to answer). With short interstimulus intervals, conventional trough-to-peak evaluation is inadequate (Lim et al., 1999), because the first of two reactions causes a diminishing bias in the estimation of the second one. The size of this bias is determined by the size of the first reaction and by the time interval between both reactions. Decomposition aims at overcoming this problem of overlapping EDA reactions.

After optimizing model coefficients for each subject, all trials were evaluated by decomposing EDA by use of the subject's individual model coefficients. Then, magnitudes of all EDA responses that were elicited within a time window of 0.5 to 4.5 s after item presentation were additively combined to a “first response” (EDA_1). The sum of EDA responses, which began between 4.5 and 8.5 s after item presentation, i.e., between 0.5 and 4.5 s after the subjects were prompted to answer, was calculated as “second response” (EDA_2). For the regression analysis, a combined response measure (EDA_sum) was calculated by adding both components per trial. For each time window, the decomposed responses were transformed into their equivalent in µS according to the subject's individual electrodermal response template.

Respiratory data were low-pass filtered (10 dB at 2.8 Hz); respiration line length (RLL) was automatically computed over a time interval of 15 s after trial onset. The RLL measure integrates information about frequency and depth of respiration. The method was derived from Timm (1982) and modified by Kircher and Raskin (2003). Respiratory data from one subject (from the *social* group) were discarded due to a technical failure. For analysis, raw data from both respiratory channels were averaged.

Electrocardiogram data obtained from three subjects (one from the *text* group, two from the *social* group) had to be excluded from analysis because of technical failure or arrhythmia. After notch filtering at 50 Hz, *R*-wave peaks were automatically detected and visually controlled. The *R-R* intervals were transformed into HR and real-time scaled (Velden and Wölk, 1987). The HR during the last second before trial onset served as pre-stimulus baseline. The phasic heart rate (pHR) was calculated by subtracting this baseline value from each second-per-second poststimulus value. For extracting the trial-wise information of the phasic HR, the mean change in HR within 15 s after trial onset, compared to the pre-stimulus baseline, was calculated (see Verschueren et al., 2007; Gamer et al., 2008).

Finger pulse waveform length (FPWL) data from five subjects (three from the *text* group, two from the *social* group) had to be discarded from analysis because of insufficient signal quality. The FPWL within the first 15 s after trial onset was calculated from the finger pulse waveform and then subjected to further analyses (Elaad and Ben-Shakhar, 2006). It comprises information about both HR and pulse amplitude.

In order to compare indicators of arousal between-groups, we additionally computed the individual averages of non-standardized SCL and HR at trial onsets. SCL and HR data were averaged over the last second before the onset of a CIT question (i.e., 3.5–4.5 s before item onset).

A within-subject standardization of measured values has been proposed by Lykken and Venables (1971). Here, according to

Ben-Shakhar (1985), Gamer et al. (2006), and Gronau et al. (2005), the physiological measures are *z*-transformed for each subject and for each data channel. All *probe* and *irrelevant* trials (but not *neutral* trials and the first trials of each stimulus category) were used to calculate individual means and standard deviations. The *z*-transformed values were used in subsequent statistical analyses.

STATISTICAL ANALYSIS

Statistical analyses were performed with SYSTAT, Version 13 (SYSTAT Software, Inc., Monte Carlo).

For each measure, mean responses to *probe* vs. *irrelevant* items were compared using one-tailed *t*-tests (matched samples), separately for *text* and *social* group. An additional *t*-test was performed to test whether the *probe*-minus-*irrelevant* response differences were enhanced in the *social* as compared to the *text* group.

Significance level was set to 0.05; Cohen's *d* was calculated as estimate of effect size (Cohen, 1988; Rosnow and Rosenthal, 1996).

Besides investigating the effects of the different questioning formats on each of the physiological measures, the capability of detecting concealed information in both groups was of interest from an applied perspective. For this purpose, the validity of each data channel and the validity of an optimized combination of the measures (EDA_sum, pHr, RLL, and FPWL) were analyzed using a binary logistic regression analysis. Because all participants in this study had deed-related knowledge, responses of a hypothetical group of "innocent" subjects were simulated according to Meijer et al., 2007; simulated trial-by-trial values were randomly drawn from a standard normal distribution.

Binary logistic regression analyses were performed with inclusion of each of the measures and with a fixed inclusion of all four measures (which in contrast to a stepwise inclusion prohibits that the included measures differ between-groups). A cross-validation was run using the hold-one-out method, separately for the *text* group and the *social* group: each subject's classification as "guilty" or "innocent" was based on a combination of his or her standardized differential physiological responses with weights calculated from all other "guilty" and "innocent" subjects. The receiver-operating characteristic (ROC) allows to estimate the capability of differentiating guilty from innocent participants for all possible cut-off points and for different dependent measures and their combination. The area under the ROC curve varies between 0

and 1 with a chance level of 0.5 and serves as an overall index of detection accuracy (Bamber, 1975; Ben-Shakhar and Elaad, 2003; Gronau et al., 2005).

RESULTS

MEMORY TEST

In the memory test, 99.2% of the *probe* items were identified correctly (99.0% in the *social* and 99.3% in the *text* group). Categories with false identification of the *probe* item were discarded from evaluation. (Note that restricting the analyses to categories with correct probe identification, as well as the exclusion of data from non-compliant or physiologically hyporesponsive participants, which are standard procedures within the experimental context, can lead to an inflation of effect sizes and detection rates when transferred to real-life CIT investigations.

OVERVIEW OF PSYCHOPHYSIOLOGICAL MEASURES

Preceding data standardization and test statistics, descriptive statistics based on raw scores are presented. **Table 1** summarizes means and standard errors of means of raw scores for each data channel separately for both groups.

Figure 2 illustrates the differential responses to *probe* vs. *irrelevant* items for both groups. Response differences (*z*-scores) between probe and irrelevant trials are depicted for each of the physiological measures.

SKIN CONDUCTANCE

Figure 3 shows the averaged intra-trial course of skin conductance depicting grand means for trials with *probe* and *irrelevant* items separately for both groups. The grand means show two strong EDA response components with an onset and peak asynchrony of 4 s, which is in accordance with the 4-s delay between item onset and prompt to answer. Response amplitudes to *probe* items exceeded those to *irrelevant* items by far in both groups, with no apparent difference between-groups. The additional EDA response, which was observed 3.5 s before the response to item onset, can be ascribed to the onset of the question text, or the face and voice respectively. An exploratory analysis of this component using *t*-tests (corresponding with the *t*-tests performed on all other measures) revealed no significant difference between item types and no group difference for the *probe*-minus-*irrelevant* differential response.

Table 1 | Means and standard errors of means (SEM) of raw scores for each data channel.

	<i>Text</i> group				<i>Social</i> group			
	<i>Probe</i> items		<i>Irrelevant</i> items		<i>Probe</i> items		<i>Irrelevant</i> items	
	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM
EDA_1 (nS)	304	38	141	16	270	48	109	14
EDA_2 (nS)	320	41	233	33	283	34	186	22
pHR (1/min)	-1.41	0.45	0.41	0.24	-3.39	0.37	-0.90	0.30
RLL (arb. units)	1755	127	2019	151	1720	145	2005	163
FPWL (arb. units)	181	20	207	22	163	13	198	15

Responses to *probe* and *irrelevant* items are listed separately for *text* and *social* group.

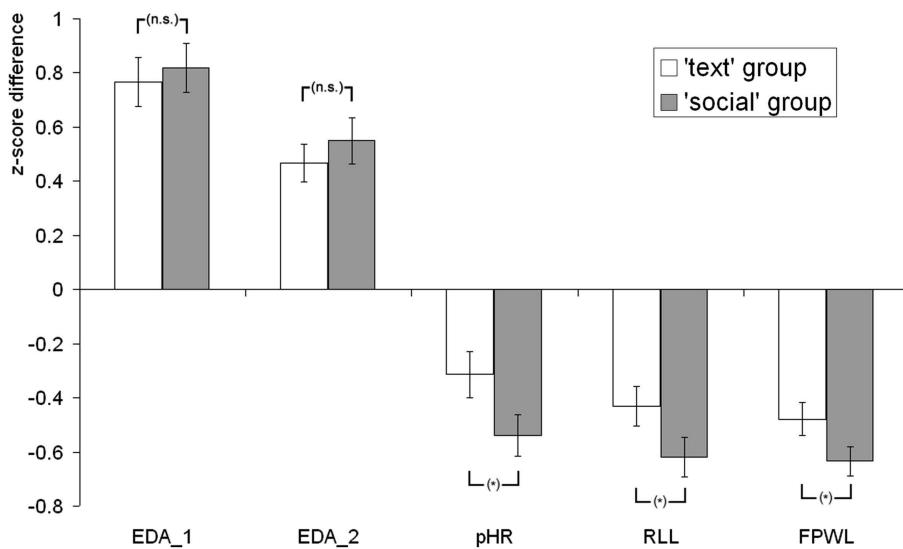


FIGURE 2 | Differential responses (z-scores) to probe vs. irrelevant items: for the *text* and the *social* group, standardized response differences are depicted for first electrodermal reaction (EDA_1), second electrodermal reaction (EDA_2), phasic heart rate (pHR),

respiration line length (RLL), and finger pulse waveform length (FPWL). Error bars represent the SEM; *indicate the level of significance of the group difference (*text* vs. *social* group; "n.s.": not significant; $p < 0.05$).

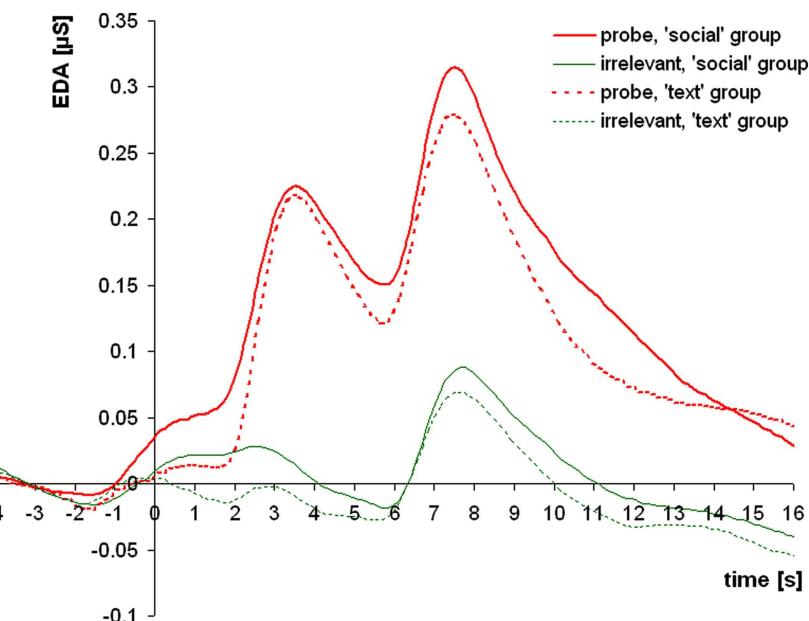


FIGURE 3 | Grand means of skin conductance responses to probe and irrelevant items for the *text* group and the *social* group. After a small initial response to the appearance of the question (text or face and

voice, respectively), two subsequent electrodermal responses of interest (EDA_1 and EDA_2) follow the item presentation and the prompt to answer.

EDA_1 responses were greater to *probe* than to *irrelevant* items in the *text* group ($t_{32} = 8.60$; $p < 0.001$; $d = 1.50$) as well as in the *social* group ($t_{33} = 9.01$; $p < 0.001$; $d = 1.55$). The between-groups *t*-test for EDA_1 response differences did not reveal greater *probe*-minus-*irrelevant* response differences in the *social* as compared to the *text* group ($t_{65} = -0.40$; $p > 0.1$).

Analogously, EDA_2 responses were greater to *probe* than to *irrelevant* items in the *text* group ($t_{32} = 6.77$; $p < 0.001$; $d = 1.18$) as well as in the *social* group ($t_{33} = 6.37$; $p < 0.001$; $d = 1.09$). The between-groups *t*-test for EDA_2 response differences did not reveal greater *probe*-minus-*irrelevant* response differences in the *social* as compared to the *text* group ($t_{65} = -0.74$; $p > 0.1$).

An additional ANOVA for probe vs. irrelevant response differences, which included a factor “time” distinguishing between the first and the second electrodermal response component and a factor “group” distinguishing between text and social group, was performed. This analysis did not reveal an interaction between “time” and “group” ($p > 0.1$), indicating similar response patterns of the two electrodermal response components. For the logistic regression analysis, both components were then additively combined in a single measure: EDA_sum. EDA_sum responses were also greater to probe than to irrelevant items in the text group ($t_{32} = 9.59$; $p < 0.001$; $d = 1.67$) as well as in the social group ($t_{33} = 9.48$; $p < 0.001$; $d = 1.63$). Probe-minus-irrelevant response differences for EDA_sum were not greater in the social as compared to the text group ($t_{65} = -0.41$; $p > 0.1$).

RESPIRATION

Respiration line length values were smaller after *probe* than after *irrelevant* items in the *text* group ($t_{33} = -5.93$; $p < 0.001$; $d = -1.02$) as well as in the *social* group ($t_{33} = -8.51$; $p < 0.001$; $d = -1.46$). The between-groups *t*-test for RLL differences revealed greater *probe*-minus-*irrelevant* RLL differences in the *social* as compared to the *text* group ($t_{66} = 1.82$; $p < 0.05$; $d = 0.44$).

HEART RATE

Heart rate decelerations were more pronounced after *probe* than after *irrelevant* items in the *text* group ($t_{32} = -3.64$; $p < 0.001$; $d = -0.63$) as well as in the *social* group ($t_{32} = -6.95$; $p < 0.001$; $d = -1.21$). The between-groups *t*-test for pHR differences revealed greater *probe*-minus-*irrelevant* pHR differences in the *social* as compared to the *text* group ($t_{64} = 1.94$; $p < 0.05$; $d = 0.48$).

FINGER PULSE

Finger pulse waveform length values were smaller after *probe* than after *irrelevant* items in the *text* group ($t_{31} = 7.88$; $p < 0.001$; $d = -1.39$) as well as in the *social* group ($t_{31} = 11.82$; $p < 0.001$; $d = -2.09$). The between-groups *t*-test for FPWL differences revealed greater *probe*-minus-*irrelevant* FPWL differences in the *social* as compared to the *text* group ($t_{62} = 1.93$; $p < 0.05$; $d = 0.48$).

TONIC MEASURES OF AROUSAL

When comparing indicators of arousal between-groups, SCL appeared higher in the *text* group ($5.03 \pm 2.41 \mu\text{S}$) than in the *social* group ($4.15 \pm 1.66 \mu\text{S}$). This was contrary to the expectation and would have reached statistical significance in case of an inverted *a priori* hypothesis ($t_{65} = -1.730$, $p = 0.044$). Inspection of the raw data indicated that this result was due to an initially enhanced EDA level in the *text* group that was preserved throughout the entire examination. HR appeared higher in the *social* group ($76.02 \pm 26.11 \text{ bpm}$) than in the *text* group ($74.90 \pm 21.53 \text{ bpm}$), but also this was not statistically significant ($t_{69} = -0.196$, $p = 0.845$).

RECEIVER-OPERATING CHARACTERISTIC

Binary logistic regression analyses were performed to classify the subjects (half “guilty” participants, and half hypothetical

“innocents”) as “guilty” or “innocent”; hence, the *a priori* probability was set to 0.5. Separately for the *text* group and the *social* group, regression models were calculated with inclusion of each individual physiological measure as well as with fixed inclusion of EDA_sum, pHR, RLL, and FPWL. The classification performance was shrinkage-corrected using the hold-one-out method (which, in turn, resulted in different regression coefficients for each subject). The different rates of false-positive (classification of an “innocent” subject as “guilty”) and false-negative outcomes (classification of a “guilty” subject as “innocent”) obtained under variation of the cut-off point for decision were calculated separately for the *text* group and the *social* group.

Table 2 shows the areas under ROC and their confidence intervals for each of the single measures and the shrinkage-corrected areas under ROC for the optimal-weight combination of EDA_sum, pHR, RLL, and FPWL. (Note that for single measures the ROC values are equivalent to those obtained without a logistic regression analysis.)

Figure 4 shows the ROC curves for the *text* group and the *social* group with the optimal-weight combination of the four physiological measures after shrinkage-correction.

While no difference between-groups in test validity is apparent with EDA, the single cardiovascular and respiration measures, and also the optimal-weight combination of measures, yielded apparently greater areas under ROC for the *social* group than for the *text* group. According to the large confidence intervals however, none of these between-groups differences turned out significant in a bootstrap analysis ($p > 0.05$ for FPWL; $p > 0.1$ for all other measures and for the optimal-weight combination).

PSYCHOPATHIC PERSONALITY INVENTORY – REVISED

From the *PPI-R*, the individual sum scores and (for exploratory purposes, not reported) the nine subscale scores were calculated. *PPI-R* data from the two participants precluded from physiological analysis were treated as missing data.

For female participants ($N = 37$), the *PPI-R* sum scores of 277.65 ± 21.49 were lower than in the reference sample (313.33 ± 25.45 ; Alpers and Eisenbarth, 2008; $p < 0.001$). For male participants ($N = 32$), the sum scores of 299.32 ± 27.25 were also lower than in the reference sample: 325.42 ± 24.92 ; $p < 0.001$).

Correlation coefficients between *T*-transformed *PPI-R* sum scores and *probe*-minus-*irrelevant* response differences for each physiological data channel were calculated for the *social* and the *text* group and across groups. Correlation coefficients for EDA_sum were 0.05 across groups, -0.01 for the *social* and 0.12 for the *text* group. The according values were -0.09 , 0.00 , and -0.18 for pHR, -0.03 , -0.01 , and -0.05 for FPWL, and -0.10 , -0.01 , and -0.20 for RLL, respectively. None of these correlations was statistically significant (all $p > 0.1$, uncorrected), indicating that differential responding in the CIT was not found to be moderated by *PPI-R* sum scores in either group or across groups. An additional bootstrap analysis was performed to assess confidence intervals for the correlation coefficients in either group; none of the group differences in correlation coefficients was significant ($p > 0.1$ each, uncorrected), which suggests that the enhanced differential responding found in the *social* group for pHR, RLL, and FPWL was not moderated by *PPI-R* sum scores.

Table 2 | Area under the receiver-operating characteristic (ROC) curves and 95% confidence intervals for a differentiation of guilty vs. hypothetical innocent subjects.

Included parameters	Area under the ROC curve and 95% confidence intervals			
	Text group		Social group	
	Area	Confidence interval	Area	Confidence interval
SINGLE MEASURES:				
EDA_sum	0.885	0.784–0.965	0.901	0.821–0.965
pHR	0.742	0.618–0.862	0.813	0.698–0.906
RLL	0.748	0.627–0.860	0.832	0.724–0.922
FPWL	0.816	0.703–0.909	0.910	0.831–0.973
OPTIMAL-WEIGHT COMBINATION:				
EDA_sum + pHR + RLL + FPWL	0.922	0.845–0.977	0.971	0.919–1.000

Shrinkage-corrected values are listed for an inclusion of each single physiological measure and for an optimal-weight combination of EDA_sum, pHR, RLL, and FPWL.

DISCUSSION

Social influences on physiological responding in the CIT are insufficiently explored. The present study compared two computer-based CIT conditions with respect to differential physiological responses: one condition used a text-based interrogation; the other included face and voice as social stimuli into the interrogation. A psychopathy questionnaire was administered in order to investigate personality influences on physiological responding and to explore a possible interaction of psychopathy with the impact of the social stimuli.

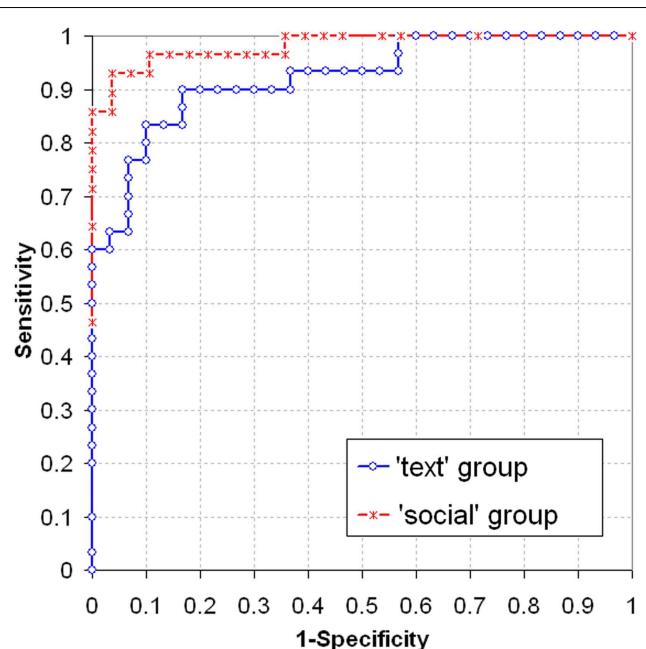
OVERALL CIT EFFECTS

For each physiological measure, in either of the two interrogation conditions as well as across conditions, significant response differences with large effect sizes were found between *probe* and *irrelevant* items.

Finger pulse waveform length yielded the greatest overall effect size, which is somewhat uncommon among most CIT studies. Elaad and Ben-Shakhar (2006), however, reported “that detection accuracy with the FPWL was at least as good as the accuracy obtained with (...) respiration changes and skin conductance responses.” Similarly, Vandenbosch et al. (2009) found CIT accuracy with FPWL as high as with EDA and better than with RLL, pHR, and finger pulse amplitude. Given that an adequate scoring of FPWL depends on sufficient signal quality, it might well be that influences such as the surrounding temperature before the experiment or the delay between temperature customization and CIT initiation, which are commonly not reported, differently influence finger pulse amplitude (and thereby signal quality and CIT accuracy) in different studies.

The time courses of skin conductance showed two temporally distinct response components with a delay of 4 s. Both EDA components, one after stimulus presentation, the other after the prompt to answer, showed large response differences between *probe* and *irrelevant* trials, with the first component yielding the greater effect size, which is in line with earlier studies (e.g., Ambach et al., 2008).

From a “detection” perspective, each of the measures in either group was capable of significantly differentiating “guilty” from

**FIGURE 4 | ROC curves for text group and social group with the predictors EDA_sum, pHR, RLL, and FPWL.**

hypothetical “innocent” subjects; ROC area values are in line with other studies. With an optimized linear combination of measures, the ROC area was 0.946 across groups, which reflects an adequate overall CIT accuracy.

TEXT CONDITION VS. FACE AND VOICE CONDITION

In a between-subjects manipulation, two CIT conditions differed in the way the CIT questions were presented (*text* group: text on the screen; *social* group: voice via speakers plus face on the screen); depiction of CIT items was identical.

Differential responding in pHR, RLL, and FPWL, indicating cardiac, pulmonary, and vascular functioning, differed significantly between conditions, which met the *a priori* expectation:

differential responding to *probe* vs. *irrelevant* items was enhanced in the condition with social stimuli, as compared to the text presentation. Contrary to our expectation, differential electrodermal responses did not mirror this finding; electrodermal response differences neither differed significantly between conditions for the first nor the second component (nor for the two components combined).

The classification of “guilty” and (hypothetical) “innocent” subjects by means of an optimized linear combination of standardized measures yielded ROC area values of 0.922 in the *text* group and 0.971 in the *social* group. Although the difference in curves between conditions appears prominent visually, and although the size of the underlying group effect differed significantly between-groups for three of the measures, this group difference in ROC curves was not statistically significant. The inclusion of data from “innocent” participants (be they real or hypothetical) entailed additional error variance which obscured the significant group difference observed in the dependent measures.

POSSIBLE EXPLANATIONS AND CONFOUNDS

Given the different impact the two-part experimental manipulation (combining face and voice in the *social* condition) exerted on the physiological measures, it can be conjectured once more (see e.g., Ambach et al., 2008, 2011b) that the responses of the individual physiological channels are not reflecting a singular psychophysiological process ongoing in the CIT (such as a unitary orienting response; see Barry, 1996, 2006). The observation that all physiological measures, except the electrodermal, were affected by the experimental manipulation suggests that processes other than orienting seem to depend on the type of CIT presentation. Earlier CIT studies, which were based on the electrodermal measure (e.g., Gustafson and Orne, 1963; Elaad and Ben-Shakhar, 1989; Furedy and Ben-Shakhar, 1991; Ben-Shakhar and Elaad, 2003), document that intentional and motivational manipulations influenced differential responding also for EDA. One might conclude that this renders intention and motivation less likely the moderators of the manipulation observed in the present study. A difference between conditions in physiological measures of arousal, which might have contributed to an explanation, remained also unproven. It is further conceivable that the different presentation types directed a participant’s attention in a different manner; a voice might more forcefully direct attention to the item presented thereafter, whereas a displayed text might allow to divert attention more easily. The longer duration of voice presentation, as compared to capturing a written text, might contribute to this. An additional, more speculative explanation, refers to experimental attempts to disentangle orienting from deceptive components in a CIT (Ambach et al., 2008; Matsuda et al., 2012). The instruction to deceptively deny specific knowledge had greater effects on cardiovascular and/or respiratory measures than on EDA. One might thus speculate that face and voice, in contrast to a visual text presentation, affect the same CIT subprocess reflected in pHR, RLL, and FPWL, namely a subprocess closely related to deceptive action. Finally, the failure of EDA to replicate the findings of the other measures might be explained by a ceiling effect; in case the electrodermal system is maximally differentially activated already with textual question presentation, then social stimuli cannot be expected to enhance

differential responding. If such a ceiling effect is assumed, then a “face and voice” presentation might be particularly advantageous over a “text only” presentation when measures other than EDA are used, or when EDA, due to suboptimal test conditions, does not reach optimal detection levels (e.g., in cases of only partial crime-related knowledge of the suspects, or with the use of countermeasures).

As Bradley (2009) suggested, the orienting response can fruitfully be regarded as embedded in motivational and attentional systems active and fluctuating within an individual. Instead of debating whether the “social” manipulation in this study had an impact on orienting, motivation, attention, or emotion, it might be more groundbreaking to regard the presence or absence of social stimuli as modifying the subject’s environment in the sense that it alters the intentional, motivational, attentional, and emotional background the orienting response takes place in.

Whether it was the virtual investigator’s face or his voice presented via speakers, or the combination of both, that determined the enhancement of differential responding in the affected measures, cannot be decided from this study: following the primary aim to maximize the experimental manipulation, face and voice were planned to occur in fixed combination.

A confound of the text vs. voice question presentation, meant to contrast absence vs. presence of voice as a social stimulus, with the visual vs. auditory presentation modality is obvious. This confound is inherent whenever visually presented text is compared with spoken word: speech is essentially human; therefore, spoken text is always a social stimulus (even in case of an alienated voice). Thus, if an auditory presentation (even without face) enhanced differential responding more than a visual text presentation, the question whether this may be called a “social” effect seems subordinate to the question, by which pathways spoken text is superior to written text in the CIT.

PERSONALITY ASPECTS: PSYCHOPATHY AND THE CIT

In line with earlier studies, PPI-R scores were greater in male than in female participants. However, the overall scores (for both genders) were smaller than the normative values for the German questionnaire version (Alpers and Eisenbarth, 2008). On the other hand, Uzieblo et al. (2010) provided standard sum scores obtained from a large population sample (males: 283.00 ± 34.30 , $N = 419$; females: 266.87 ± 32.12 , $N = 256$). Taking these values into account (although obtained with a different translation of the PPI-R), the mean values obtained in the present study do not point toward a biased sample.

We did not find a correlation between the psychopathy sum score and differential physiological responding in any of the four measures; the results of explorative analyses with subscales are not reported due to their complexity and fruitlessness. Likewise, no interaction effect was found that would have pointed toward a different impact of the social stimuli in individuals with high or low psychopathy scores.

SUGGESTIONS FOR LABORATORY STUDIES

In follow-up research, the two parts of the experimental manipulation should be disentangled. Voice presentation of CIT questions should be compared with a visual text presentation, separately

from investigating the influence of additional social stimuli such as a face; a follow-up study should enrich the present design by including a condition with auditory question presentation but without a depicted face.

While it should be easy to separate the influences of face from the influence of voice, it might not be that easy to separate modality effects (i.e., auditory vs. visual question presentation) due to the social character of the presentation (social: human voice; non-social: written text); this is due to the social nature of speech, *per se*.

Beyond the social stimuli used in the present study, the importance of particular elements of social presence and interaction should be highlighted more in CIT research. Beside the impact of isolated social stimuli (e.g., a depicted pair of eyes, or the sounds of a human voice), the importance of an investigator's appearance, demeanor, and social acting deserves more attention. This will resume a line of research that lay idle for a couple of decades (see e.g., Waid and Orne, 1981), due to the desire to standardize experiments as far as possible. Balancing standardization of experimental conditions and the investigation of social influences on psychophysiology in the CIT, which can be standardized only to a limited extent, should be aimed at. An additional suggestion arises from the design of the present study. Instead of the between-subjects design used here, conditions differing with respect to "social content" could be compared within-subject, given that this can be implemented meaningfully.

SUGGESTIONS FOR APPLIED RESEARCH

Besides inferable clear guidelines for future laboratory studies, the observed results have implications for research on the field application of the CIT. Different CIT interrogation types, different CIT settings, as well as different experimenter roles (e.g., harshness, social support, expression) and other specific characteristics (e.g., gender, ethnicity) might have a different social impact on a person investigated in the CIT (see Iacono, 2000). Studying these influences, comparing different practical settings, and optimizing conditions with respect to test accuracy, should become a focus

REFERENCES

- Alpers, G. W., and Eisenbarth, H. (2008). *Psychopathic Personality Inventory-Revised (PPI-R). German Version*. Göttingen: Hogrefe.
- Ambach, W., Bursch, S., Stark, R., and Vaitl, D. (2010). A Concealed Information Test with multimodal measurement. *Int. J. Psychophysiol.* 75, 258–267.
- Ambach, W., Dummel, S., Lüer, T., and Vaitl, D. (2011a). Physiological responses in a Concealed Information Test are determined interactively by encoding procedure and questioning format. *Int. J. Psychophysiol.* 81, 275–282.
- Ambach, W., Stark, R., and Vaitl, D. (2011b). An interfering n-back task facilitates the detection of concealed information with EDA but not with cardiopulmonary physiology. *Int. J. Psychophysiol.* 80, 217–226.
- Ambach, W., Stark, R., Peper, M., and Vaitl, D. (2008). Separating deceptive and orienting components in a Concealed Information Test. *Int. J. Psychophysiol.* 70, 95–104.
- Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *J. Math. Psychol.* 12, 387–415.
- Barry, R. J. (1996). Preliminary process theory: towards an integrated account of the psychophysiology of cognitive processes. *Acta Neurobiol. Exp. (Wars)* 56, 469–484.
- Barry, R. J. (2006). Promise versus reality in relation to the unitary orienting reflex: a case study examining the role of theory in psychophysiology. *Int. J. Psychophysiol.* 62, 353–366.
- Ben-Shakhar, G. (1985). Standardization within individuals: a simple method to neutralize individual differences in skin conductance. *Psychophysiology* 22, 292–299.
- Ben-Shakhar, G., and Elaad, E. (2003). The validity of psychophysiological detection of information with the Guilty Knowledge Test: a meta-analytic review. *J. Appl. Psychol.* 88, 131–151.
- Bradley, M. M. (2009). Natural selective attention: orienting and emotion. *Psychophysiology* 46, 1–11.
- Bradley, M. T., McLaren, V. V., and Carle, S. B. (1996). Deception and nondeception in guilty knowledge and guilty actions polygraph tests. *J. Appl. Psychol.* 81, 153–160.
- Chapman, A. J. (1973). An electromyographic study of apprehension about evaluation. *Psychol. Rep.* 33, 811–814.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. San Diego, CA: McGraw-Hill.
- Coles, M. G. H., Gale, A., and Kline, P. (1971). Personality and habituation of the orienting reaction: tonic and response measures of electrodermal activity. *Psychophysiology* 8, 54–63.
- Crider, A., and Lunn, R. (1971). Electrodermal lability as a personality dimension. *J. Exp. Res. Pers.* 5, 145–150.
- Dickerson, S. S., Mycek, P. J., and Zaldivar, F. (2008). Negative social evaluation, but not mere social presence, elicits cortisol responses to a laboratory stressor task. *Health Psychol.* 27, 116–121.
- Elaad, E., and Ben-Shakhar, G. (1989). Effects of motivation and verbal response type on psychophysiological detection of information. *Psychophysiology* 26, 442–451.

- Elaad, E., and Ben-Shakhar, G. (1991). Effects of mental countermeasures on psychophysiological detection in the guilty knowledge test. *Int. J. Psychophysiol.* 11, 99–108.
- Elaad, E., and Ben-Shakhar, G. (2006). Finger pulse waveform length in the detection of concealed information. *Int. J. Psychophysiol.* 61, 226–234.
- Furedy, J. J., and Ben-Shakhar, G. (1991). The roles of deception, intention to deceive, and motivation to avoid detection in the psychophysiological detection of guilty knowledge. *Psychophysiology* 28, 163–171.
- Gamer, M. (2010). Does the Guilty Actions Test allow for differentiating guilty participants from informed innocents? A re-examination. *Int. J. Psychophysiol.* 76, 19–24.
- Gamer, M., Rill, H. G., Vossel, G., and Gödert, H. W. (2006). Psychophysiological and vocal measures in the detection of guilty knowledge. *Int. J. Psychophysiol.* 60, 76–87.
- Gamer, M., Verschuere, B., Crombez, G., and Vossel, G. (2008). Combining physiological measures in the detection of concealed information. *Physiol. Behav.* 5, 333–340.
- Glass, D. C., Gordon, A., and Henchy, T. (1970). The effects of social stimuli on psychophysiological reactivity to an aversive film. *Psychon. Sci.* 20, 255–256.
- Gronau, N., Ben-Shakhar, G., and Cohen, A. (2005). Behavioral and physiological measures in the detection of concealed information. *J. Appl. Psychol.* 90, 147–158.
- Gustafson, L. A., and Orne, M. T. (1963). Effects of heightened motivation on the detection of deception. *J. Appl. Psychol.* 47, 408–411.
- Gustafson, L. A., and Orne, M. T. (1965). Effects of perceived role and role success on the detection of deception. *J. Appl. Psychol.* 49, 412–417.
- Haley, K. J., and Fessler, D. M. T. (2005). Nobody's watching? Subtle cues affect generosity in an anonymous economic game. *Evol. Hum. Behav.* 26, 245–256.
- Hrycaiko, D. W., and Hrycaiko, R. B. (1980). Palmar sweating in an evaluative audience situation. *J. Soc. Psychol.* 111, 269–280.
- Iacono, W. G. (2000). "The detection of deception," in *Handbook of Psychophysiology*, eds J. T. Cacioppo, L. G. Tassinary, and G. G. Berntson (Cambridge: Cambridge University Press), 772–793.
- Kircher, J. C., and Raskin, D. C. (2003). *The Computerized Polygraph System II (Software Version 4.01)*. Salt Lake City, UT: Scientific Assessment Technologies.
- Lilienfeld, S. O., and Andrews, B. P. (1996). Development and preliminary validation of a self-report measure of psychopathic personality traits in noncriminal populations. *J. Pers. Assess.* 66, 488–524.
- Lilienfeld, S. O., and Widows, M. R. (2005). *Psychopathic Personality Inventory-Revised: Professional manual*. Lutz, FL: Psychological Assessment Resources.
- Lim, C. L., Gordon, E., Rennie, C., Wright, J. J., Bahramali, H., Li, W. M., et al. (1999). Dynamics of SCR, EEG, and ERP activity in an odd-ball paradigm with short interstimulus intervals. *Psychophysiology* 36, 543–551.
- Lim, C. L., Rennie, C., Barry, R., Bahramali, H., Lazzaro, I., Manor, B., et al. (1997). Decomposing skin conductance into tonic and phasic components. *Int. J. Psychophysiol.* 25, 97–109.
- Lykken, D. T. (1957). A study of anxiety in the sociopathic personality. *J. Abnorm. Psychol.* 55, 6–10.
- Lykken, D. T. (1959). The GSR in the detection of guilt. *J. Appl. Psychol.* 43, 385–388.
- Lykken, D. T. (1974). Psychology and the lie detector industry. *Am. Psychol.* 29, 725–739.
- Lykken, D. T., and Venables, P. H. (1971). Direct measurement of skin conductance: a proposal for standardization. *Psychophysiology* 8, 656–672.
- Martens, R. (1969). Palmar sweating and the presence of an audience. *J. Exp. Soc. Psychol.* 5, 371–374.
- Matsuda, I., Nittono, H., Ogawa, T. (2012). Central and autonomic responses that reflect the intention to conceal in a concealed information test. *Int. J. Psychophysiol.* 85, 325.
- Meijer, E. H., Smulders, F. T. Y., Johnston, J. E., and Merckelbach, H. L. G. J. (2007). Combining skin conductance and forced choice in the detection of concealed information. *Psychophysiology* 44, 814–822.
- O'Gorman, J. G. (1990). Individual differences in the orienting response: nonresponding nonclinical samples. *Pavlov. J. Biol. Sci.* 25, 104–110.
- Orne, M. T. (1975). Implications of laboratory research for the detection of deception. *Polygraph* 2, 169–199.
- Rosnow, R. L., and Rosenthal, R. (1996). Computing contrasts, effect sizes, and counternulls on other people's published data: general procedures for research consumers. *Psychol. Methods* 1, 331–340.
- Saxe, L. (1991). Science and the CQT polygraph: a theoretical critique. *Curr. Dir. Psychol. Sci.* 33, 69–73.
- Schilbach, L., Wohlschläger, A. M., Krämer, N. C., Newen, A., Shah, J. N., Fink, G. R., et al. (2006). Being with virtual others: neural correlates of social interaction. *Neuropsychologia* 44, 718–730.
- Sokolov, E. N. (1963). Higher nervous functions: the orienting reflex. *Annu. Rev. Physiol.* 25, 545–580.
- Sproull, L., Subramani, M., Kiesler, S., Walker, J., and Waters, K. (1996). When the interface is a face. *Int. J. Hum. Comput. Interact.* 11, 97–124.
- Stern, R. M., Breen, J. P., Watanabe, T., and Perry, B. S. (1981). Effect of feedback of physiological information on responses to innocent associations and guilty knowledge. *J. Appl. Psychol.* 66, 677–681.
- Timm, H. W. (1982). Effect of altered outcome expectancies stemming from placebo and feedback treatments on the validity of the guilty knowledge technique. *J. Appl. Psychol.* 67, 391–400.
- Uzieblo, K., Verschuere, B., van den Bussche, E., and Crombez, G. (2010). The validity of the psychopathic personality inventory-revised in a community sample. *Assessment* 17, 334–346.
- Vandenbosch, K., Verschuere, B., Crombez, G., and De Clercq, A. (2009). The validity of finger pulse line length for the detection of concealed information. *Int. J. Psychophysiol.* 71, 118–123.
- Velden, M., and Wölk, C. (1987). Depicting cardiac activity over real time: a proposal for standardization. *J. Psychophysiol.* 1, 173–175.
- Verschuere, B. (2011). "Psychopathy and the detection of concealed information," in *Memory Detection*, eds B. Verschuere, G. Ben-Shakhar, and E. H. Meijer (Cambridge: Cambridge University Press), 215–230.
- Verschuere, B., and Ben-Shakhar, G. (2011). "Theory of the concealed information test," in *Memory Detection*, eds B. Verschuere, G. Ben-Shakhar, and E. H. Meijer (Cambridge: Cambridge University Press), 128–148.
- Verschuere, B., Crombez, G., De Clercq, A., and Koster, E. H. W. (2005). Psychopathic traits and autonomic responding to concealed information in a prison sample. *Psychophysiology* 42, 239–245.
- Verschuere, B., Crombez, G., Koster, E. H. W., and De Clercq, A. (2007). Antisociality, underarousal and the validity of the Concealed Information Polygraph Test. *Biol. Psychol.* 74, 309–318.
- Waid, W. M. (1976). Skin conductance response to both signaled and unsigned noxious stimulation predicts level of socialization. *J. Pers. Soc. Psychol.* 34, 923–929.
- Waid, W. M., and Orne, M. T. (1980). Individual differences in electrodermal lability and the detection of information and deception. *J. Appl. Psychol.* 65, 1–8.
- Waid, W. M., and Orne, M. T. (1981). "Cognitive, social, and personality processes in the physiological detection of deception," in *Advances in Experimental Social Psychology*, ed. L. Berkowitz (New York, London, Toronto, Sydney, San Francisco: Academic Press), 61–101.
- Waid, W. M., Orne, M. T., and Wilson, S. K. (1979). Effects of level of socialization on electrodermal detection of deception. *Psychophysiology* 16, 15–22.
- Zajonc, R. B. (1965). Social facilitation. *Science* 149, 269–274.
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 30 August 2012; accepted: 30 October 2012; published online: 19 November 2012.

Citation: Ambach W, Assmann B, Krieg B and Vaitl D (2012) Face and voice as social stimuli enhance differential physiological responding in a Concealed Information Test. *Front. Psychology* 3:510. doi: 10.3389/fpsyg.2012.00510
This article was submitted to Frontiers in Cognitive Science, a specialty of Frontiers in Psychology.

Copyright © 2012 Ambach, Assmann, Krieg and Vaitl. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.



A functional analysis of deception detection of a mock crime using infrared thermal imaging and the Concealed Information Test

Kevin K. Park¹, Hye Won Suk², Heungsun Hwang² and Jang-Han Lee^{1*}

¹ Clinical Neuro-psychology Lab., Department of Psychology, Chung-Ang University, Seoul, South Korea

² Quantitative Methods Lab., Department of Psychology, McGill University, Montreal, QC, Canada

Edited by:

Wolfgang Ambach, Institute for Frontier Areas of Psychology and Mental Health (IGPP), Germany

Reviewed by:

Martin P. Paulus, University of California San Diego, USA

Panagiotis Tsiamyrtzis, Athens University of Economics and Business, Greece

Lara Warmelink, University of Lancaster, UK

***Correspondence:**

Jang-Han Lee, Department of Psychology, Chung-Ang University, 221 Heukseok-dong, Dongjak-gu, Seoul 156-756, South Korea.

e-mail: clipsy@cau.ac.kr

The purpose of this study was to utilize thermal imaging and the Concealed Information Test to detect deception in participants who committed a mock crime. A functional analysis using a functional ANOVA and a functional discriminant analysis was conducted to decrease the variation in the physiological data collected through the thermal imaging camera. Participants chose between a non-crime mission (Innocent Condition: IC), or a mock crime (Guilty Condition: GC) of stealing a wallet in a computer lab. Temperature in the periorbital region of the face was measured while questioning participants regarding mock crime details. Results revealed that the GC showed significantly higher temperatures when responding to crime relevant items compared to irrelevant items, while the IC did not. The functional ANOVA supported the initial results that facial temperatures of the GC elevated when responding to crime relevant items, demonstrating an interaction between group (guilty/innocent) and relevance (relevant/irrelevant). The functional discriminant analysis revealed that answering crime relevant items can be used to discriminate guilty from innocent participants. These results suggest that measuring facial temperatures in the periorbital region while conducting the Concealed Information Test is able to differentiate the GC from the IC.

Keywords: deception detection, thermal imaging, mock crime, Concealed Information Test

INTRODUCTION

Deception detection is widely used by law enforcement around the world. Although very few countries actually allow the results to be used as evidence in court, investigators frequently use lie detecting as a tool of reference during investigations. Many forms of deception detection exist, but the polygraph is the most widely used method. Unfortunately, field studies have shown that polygraph testing accuracy is in the unsatisfactory range of 72–91% (National Research Council, 2003). Among the numerous reasons for the variability in accuracy, a main drawback of polygraph testing is its dependency on the level of training and experience of the polygrapher. In other words, the accuracy of a polygraph test is greatly affected by the subjective skill of the polygrapher. Also, polygraph testing in itself can cause high levels of anxiety in subjects, which can also affect the results or even lead to false-positive conclusions. It is therefore imperative that additional means of deception detection are developed, standardized, and applied as alternative methods, or at least as secondary support to the polygraph.

Deception detection using thermal imaging (a.k.a. thermography) incorporates an infrared thermal imaging camera to measure facial skin temperature as a cue to deception. Although not yet used in law enforcement, thermal image analysis for polygraph testing has already gained a US patent (Pavlidis, 2005; Patent No: US 6854879 B2), and has obtained empirical support from

previous research with results suggesting that it has the potential to detect deception quite accurately (Pavlidis et al., 2002; Pollina et al., 2006; Tsiamyrtzis et al., 2007; Dowdall et al., 2009). In general, when a deceptive subject is being interrogated, they experience stress which activates the autonomic nervous system. This then activates the sympathetic nervous system, which is responsible for stress responses such as increased blood flow to the eyes to facilitate rapid eye movement in preparing the body for the fight-or-flight response (Pavlidis and Levine, 2001, 2002). This increased blood flow is detectable in the periorbital region of the face through thermal imaging. The periorbital regions are the symmetrical areas to the left and right of the bridge of the nose between the eyes. Previous deception detection studies that used thermal imaging also did so by measuring the temperature of the periorbital regions of the face. In these studies, average facial temperatures collected from the periorbital regions were higher during deceptive responses, compared to non-deceptive responses, thus acting as cues to deception.

An outstanding advantage of using thermal imaging is that it is non-invasive, in that no sensors are attached to the subject (Arora et al., 2008). While typical polygraphs require numerous contact sensors, thermal imaging has none, making it more natural and comfortable. Research in psychophysiology has shown that contact sensors (i.e., polygraph sensors) can compromise comfort, which can effect physiological measurement (Yankee, 1965),

as well as deception detection procedures (Pavlidis and Levine, 2002). Another advantage is that raw thermal data can be saved for later analysis, in the case that a better more accurate analysis method is developed in the future (Pavlidis and Levine, 2001). In addition, thermal imaging cameras generally look like video cameras, meaning deception detection could take place without the subject even realizing it is happening, which can prevent unwanted attempts at countermeasures.

While detecting deception with thermal imaging has many advantages, it also has certain disadvantages, one of them being that it is sensitive to the environment and changes in the environment. In particular, it is sensitive to ambient temperatures and humidity levels (Hermans-Killam, 2002), as well as changes in the distance between the subject and the thermal imaging camera lens (Jones and Plassmann, 2002). Unlike measuring body temperatures to detect sick people at an airport, detecting deception must measure very small changes in skin surface temperature, and therefore such sensitivity may have a critical effect on the measurement results. Therefore, in order to control for these possible variables, the present study conducted the thermal imaging measurements in a highly controlled experimental environment.

To detect deception, whether using a polygraph or thermal imaging, a method of questioning is needed. Although in the field the Control Question Test (CQT; Reid, 1947) is the most widely used questioning technique (Meijer and Verschueren, 2010), it is criticized by researchers for its lack of theoretically based empirical evidence (Ben-Shakhar, 2008; Iacono, 2008). Unlike the CQT, the Concealed Information Test (CIT; a.k.a. Guilty Knowledge Test or GKT; Lykken, 1959) is empirically supported as a physiologically sound method of questioning (Ben-Shakhar and Furedy, 1990; Elaad, 1998; MacLaren, 2001; Ben-Shakhar and Elaad, 2003). The present study detected deception under controlled experimental conditions, and therefore utilized the CIT instead of the CQT to maintain a theoretically based experimental process of deception detection. In addition, because the second experimenter (the interviewer) was not trained in interrogation, the CIT is ideal in that it is a standardized, easily replicated procedure that does not require professional training, as does the CQT (Ben-Shakhar and Elaad, 2002).

Unlike field studies where the interviewees are suspects to actual crimes, participants in experiments are typically average citizens or students, and therefore a mock crime is needed. Guilty participants commit a crime, and innocent participants enact a similar non-criminal task, or are simply given information about the crime. To motivate participants and provide an incentive to be judged innocent, they are given a reward (e.g., monetary compensation, academic credits) upon successful deception, or punishment (e.g., monetary penalties, academic tasks) for failing to deceive. In a meta-analytic study of mock crime research, the incentive to motivate deception was a main variable that affected the outcome of deception detection (Kircher et al., 1988). Therefore, in the present study, participants in the guilty condition were told they would receive triple the original participation fee upon success, but would receive nothing if they failed, incorporating both award and punishment.

To further increase anxiety during the mock crime, guilty participants were to commit theft and eliminate evidence of their crime in a public computer lab. Innocent participants had to go to the same computer lab and send out an email, which allowed the blind experimenter to ask questions that were relevant to both groups, but only the guilty participants would possess crime-relevant information. Further details regarding the mock crime scenario are explained in the method section.

As with most physiological data, skin surface temperatures measured using thermal imaging could be thought of as functional data, and was therefore further analyzed using a functional ANOVA (Ramsay and Silverman, 2006) and a functional discriminant analysis (Ramsay and Silverman, 2006). An important property that distinguishes functional data from multivariate data is the existence of a smooth curve assumed to generate the data. Functional data assumes that an underlying function gives rise to the observed data, and that the underlying function is smooth so that adjacent data values tend to be similar to some extent and not too different from each other. In other words, adjacent data values provide overlapping information, not independent information.

A functional ANOVA is a functional extension of an ANOVA, in which the response variable is a function and predictor variables are categorical. A functional ANOVA was used to see if facial temperatures of participants were affected by guilt or innocence and/or whether they were answering crime-relevant or crime-irrelevant questions. The functional discriminant analysis is a functional version of Fisher's linear discriminant analysis, which seeks to find components, or weighted integrations of functions that separate multiple groups of observations as much as possible. The functional discriminant analysis was used to see how well the facial temperature data was able to differentiate guilty participants from the innocent participants. Further details regarding the functional ANOVA and the functional discriminant analysis are explained in the method section and the appendix of this study.

The aim of the present study was to detect deception in participants who conducted a realistic mock crime using infrared thermal imaging and a simplified facial tracking method, along with the CIT method of questioning. The purpose of the study was to (a) detect deception using thermal imaging through a simplified method, (b) in a more controlled environment, (c) using the most realistic mock crime possible, (d) using the most optimal method of statistical analyses, and (e) replicate the results of the previous studies that have done so in the past. It was predicted that the guilty participants would be differentiable from the innocent participants, in that the guilty condition would show an increase in facial temperatures of the periorbital regions when responding to crime-relevant sub-questions compared to the irrelevant sub-questions of the CIT, while the innocent condition (IC) would show no significant difference between the two. It was also predicted that using the same thermal imaging data, a functional ANOVA would reveal similar results, supporting the initial analysis, and also that a functional discriminant analysis would be able to differentiate the guilty from the innocent participants from their facial temperatures in the periorbital regions.

METHOD

PARTICIPANTS

A total of 34 participants were recruited from an online bulletin board on a university website. The bulletin board entry stated that participants were being recruited for a psychology experiment on measuring facial temperatures using thermal imaging, and would be paid \$10 for their participation. All participants read and signed a written consent form agreeing to participate in the experiment. One participant was unable to finish the experimental procedure, and the thermal imaging data from three participants was incomplete and had to be discarded. This left the data of 30 participants (17 male, 13 female), between the ages of 18 and 30 ($M = 22.74$, $SD = 2.77$), for the final data analyses.

MATERIALS

Apparatus

Thermal imaging. To record the facial temperatures of the participant's faces during the experiment, an Infrared Thermography H2640 infrared thermal imaging camera (NEC Avio Infrared Technologies Co. Ltd., Japan) with 320×240 pixel resolution and heat resolution of 0.08°C ($\pm 2\%$ accuracy) at 30 Hz mounted on an industrial strength tripod (SLIK Corporation, Japan) was used. The thermal imaging camera was placed so that the lens of the camera was 100 cm (± 1 cm) from the participant's face, which is the distance that the thermal imaging camera manufacturer suggested as the optimal recording distance for measuring human skin surface temperatures. The thermal imaging camera was connected to an Xnote P300-TP8WK laptop computer (LG Electronics, Korea). A digital thermometer/hygrometer was placed directly under the thermal imaging camera, and experiments were conducted at a constant room temperature of 21.0°C ($\pm 0.25^\circ\text{C}$) and 65% ($\pm 2\%$) humidity.

Webcam. To provide a CCTV security camera at the computer lab where the mock crime would be taking place, a Quick Cam® Ultra Vision SE webcam (Logitech, USA) was mounted at the front of the computer lab. The webcam was connected to a desktop computer at the desk where a confederate acting as the computer lab assistant was sitting. This webcam not only acted as a CCTV security camera which the participants conducting the mock crime had to deactivate, it also allowed the experimenters in the psychology laboratory to view what was happening in the computer lab while the mock crime was taking place.

The red wallet. A bright-red, faux leather, woman's wallet with gold-plated trimming was used as the target object that the participants conducting the mock crime had to steal. The wallet was a three-way folding style wallet with a few credit cards, some business cards, and some monetary bills placed in it to make it look and feel as realistic as possible.

Health questionnaire

A short questionnaire was designed to ask participants whether they were sick, taking any kind of medication, had any history of thyroid problems which may affect body temperature control, or were currently visiting the hospital for any of the above

reasons. This questionnaire was conducted before the experiment to screen out any possible participants who may not show "normal" physiological or temperature related responses to the experimental procedures.

Concealed Information Test

While recording the thermal imaging data, participants were asked a series of questions to detect deception. Each question begins with a main primary question, followed by a series of five secondary sub-questions containing different possible answers to the original main question. For example, a main question was "What was the item you stole from the computer lab?" and was followed by sub-questions such as "Was it a watch?" "Was it a ring?" and "Was it a wallet?" According to the theories underlying the CIT, if the participant did actually steal the wallet, then he or she would have critical knowledge regarding the mock crime, which in this case would be the wallet. Thus, when a guilty participant is asked if the item stolen was a wallet, their facial temperature response to this sub-question would be different from the other sub-questions presented. Three main questions were asked, but there were only two actual questions. The first question was asked a second time after the second question to conform to the standard practice of the CIT, with the sub-questions being in a different order from the first time to eliminate any ordering effects (see Appendix section "Concealed Information Test Questioning Protocol" for the questioning protocol). The sub-questions that were not relevant to the crime are irrelevant items (IR), and the crime relevant sub-questions are relevant items (RE). After each sub-question was asked, a period of 10 s was allowed to pass before the next question was asked. This was to allow the participants' facial temperatures to recover from any fluctuations that may have occurred from the previous question. The questioning session lasted approximately 6–8 min.

PROCEDURE

Upon arrival, participants first read and signed a written consent form, followed by the health questionnaire. The experimenter then entered the room carrying two colored envelopes and some yellow Post-it notes. The experimenter explained the experiment and what the participant would have to do (see Appendix section "Envelope Selection Explanation"). The participants then chose whether they would conduct the guilty (red) or innocent (blue) mission. The experiment was purposefully designed this way to increase the level of involvement the participants would feel. Considering most crimes are committed by choice, and not by force, it was expected that this method would produce a more realistic response during deception detection. Participants were informed that if they successfully accomplished the red mission, they would receive a reward three times the amount of the original \$10 participation fee of \$30, but if they failed they would receive nothing. Here, success meant stealing the wallet and eliminating the evidence without being discovered or questioned by anyone, and not being detected of deception by the second experimenter (interrogator). They were also informed that if they successfully accomplished the blue mission, they would receive a \$10 participation fee. Here, success meant sending an email at the computer lab, leaving a message to the computer lab assistant, and not

being accused of deception by the second experimenter. Although the participants were informed that failing the missions would result in receiving nothing, all participants were to receive a small token of appreciation of \$5 for participating in the experiment, regardless of the outcome.

The participants selected an envelope, read the instruction sheet inside, and were also given a small Post-it note to write down whatever information they thought was important in completing the mission. The Post-it notes were small and did not have enough space to copy all the mission information verbatim. This forced participants to summarize the information in their own words. It was expected that this process would further increase personal involvement in the mock crime, leading to a feeling of having planned certain aspects of the mission themselves, or a feeling of having taken part in the plotting of the crime. The mission documents contained the following instructions:

Blue mission:

1. Go to the Computer Lab located on the 5th floor of this building.
2. Sit at any seat and send an email to kekipark@hotmail.com.
3. In the body of the email, write your name, age, and mobile phone number.
4. When the Computer Lab Assistant leaves his seat, go to his computer and leave a note using the Notepad program on his computer saying "Hello, I used one of the computers. Thank you."
5. Return to the Psychology Laboratory.

Red mission:

1. Go to the Computer Lab located on the 5th floor of this building.
2. Locate the girl sitting in the 3rd row from the front of the room, and the 4th seat from the windows.
3. When the girl leaves her seat, go to her seat and steal a red wallet from her bag.
4. When the Computer Lab Assistant leaves his seat, go to his computer and shutdown the webcam program that is running on the desktop. This will automatically delete the video file being saved.
5. Return to the Psychology Laboratory.

After participants finished writing their notes, they were given detailed explanations on how to get to the computer lab. They were then told to go the computer lab, execute the mission, and immediately return as soon as they were done.

Upon returning, participants were asked if they had successfully completed their mission, and guilty participants were asked for the wallet. The participants were then taken to a temperature and humidity controlled measurement room where the thermal imaging camera was set up. A second experimenter blind to the participants' mission selection informed the participants that although she was aware that a crime had taken place in the computer lab, she had no knowledge of who the perpetrator was. She then explained that she would ask a series of questions in an attempt to figure out whether the participant committed the crime or not.

Before questioning, participants relaxed for 2 min to adjust to the room. Afterwards, the first experimenter came into the room to adjust the thermal imaging camera to record a 1 min baseline reading. The participants were told that the camera was a video camera, and that the interview would be recorded and later analyzed, so to remain as motionless as possible during questioning, and to maintain eye-contact with the camera until the questioning ended. The second experimenter was seated facing the participant at a right angle and was outside the field of view of the participant. As questioning began, the first experimenter began recording the thermal data from outside the room with a laptop computer connected to the thermal imaging camera. The entire experiment lasted approximately 45 min to 1 h, including the questionnaires, explanation and task selection, the mock crime, and the questioning session. When finished, participants were thanked, debriefed, and asked not to disclose any information regarding the experiment until the end of the experiment period, to prevent contaminating future participants.

EXPERIMENTAL DESIGN

There were two experimental conditions in this study: 18 participants in the Guilty Condition (GC) which selected the red envelope and committed a mock crime, and 12 participants in the IC which selected the blue envelope and acted as the control condition. Therefore, in order to differentiate which participants were in the GC and which were in the IC, the average of the maximum temperature values in the periorbital region while responding to the RE questions were compared to the values while responding to the IR questions. Although the first primary question was asked twice, and the second primary question asked once, each repetition of the first primary question was treated as an individual primary question in the analysis. Therefore, there were a total of three primary questions in the analysis. The mean temperature values of the RE and the IR items for each condition were compared using a paired-samples *t*-test. A significant increase in mean temperature value for the responses to the RE compared to the IR sub-questions would signal that the participant possessed concealed knowledge regarding the mock crime.

DATA COLLECTION AND ANALYSES

Initial analysis

The thermal image data used to analyze the facial temperature readings were collected from the periorbital region of the face. This is the area between the eye and the bridge of the nose on either side of the nose. As shown in **Figure 1**, an area of interest (AOI) was designated to cover the periorbital regions, but not the actual eye itself. An AOI is an area designated by the user of the thermal imaging software from which maximum or minimum temperatures are collected and analyzed. AOIs are designated in order to avoid including areas of the face which are always the hottest regions regardless of the situation, such as the eye sockets and the inside of the mouth. The maximum temperature point within the AOI was recorded during each frame of recording (30 frames per second). The mean temperature value corresponding to each response was the average of the maximum temperature point during the 10 s of response time given after each sub-question was asked. The 10 s of response time

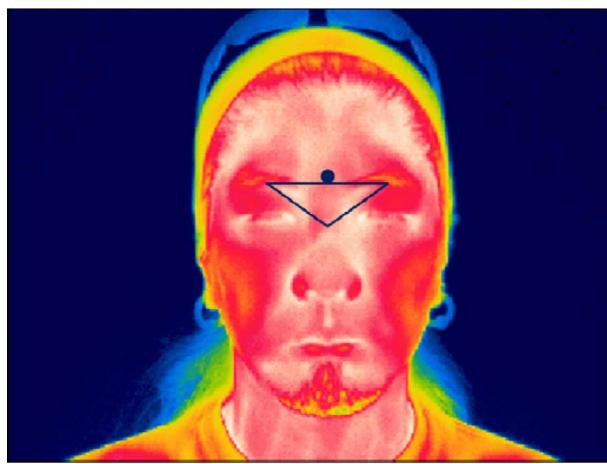


FIGURE 1 | A sample thermal image showing the Area of Interest (AOI) covering the periorbital regions (left and right corners of the inverted triangle) and the metallic tracking sticker (black dot).

started at the end of the last word of each sub-question. When an AOI is selected, the thermal imaging software automatically tracks this designated region of the face and follows it when the participant moves their face. However, to increase tracking accuracy, a small metallic sticker was placed above the bridge of the nose which in thermal imaging appears as a black dot relative to the skin (see **Figure 1**). Therefore, when the AOI was set to follow the black dot, tracking was extremely accurate as long as the participants did not tilt their head from side to side at an angle or turn their head to the left or right. None of the participants tilted or turned their heads during measurement. The point of maximum temperature was always measured from within the designated AOI.

To compare mean facial temperature values between conditions, independent-samples *t*-tests were conducted, and to compare between RE and IR sub-questions, paired-samples *t*-tests were conducted, all using SPSS 17.0 for Windows.

Functional ANOVA and functional discriminant analysis

Facial temperatures were measured for a duration of 10 s beginning after each question was posed by the experimenter. Therefore, the data consisted of 450 time series, or functions (30 participants \times 15 questions), measured over 300 time points (10 s \times 30 Hz), and three participants were eliminated from the analysis due to severe noise in their signals. Due to the limitation of computational power, the number of time points needed to be reduced to conduct the functional ANOVA, and therefore one of every five time points was used so that the number of time points per question was decreased to 60 (10 s \times 6 Hz). A total of 450 functions measured over 60 time points were analyzed. **Figures 2A,B** display the raw data of one guilty subject (subject 1) measured while answering three relevant questions and 12 irrelevant questions, respectively. Similarly, **Figures 2C,D** show the raw data of one innocent subject (subject 3) measured while answering three relevant questions and 12 irrelevant questions, respectively. Before any analyses were

conducted, the original functions were smoothed by the roughness penalty smoothing method with $\lambda = 10$ (see Appendix section “Smoothing: Roughness Penalty Smoothing Method” for more details on smoothing and Appendix section “Functional ANOVA” for details on the functional ANOVA). **Figure 3** displays the smoothed data corresponding to the raw data shown in **Figure 2**.

The functional discriminant analysis estimates a weight function, instead of a vector of weight, which separates multiple groups of functions as much as possible (see Appendix section “Fisher’s Linear Discriminant Analysis” for the technical details of the Fisher’s linear discriminant analysis and Appendix section “Functional Discriminant Analysis” for details on the functional discriminant analysis applied). The data used in the functional discriminant analysis consisted of 450 time series, or functions, (30 participants \times 15 questions) measured over 300 time points (10 s \times 30 Hz). The data measured for RE questions and IR questions was analyzed separately.

RESULTS

HEALTH QUESTIONNAIRE

No participant reported any medical problems in the Health Questionnaire.

BASELINE FACIAL TEMPERATURES

An independent-samples *t*-test revealed no significant differences in baseline facial temperature readings between the GC ($M = 35.83$, $SD = 0.59$), and the IC ($M = 36.03$, $SD = 0.78$), $t_{(28)} = -0.78$, $p = 0.44$. There were also no significant differences between male and female participants, or between their ages. These results show that there was no significant facial temperature difference between the participants in the GC and the IC before the experiment began.

FACIAL TEMPERATURE CHANGE VALUES

The thermal imaging camera measured temperatures at 30 frames per second. For each sub-question asked, temperature values of the hottest point within the AOI were recorded for a period of 10 s starting at the moment the experimenter ended her question. These temperature values were averaged, resulting in a mean facial temperature value for the RE items and the IR items for each participant. To obtain a facial temperature change value (FTCV) for the RE and IR items, baseline facial temperatures was subtracted from the mean temperature values.

Before performing paired-samples *t*-tests as described in the following sections, the normality assumption was tested which should be satisfied for a paired-samples *t*-test to be conducted. First, scatter plots, Q-Q plots, and boxplots of the FTCV scores of the four conditions (GC-RE, GC-IR, IC-RE, and IC-IR) were examined, and are presented in **Figures 4, 5, and 6**. As shown in **Figures 5** and **6**, the FTCV scores of subject 31 of the GC for both the RE and IR questions seemed to be deviated from normal distributions. Therefore, the Shapiro-Wilk test was performed to statistically test the null hypothesis that the FTCV scores in each of the four conditions came from a normal distribution. Results of the Shapiro-Wilk tests were not significant [for GC-RE, $W_{(18)} = 0.94$, $p = 0.33$; for GC-IR, $W_{(18)} = 0.94$, $p = 0.30$; for IC-RE,

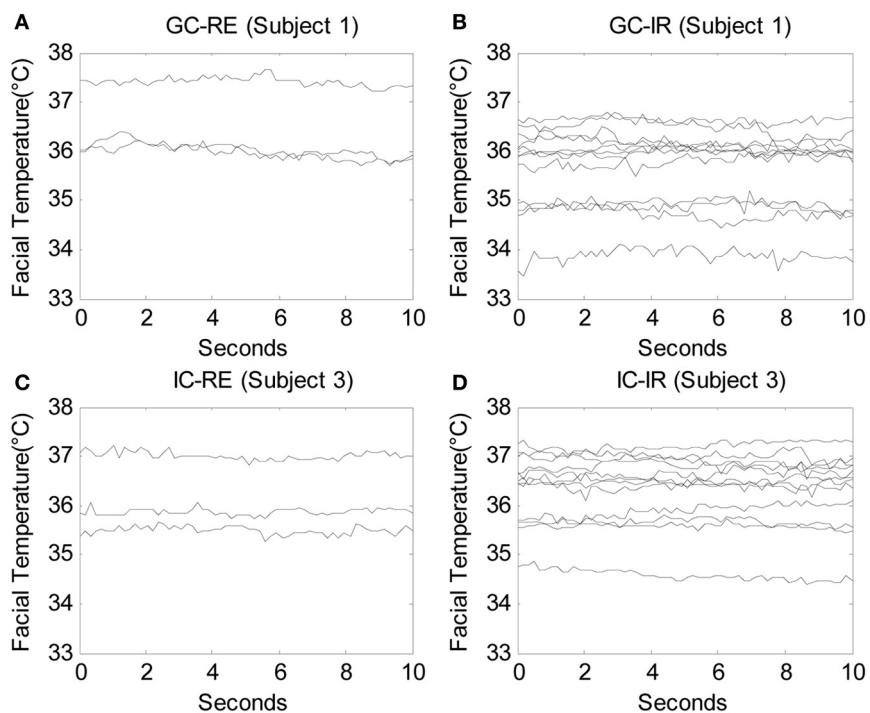


FIGURE 2 |The raw facial temperature data of one guilty subject (subject 1) answering (A) 3 relevant questions and (B) 12 irrelevant questions, and those of one innocent subject (subject 3)

answering (C) 3 relevant questions and (D) 12 irrelevant questions, in which each line indicates the facial temperature for each question.

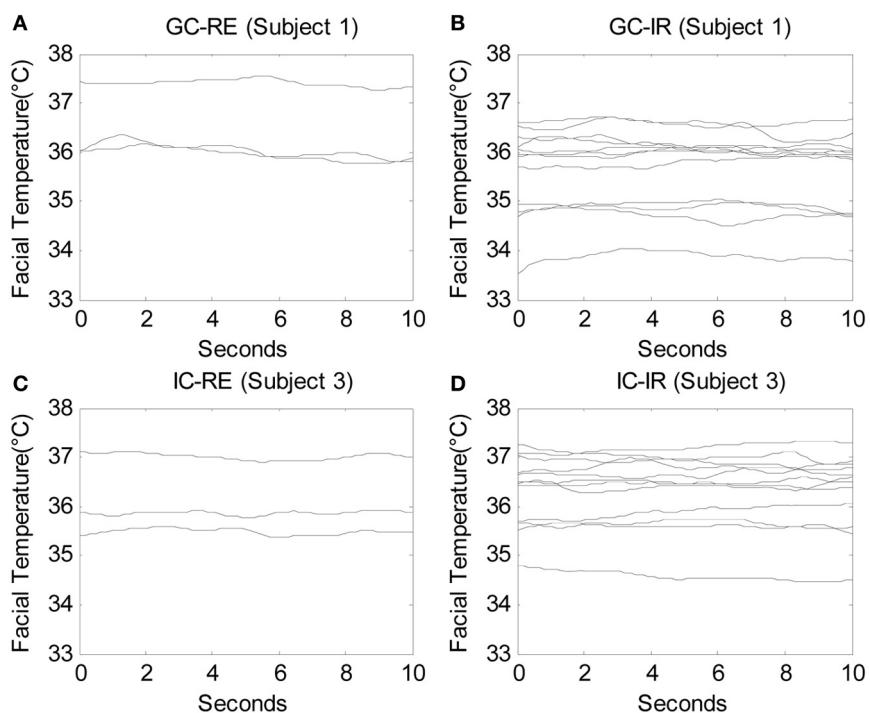


FIGURE 3 |The smoothed facial temperature data under $\lambda = 10$ of one guilty subject (subject 1) answering (A) 3 relevant questions and (B) 12 irrelevant questions, and those of one innocent subject

(subject 3) answering (C) 3 relevant questions and (D) 12 irrelevant questions, in which each line indicates the facial temperature for each question.

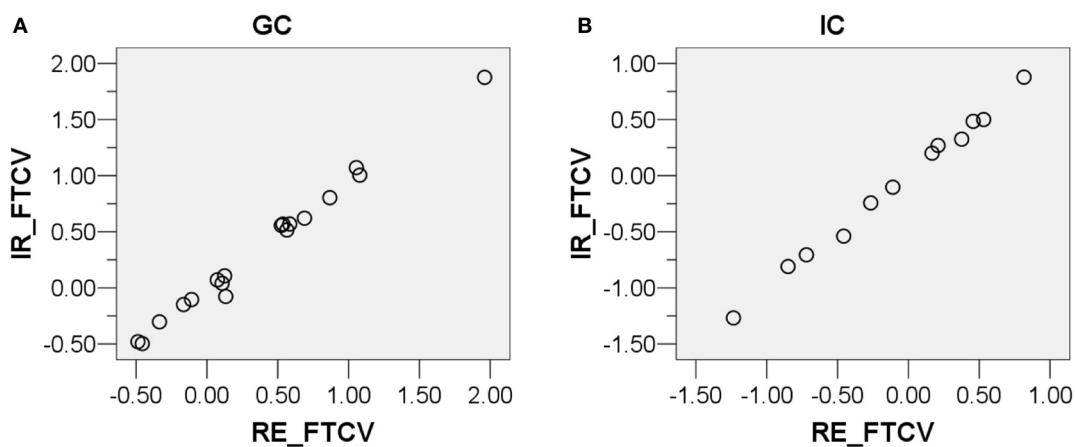


FIGURE 4 | Scatter plots of the distributions of the facial temperature change values of the (A) Guilty Condition responding to relevant sub-questions (X-axis) and irrelevant sub-questions (Y-axis), and (B)

Innocent Condition responding to relevant sub-questions (X-axis) and irrelevant sub-questions (Y-axis). The facial temperature change value in degrees Celsius ($^{\circ}\text{C}$).

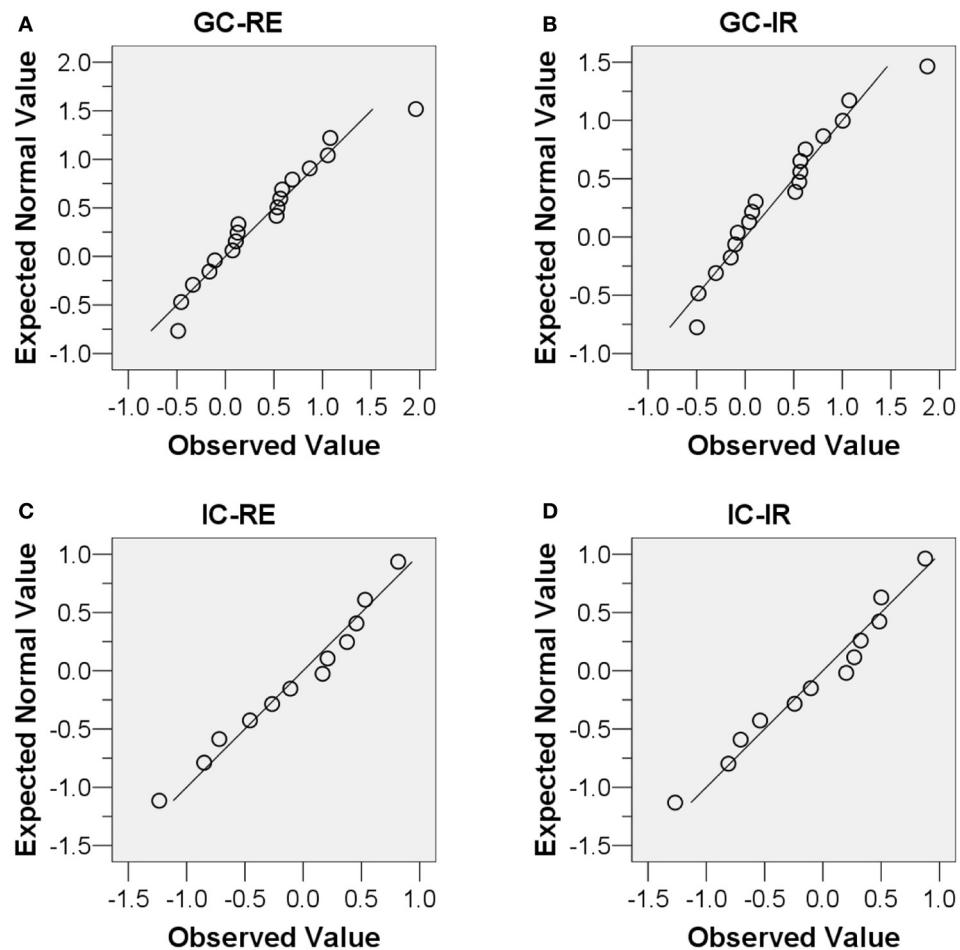


FIGURE 5 | Normal Q-Q plots of the four conditions. (A) Guilty Condition responding to relevant sub-questions, (B) Guilty Condition responding to irrelevant sub-questions, (C) Innocent Condition responding to relevant sub-questions, and (D) Innocent Condition responding to irrelevant sub-questions.

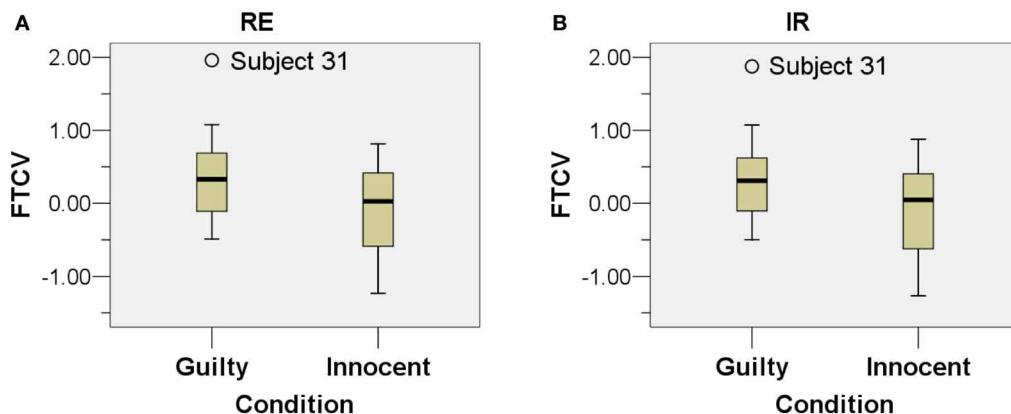


FIGURE 6 | Boxplots of the two types of questions. (A) relevant sub-questions and **(B)** irrelevant sub-questions. The X-axis indicates the condition (Guilty Condition and Innocent Condition) and the Y-axis indicates the facial temperature change value in degrees Celsius ($^{\circ}\text{C}$).

$W_{(12)} = 0.96, p = 0.84$; for IC-IR $W_{(12)} = 0.96, p = 0.82$], indicating that the normality assumption was not violated in any of the four conditions.

GUILTY CONDITION

A directional paired t -test for the GC revealed a significant difference between the FTCVs for the RE questions ($M = 0.40, SD = 0.62$) and the IR questions ($M = 0.37, SD = 0.61$), $t_{(17)} = 1.91, p < 0.05$. However, when utilizing the Bonferroni correction to control for an experimentwise error rate, the results were no longer significant and only showed a trend ($p < 0.10$) toward temperature responses to the crime relevant sub-questions being higher than the temperature responses to the crime irrelevant sub-questions.

INNOCENT CONDITION

A directional paired-samples t -test for the IC revealed no significant difference between the FTCVs for the RE questions ($M = -0.17, SD = 0.59$) and the IR questions ($M = -0.17, SD = 0.59$), $t_{(11)} = -0.04, p > 0.05$. As expected, there were no differences in the temperature responses to crime relevant and irrelevant sub-questions.

The above results show that there was no significant difference in FTCV values between RE and IR responses in the IC, yet there was a noticeable trend in the values between RE and IR responses in the GC. These analyses were conducted using t -tests which analyze the data by comparing mean values. However, the data of the present study are time-based values, and a comparison of means may have been a meticulous enough approach. Important information may have been lost or overlooked during the process of averaging out this chronological data. Mean values summarize the data measured over continuous time points as mingle measures, and it may not, in this case, have been enough to consider only mean values to capture all of the characteristics that reflect a group difference. Therefore, a functional ANOVA, which uses all of the values measured in its analysis, was utilized to evaluate all of the existing data in its entirety in greater detail, as well as prevent any

loss of information that may have occurred from a simple mean comparison.

FUNCTIONAL ANOVA AND FUNCTIONAL DISCRIMINANT ANALYSIS

The researchers who conducted the additional analyses did not participate in the actual experiment, and were only provided with the raw thermal data. This eliminated any researcher biases that may have affected the results.

The functional ANOVA examined the main effect of condition (GC/IC), the main effect of relevance (RE/IR), and the interaction effect of condition and relevance. **Figure 7A** presents the mean facial temperature over the 10 s for the four different conditions. From top to bottom, the four lines indicate GC-RE, IC-IR, IC-RE, and GC-IR. We can see that guilty participants manifested higher facial temperature for RE questions than IR questions. **Figure 7B** shows the significant main effect of condition, which indicates that guilty participants manifested lower facial temperatures when answering IR questions over the 10 s by around 0.55°C . **Figure 7C** presents the significant main effect of relevance, which indicates that innocent participants manifested lower facial temperatures when answering RE questions compared to IR questions over the 10 s by around 0.33°C . **Figure 7D** shows that the interaction effect of condition and relevance was significant, which indicates that the facial temperature of guilty participants answering RE questions was significantly higher than what could be predicted from the sum of the two main effects by 0.9°C . This suggests that facial temperature is affected by the interaction between condition (GC/IC) and relevance (RE/IR), meaning that guilty participants showed higher facial temperature when answering RE questions than IR questions whereas innocent participants did not.

The functional discriminant analysis analyzed 90 functions ($30 \text{ participants} \times 3 \text{ relevant questions}$) for the RE questions based on a weight function estimated with penalty parameter $\rho = 10$ determined by the leave-one-out cross-validation. Before the analysis, each function was baseline corrected by subtracting the corresponding baseline temperature. The 90 functions

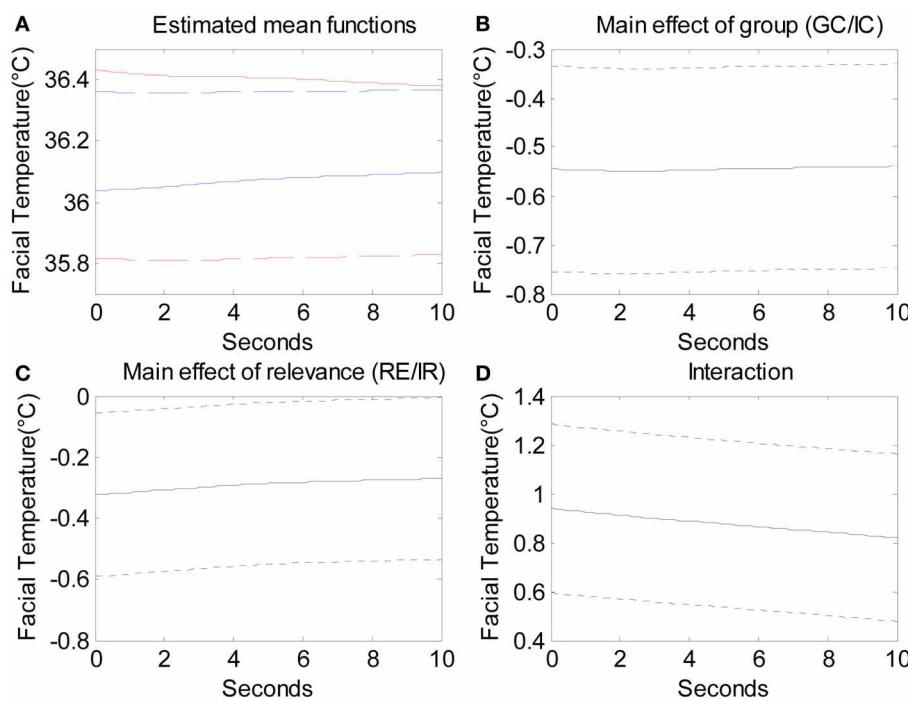


FIGURE 7 | (A) The estimated mean functions of the four conditions. From the top, solid red line indicates the mean function of GC-RE, blue dashed line IC-IR, blue solid line IC-RE, and red dashed line GC-IR. **(B)** The main effect of condition (GC/IC; solid line) with 95% pointwise confidence

interval (dotted lines). **(C)** The main effect of relevance (solid line) with 95% pointwise confidence interval (dotted lines). **(D)** The interaction effect of condition and relevance (solid line) with 95% pointwise confidence interval (dotted lines).

were classified into two groups based on the weight function, and 98.89% (89 out of 90) were correctly classified (misclassification rate = 1.11%). This result indicates that facial temperatures measured while answering a RE question can be used to differentiate whether a participant is in the GC or IC.

For IR questions, 360 functions (30 participants \times 12 irrelevant questions) were analyzed based on a weight function estimated with penalty parameter $\rho = 10^6$ which was also determined by the leave-one-out cross-validation. Again, before the analysis, each function was baseline corrected. When the 360 functions were classified into two groups based on this weight function, 68.89% (248 out of 360) were correctly classified (misclassification rate = 31.11%) which is only slightly higher than chance. This result indicates that facial temperatures measured for IR questions does not effectively discriminate guilty and innocent participants.

DISCUSSION

The present study utilized infrared thermal imaging with the CIT to detect the deception of participants who committed a mock crime. However, because there are certain limitations in using thermal imaging in the field, such as environmental factors and participant movement, the present study aimed to overcome these limitations by conducting a laboratory based experiment that would control for such variables. In accordance to conducting a lab based study, the present study further utilized deception detection techniques that were best suited for

research purposes. One of which was to use the CIT method of questioning, which is based on empirical evidence, and the other being a highly realistic mock crime scenario. In addition, a new and simple means of tracking the facial movement of the participants during thermal image measurement to minimize temperature variances due to head movement was also developed.

Results revealed that the average maximum skin surface temperatures recorded in the periorbital regions of the guilty participants were, as expected, significantly higher while responding to RE items compared to IR items. In contrast, and also as expected, there were no significant temperature differences between the RE and IR items measured from the innocent participants. These results are in line with the previous results of studies which used thermal imaging to detect deception (Pavlidis et al., 2002; Pollina et al., 2006; Tsiamyrtzis et al., 2007; Dowdall et al., 2009). However, the facial tracking process necessary to accurately measure facial skin temperatures used in the present study was drastically simplified in comparison to those used in previous research. Instead of relying on high-tech computer programming, a more analogue method of tracking was developed and was successfully applied.

The results of this study support past research that the CIT is indeed an effective method of questioning for deception detection, assuming the appropriate circumstances apply, which in this case was that the interviewer possessed information regarding evidence that only the guilty participants knew, and the

innocent participants did not. The CIT was conducted with no pre-interview or any other type of interviewee preparation, other than informing the subject that they would be asked a few questions regarding a crime that had been committed. This allowed for an extremely short questioning session, the interviewer needed no information about the participant to conduct the session, no pre-interview or rapport building was necessary, and the interviewer needed no special training to conduct the questioning session. Therefore, when applicable, the CIT seems to be a much more efficient means of questioning than the CQT.

The mock crime used in the present study was also highly effective at making the participants feel as if they were actually committing a crime. How the participants felt during the mock crime was not systematically measured, yet it was clear to the experimenter that most of the participants were highly anxious about conducting the mock crime, as well as receiving the deception detection procedure. Examples of this were, but not limited to, participants' hands shaking when they returned from conducting the mock crime, participants not being able to steal the wallet and returning empty handed (but eventually going through with it), and in one extreme case the participant gave up and decided not to participate in the study after attempting the mock crime. The combination of the public location, having to dig through a stranger's bag for a wallet, and having to eliminate evidence at the computer lab assistant's computer seemed to have provided enough immersiveness to make the participants believe they were actually doing something illegal.

In addition to the initial statistical analyses, additional analyses were conducted to further examine the results which revealed only a trend toward the predictions of the present study. As predicted, and in line with the trend found in the initial analyses, the additional analyses conducted using a functional ANOVA were able to show that facial temperatures in the periorbital regions of guilty participants were significantly higher while responding to RE questions compared to IR questions, but not in the innocent participants. This result not only supports the results of previous studies, but also increased the ecological validity of the experiment by displaying consistent results even when analyzed through different statistical methods by researchers who did not participate in the experiment itself. A functional discriminant analysis was also able to discriminate between the guilty and innocent participants at a classification rate of 98.89%. This result provides support for the potential that thermal imaging has in detecting deception, or at the very least supplementing existing methods of deception detection to increase their accuracy.

Certain limitations applied to the present study. First, the thermal imaging camera used was not the highest resolution camera available. There are other thermal imaging cameras currently available with greater resolution, which may produce more accurate measurements. Second, the participants were given a choice to choose between the GC and IC in order make the mock crime scenario more immersive. Although there were no significant differences in age, gender, health, or baseline temperatures between the two conditions, it is possible that other dissimilarities may have had an effect on the results, such as personality differences or intelligence. Had such information been measured prior to the condition selection procedure, it could have provided valuable

information as to which participants chose the guilty condition and how they may have differed from the participants in the IC. Third, although the thermal imaging procedure was non-invasive compared to all the sensors of a polygraph, due to the fact that the participants were told not to move and maintain eye-contact with the camera during the questioning session, and that they had to have a small metallic sticker placed on their forehead, the procedure was not totally free of constraints. To overcome this limitation, the development of an advanced tracking method will be necessary. Such a tracking method would allow for a more realistic study where the participants would be able to move freely during measurement. A more sophisticated tracking method could also prevent any changes in ratio between the AOI and the size of the participants' thermal image from moving back and forth in relation to the thermal imaging camera. Fourth, the study was conducted during the middle of summer, which may have led to less emphasized temperature differences between the innocent and guilty participants. In other words, the entire sample's baseline temperatures may have been higher than normal, leading to smaller increases in temperatures for the deceptive participants' facial temperature responses. Fifth, the number of participants in the study was relatively small. Even though the number was sufficient to conduct the statistical analyses without technical issues, future research should increase the number of participants to further increase the reliability of the results. Sixth, the participants were allowed to choose whether they wanted to engage in a mock-crime involving monetary risk, or a relatively risk-free task. The present study was conducted this way to further immerse the guilty participants into feeling as if they were really involved in the crime. Although the participants were random university students, allowing them to choose their own task forced the study to sacrifice a certain amount of control afforded by random allocation. However, in reality, most criminals decide for themselves whether they should or should not commit criminal behavior, and therefore this freedom of choice may have increased ecological validity. A final limitation is that the present study used the value of the hottest single pixel of each frame from within the AOI for the analyses and from a statistical point of view this is not a very robust approach.

The results of the present study have demonstrated three main findings. First, it has provided support for previous studies that have utilized thermal imaging to detect deception, but in an experimental environment further controlling for temperature, humidity, and unnecessary body movement, in a much more simple and effective manner. Second, it has provided support for previous studies that claim the CIT is a more efficient questioning method requiring little to no training. Finally, a mock crime was designed that seems highly effective at providing a realistic crime experience, without placing anyone involved at risk or danger.

In conclusion, the present study has shown that using thermal imaging to detect deception has realistic and applicable potential to be utilized in modern day law enforcement. However, standardization of the equipment, methodology, and data analysis techniques are necessary before any kind of field application can be expected. Future research on deception detection using thermal imaging should place emphasis on three areas. First, developing a more advanced facial tracking method. Second, a

simpler way of analyzing the thermal data collected, to make detecting faster and more accurate, yet easier to apply to real-world circumstances. Third, conducting research using the most high-resolution thermal imaging equipment. This may produce not only more accurate results, but even allow for the discovery of previously unknown physiological changes in facial skin temperatures or facial temperature changing regions that can also act as cues during deception detection. Finally, taking a more robust

approach in the statistical analyses of the maximum temperature values by analyzing not one pixel, but an area of pixels from the thermal images.

ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government MEST, Basic Research Promotion Fund (NRF-2011-013-B00117).

REFERENCES

- Arora, N., Martins, D., Ruggerio, D., Tousimis, E., Swistel, A. J., Osborne, M. P., et al. (2008). Effectiveness of a noninvasive digital infrared thermal imaging system in the detection of breast cancer. *Am. J. Surg.* 196, 523–526.
- Ben-Shakhar, G. (2008). The case against the use of polygraph examinations to monitor post-conviction sex offenders. *Legal Criminol. Psychol.* 13, 191–207.
- Ben-Shakhar, G., and Elaad, E. (2002). “The guilty knowledge test (GKT) as an application of psychophysiology: future prospects and obstacles,” in *Handbook of Polygraph Testing*, ed M. Kleiner (San Diego, CA: Academic Press), 87–102.
- Ben-Shakhar, G., and Elaad, E. (2003). The validity of psychophysiological detection of information with the Guilty Knowledge Test: a meta-analytic review. *J. Appl. Psychol.* 88, 131–151.
- Ben-Shakhar, G., and Furedy, J. J. (1990). *Theories and Applications in the Detection of Deception: A Psychophysiological and International Perspective*. New York, NY: Springer-Verlag Publishing.
- Craven, P., and Wahba, G. (1979). Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer. Math.* 31, 377–403.
- de Boor, C. (2001). *A Practical Guide to Splines, Revised Edn*. New York, NY: Springer.
- Dowdall, J., Pavlidis, I., and Tsiamyrtzis, P. (2009). “Coalitional tracker for deception detection in thermal imagery,” in *Augmented Vision Perception in Infrared: Algorithms and Applied Systems, Advances in Pattern Recognition*, ed R. I. Hammoud (London: Springer-Verlag), 113–137.
- Elaad, E. (1998). The challenge of the concealed knowledge polygraph test. *Expert Evid.* 6, 161–187.
- Hermans-Killam, L. (2002). *What Is Infrared?* Pasadena, CA: California Institute of Technology, Infrared Processing and Analysis Center. Retrieved from November 20, 2002. Available online at: <http://www.ipac.caltech.edu/Outreach/Edu/infrared.html>
- Iacono, W. G. (2008). Effective policing – understanding how polygraph tests work and are used. *Crim. Justice Behav.* 35, 1295–1308.
- Jones, B. F., and Plassmann, P. (2002). Digital infrared thermal imaging of human skin. *IEEE Eng. Med. Biol. Mag.* 21, 41–48.
- Kircher, J. C., Horowitz, S. W., and Raskin, D. C. (1988). Meta-analysis of mock crime studies of the control question polygraph technique. *Law Hum. Behav.* 12, 79–90.
- Lykken, D. T. (1959). The GSR in the detection of guilt. *J. Appl. Psychol.* 43, 385–388.
- MacLaren, V. V. (2001). A quantitative review of the guilty knowledge test. *J. Appl. Psychol.* 86, 674–683.
- Meijer, E. H., and Verschueren, B. (2010). The polygraph and the detection of deception. *J. Forensic Psychol. Pract.* 10, 325–338.
- National Research Council. (2003). *The Polygraph and Lie Detection*. Washington, DC: The National Academic Press.
- Pavlidis, I. (2005). U.S. Patent No. 6, 854, 879B2. Morristown, NJ: U.S.
- Pavlidis, I., Eberhardt, N. L., and Levine, J. A. (2002). Seeing through the face of deception. *Nature* 415:35. doi: 10.1038/415035a
- Pavlidis, I., and Levine, J. (2002). Thermal image analysis for polygraph testing. *IEEE Eng. Med. Biol. Mag.* 21, 56–64.
- Pavlidis, I., and Levine, J. (2001). “Monitoring of periorbital blood flow rate through thermal image analysis and its application to polygraph testing,” in *Proceeding of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology* (Istanbul, Turkey).
- Pollina, D. A., Dollins, A. B., Senter, S. M., Brown, T. E., Pavlidis, I., Levine, J. A., et al. (2006). Facial skin surface temperature changes during a “Concealed Information” Test. *Ann. Biomed. Eng.* 7, 1182–1189.
- Ramsay, J. O., Hooker, G., and Graves, S. (2009). *Functional Data Analysis with R and Matlab*. New York, NY: Springer.
- Ramsay, J. O., and Silverman, B. W. (2006). *Functional Data Analysis*. New York, NY: Springer.
- Reid, J. E. (1947). A revised questioning technique in lie detection tests. *J. Crim. Law Criminol.* 37, 542–547.
- Tsiamyrtzis, P., Dowdall, J., Shastri, D., Pavlidis, I. T., Frank, M. G., and Ekman, P. (2007). Imaging facial physiology for the detection of deceit. *Int. J. Comput. Vis.* 71, 197–214.
- Yankee, W. J. (1965). An investigation of sphygmomanometer discomfort thresholds in polygraph examinations. *Police* 9, 12–18.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 25 June 2012; accepted: 21 February 2013; published online: 07 March 2013.

*Citation: Park KK, Suk HW, Hwang H and Lee J-H (2013) A functional analysis of deception detection of a mock crime using infrared thermal imaging and the Concealed Information Test. *Front. Hum. Neurosci.* 7:70. doi: 10.3389/fnhum.2013.00070*

Copyright © 2013 Park, Suk, Hwang and Lee. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.

APPENDIX

ENVELOPE SELECTION EXPLANATION

"This experiment involves you conducting a mission. There are two different missions you will be choosing from. Both missions will be conducted at the same location. It is a public place with people who are not involved with the experiment. There may only be a small number of people, or quite a few.

In my hands are two different envelopes. The blue envelope contains a legal mission and the red envelope contains an illegal mission. If you choose the blue mission, you will simply do what the mission says, and as long as you correctly do what the mission states, you will receive \$10. However, if you choose the red mission, you will only receive a reward if you successfully complete the two stages of the mission.

To complete the first stage, you will have to complete the stated illegal act without anyone else knowing other than myself. If anyone else discovers that you are committing an illegal act, you automatically fail the mission. After successfully completing the mission, you will be questioned while having your bio-signals measured.

To complete the second stage of the mission, you must fool the questioner into thinking you did not commit any illegal acts, without being detected. This means that you must not only convince the questioner, but you must also trick the bio-signal measurement equipment as well. So, if you successfully complete both stages of the red mission, you will be rewarded \$10 for participating in the experiment, as well as a bonus reward of \$20, for a total of \$30. However, if you fail either stage of the red mission, you will receive nothing, and go home empty handed.

If you have any questions for me, please feel free to ask. If not, select an envelope now."

SMOOTHING: ROUGHNESS PENALTY SMOOTHING METHOD

Let y_t denote the value of facial temperature at time t . In functional data analysis, the observed temperature y_t , is regarded as a realization of an underlying continuous smooth temperature function x , rather than merely as a sequence of discrete observations. More specifically, we assume the following model,

$$y_t = x(t) + \varepsilon(t) \quad (\text{A1})$$

where $x(t)$ is the value of the underlying smooth function evaluated at time t and $\varepsilon(t)$ is a perturbation at time t that causes the observed data y_t to look rough. The first step of functional data analysis is to estimate the smooth function x , which requires a smoothing method to convert the observed values y_t to a function x with values $x(t)$ computable for any desired time point t .

Smoothing methods based on so-called basis expansion procedures represent a function x as a weighted sum of well-known basis functions ϕ_k

$$x(t) = \sum_{k=1}^K c_k \phi_k(t), \quad (\text{A2})$$

where K is the number of basis functions and c_k is the coefficient for the k th basis function. By estimating the coefficients c_k , we can represent a complicated-looking function as a linear combination

of well-known basis functions, which aids further analysis. In particular, it is convenient to estimate derivatives of a function expressed by a basis expansion because the first derivative of x , Dx , can be expressed as

$$Dx(s) = \sum_{k=1}^K c_k D\phi_k(s), \quad (\text{A3})$$

where $D\phi_k(s)$ is the first derivative of basis function ϕ_k at time s . More generally, the derivative of order m of function x at time s will be given as

$$D^m x(s) = \sum_{k=1}^K c_k D^m \phi_k(s), \quad (\text{A4})$$

where D^m denotes the derivative of order m . This property will be useful in estimating the coefficients c_k , which will be clear soon.

There are many popular bases that are widely used in practice. Most functional data analyses are known to involve either a Fourier basis for periodic data or a B-spline basis for non-periodic data (Ramsay and Silverman, 2006). We used the B-spline basis (de Boor, 2001) for smoothing the data because facial temperature can be considered non-periodic.

The remaining problem is how to estimate the coefficients c_k . We applied a penalized least-squares method that is considered more powerful and versatile than other methods such as least-squares smoothing, kernel smoothing, and local polynomial fitting approaches (Ramsay and Silverman, 2006). The objective function of the penalized least-squares method is given as the following.

$$L = \sum_{t=1}^T \left[y_t - \sum_{k=1}^K c_k \phi_k(t) \right]^2 + \lambda \int [D^2 x(s)]^2 ds \quad (\text{A5})$$

Basically, the coefficients are obtained by minimizing the sum of squared differences between observed values y , and function values x as shown in (A6), the first term of the objective function.

$$\sum_{t=1}^T \left[y_t - \sum_{k=1}^K c_k \phi_k(t) \right]^2 \quad (\text{A6})$$

The role of the second term in the objective function is to control the roughness of the function x . The squared second derivative $[D^2 x(s)]^2$ of function x at time s is called its curvature at s , since a straight line, which has no curvature, will have a zero second derivative. Therefore, a function's roughness measured across all time points s is the integrated squared second derivative as given in (A7).

$$\int [D^2 x(s)]^2 ds \quad (\text{A7})$$

By applying (4) this roughness can be rewritten as

$$\int \left[\sum_{k=1}^K c_k D^2 \phi_k(s) \right]^2 ds \quad (\text{A8})$$

which shows that the basis expansion is useful in estimating coefficients c_k ; it expresses the roughness as a linear combination of the second derivative of well-known basis functions.

By minimizing the objective function (A5), we wish to obtain two conflicting goals in curve estimation: model fit and generalizability. By minimizing (A6), we want to obtain the estimated curve with a good fit to the data in terms of minimizing the residual sum of squares. On the other hand, by minimizing (A7), we do not want the fit to be too good as to be excessively wiggly and overfit the data. The balance between these two conflicting goals in the objective function can be controlled by the smoothing parameter λ . As λ gets bigger, the objective function will place more emphasis on the smoothness and less on fitting the data. Therefore, as λ approaches infinity, the estimated curve will approach the standard linear regression which has $\int [D^2x(s)]^2 ds = 0$. In contrast, as λ becomes smaller, less penalty is placed on the curvature, and as λ approaches zero, the estimated curve approaches an interpolant to the data, which passes exactly through all the given data points. There exist several methods to choose the smoothing parameter and we applied the generalized cross-validation (GCV) method proposed by Craven and Wahba (1979). The basic idea under GCV is to compute a measure of mean squared error over a range of values of λ and choose the value that gives its minimum. In this analysis, we tried 11 different values of λ ($\log_{10} \lambda = -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5$) and obtained $\lambda = 10^1$ as the optimal value.

FUNCTIONAL ANOVA

After obtaining a smooth function x , we can perform a functional ANOVA. The form of the functional ANOVA model for analyzing our data can be given as

$$\text{Temp}_i(t) = \mu(t) + \alpha(t) + \beta(t) + \gamma(t) + \delta_i(t) + e_i(t) \quad (\text{A9})$$

where, $\text{Temp}_i(t)$ is the facial temperature for participant i evaluated at time t , $\mu(t)$ is the mean facial temperature of innocent participants answering irrelevant questions evaluated at time t , $\alpha(t)$ is the main effect of group (i.e., the difference between the mean face temperature of innocent participants and that of guilty participants when they are answering irrelevant questions), $\beta(t)$ is the main effect of relevance (i.e., the difference between the mean face temperature of innocent participants answering irrelevant questions and that of innocent participants answering relevant questions), $\gamma(t)$ is the interaction between group and relevance (i.e., the difference between the mean face temperature of guilty participants answering irrelevant questions and that of guilty participants answering relevant questions), $\delta_i(t)$ is the participant specific effect, and $e_i(t)$ is a residual function.

With this model, we want to test whether there is a significant effect of being in the guilty group, being in the relevant condition, and/or simultaneously being in the guilty group and relevant condition, on facial temperatures measured over time. In order to estimate parameters (or functions), we need to construct a matrix Z where N is the total number of functions ($N = 150$, 30 participants \times 5 questions). The rows corresponding to the participants in group 1 (innocent) and condition 1 (irrelevant) will have $[1\ 0\ 0\ 0]$, the rows corresponding to the participants group 2 (guilty) and condition 1 (irrelevant) will have $[1\ 1\ 0\ 0]$,

the rows corresponding to the participants in group 1 (innocent) and condition 2 (relevant) will have $[1\ 0\ 1\ 0]$, and the rows corresponding to the participants in group 2 (guilty) and condition 2 (relevant) will have $[1\ 1\ 1\ 1]$ in the first four columns. In the next 18 columns, subject k in group 1 (guilty) will have a row vector whose k th element is one and all the other elements are zero. In the next 12 columns, subject k in group 2 (innocent) will have a row vector whose k th element is one and all the other elements are zero. Then the model (A9) has the equivalent formulation as linear regression as shown in (A10),

$$\mathbf{x}(t) = \mathbf{Z}\boldsymbol{\beta}(t) + \mathbf{e}(t) \quad (\text{A10})$$

where $\mathbf{x}(t)$ is N by 1 vector of facial temperature, $\boldsymbol{\beta}(t) = [\mu(t), \alpha(t), \beta(t), \gamma(t), \delta_1(t), \dots, \delta_N(t)]'$, and $\mathbf{e}(t)$ is N by 1 vector of residuals. In order to deal with linear dependency in the columns of the design matrix, we need to augment the rows of the design matrix to enforce the sum of the participant specific effects in each group to be zero.

The regression coefficients $\boldsymbol{\beta}(t)$, are the parameters that we want to estimate and they can be estimated by minimizing the following objective function.

$$\begin{aligned} L_\beta = & \int ([\mathbf{x}(s) - \mathbf{Z}\boldsymbol{\beta}(s)]'[\mathbf{x}(s) - \mathbf{Z}\boldsymbol{\beta}(s)])ds \\ & + \lambda_\beta \int ([D^2\boldsymbol{\beta}(s)]'[D^2\boldsymbol{\beta}(s)])ds \end{aligned} \quad (\text{A11})$$

This objective function has the same structure as the objective function (A5) of the penalized least-squares smoothing. Basically, the regression coefficients are obtained by minimizing sum of squared differences between the function value $\mathbf{x}(s)$ and the predicted value $\mathbf{Z}\boldsymbol{\beta}(s)$ over all possible values of s , i.e., minimizing (A12), the first term of the objective function.

$$\int ([\mathbf{x}(s) - \mathbf{Z}\boldsymbol{\beta}(s)]'[\mathbf{x}(s) - \mathbf{Z}\boldsymbol{\beta}(s)])ds \quad (\text{A12})$$

The second term of the objective function is introduced to control the roughness of the regression coefficient function $\boldsymbol{\beta}$. The inner product of the second derivative of $\boldsymbol{\beta}(s)$, $[D^2\boldsymbol{\beta}(s)]'[D^2\boldsymbol{\beta}(s)]$, is defined as the curvature of the function $\boldsymbol{\beta}$ at time s . Therefore, the function's roughness measured across all argument values s is the integrated curvature as given in (A13).

$$\int ([D^2\boldsymbol{\beta}(s)]'[D^2\boldsymbol{\beta}(s)])ds \quad (\text{A13})$$

As $\boldsymbol{\beta}$ can be considered a smooth function, we used a basis expansion method to represent $\boldsymbol{\beta}$. Let \mathbf{B} denote a 2 by K_β matrix of coefficients, where K_β is the number of basis functions, and $\boldsymbol{\theta}(s)$ denote a K_β by 1 vector of basis functions at time s . We used the B-spline basis for smoothing the function of regression coefficients. The regression coefficients can now be represented as (A14) by the basis expansion.

$$\boldsymbol{\beta}(s) = \mathbf{B}\boldsymbol{\theta}(s) \quad (\text{A14})$$

Therefore, the objective function (A11) can be rewritten as (A15) and estimating regression coefficients β is equivalent to estimating the coefficients of the basis expansion \mathbf{B} .

$$\begin{aligned} L_\beta = & \int ([\mathbf{x}(s) - \mathbf{Z}\mathbf{B}\theta(s)]'[\mathbf{x}(s) - \mathbf{Z}\mathbf{B}\theta(s)])ds \\ & + \lambda_\beta \int ([\mathbf{B}D^2\theta(s)]'[\mathbf{B}D^2\theta(s)])ds \end{aligned} \quad (\text{A15})$$

The coefficient \mathbf{B} , can be estimated by minimizing L_β with respect to \mathbf{B} , which is given as $\text{vec}(\mathbf{B}) = [\mathbf{J}_{\theta\theta} \otimes (\mathbf{Z}'\mathbf{Z}) + \mathbf{R} \otimes \lambda_\beta \mathbf{I}]^{-1} \text{vec}(\mathbf{Z}'\mathbf{C}\mathbf{J}_{\phi\theta})$, where $\text{vec}(\mathbf{B})$ is an operator that stacks the columns of the matrix \mathbf{B} in one column vector, \otimes is the Kronecker product, \mathbf{C} is the matrix of coefficients in (A2), $\mathbf{J}_{\theta\theta} = \int \theta(s)\theta(s)'ds$, $\mathbf{J}_{\phi\theta} = \int \phi(s)\theta(s)'ds$, and $\mathbf{R} = \int D^2\theta(s)D^2\theta(s)'ds$.

The smoothing parameter λ_β again plays a role to control the level of the roughness of the estimated function of $\theta(s)$. As λ_β gets bigger, the estimated curve of $\theta(s)$ will be smoother. In order to choose the value of λ_β , we applied a cross-validation method proposed by Ramsay et al. (2009). The basic idea of this method is to compute the cross validated integrated squared error over a range of values of λ_β and choose the value that gives its minimum. In this analysis, we tried 13 different values of λ_β ($\log_{10} \lambda_\beta = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9$) and obtained $\lambda_\beta = 10^6$ as the optimal value.

FISHER'S LINEAR DISCRIMINANT ANALYSIS

The purpose of Fisher's linear discriminant analysis is to seek components (or linear combinations) of measured variables that are efficient for discrimination. Suppose that we have a set of N observations (or objects) of p dimensions denoted by $\mathbf{x}_1, \dots, \mathbf{x}_N$, which consists of C subsets, N_1 in the subset S_1 , N_2 in the subset S_2, \dots, N_C in the subset S_C . If we form a linear combination of the p variables in \mathbf{x}_i , it can be represented by

$$y_i = \mathbf{w}' \mathbf{x}_i \quad (\text{A1})$$

where \mathbf{w} indicates a weight vector of order p and y_i indicates the constructed component of observation i . If the norm of \mathbf{w} is one, constructing a component by $\mathbf{w}'\mathbf{x}_i$ is equivalent to projecting \mathbf{x}_i onto a line in the direction of \mathbf{w} geometrically. We want to find the best direction \mathbf{w} that separates C subsets as much as possible. The measure of the separation between the projected data y can be measured by the variance of the means of the C subsets. If $\bar{\mathbf{x}}$ is the grand mean of \mathbf{x}_i given by

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i \quad (\text{A2})$$

then the grand mean for the corresponding projected data is given by

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i = \frac{1}{N} \sum_{i=1}^N \mathbf{w}' \mathbf{x}_i = \mathbf{w}' \left(\frac{1}{N} \sum_{i=1}^N \mathbf{x}_i \right) = \mathbf{w}' \bar{\mathbf{x}}. \quad (\text{A3})$$

which is the projected grand mean. Likewise, if $\bar{\mathbf{x}}_c$ is the sample mean of \mathbf{x} in subset c given by

$$\bar{\mathbf{x}}_c = \frac{1}{N_c} \sum_{\mathbf{x}_i \in S_c} \mathbf{x}_i \quad (\text{A4})$$

then the sample mean for the corresponding y 's in subset c is given by

$$\bar{y}_c = \frac{1}{N_c} \sum_{y_i \in S_c} y_i = \frac{1}{N_c} \sum_{\mathbf{x}_i \in S_c} \mathbf{w}' \mathbf{x}_i = \mathbf{w}' \left(\frac{1}{N} \sum_{\mathbf{x}_i \in S_c} \mathbf{x}_i \right) = \mathbf{w}' \bar{\mathbf{x}}_c. \quad (\text{A5})$$

which is the projected subset mean. The variation among the projected means can be calculated by

$$\begin{aligned} s_B^2 &= \sum_{c=1}^C N_c (\bar{y}_c - \bar{y})^2 \\ &= \sum_{c=1}^C N_c (\mathbf{w}' \bar{\mathbf{x}}_c - \mathbf{w}' \bar{\mathbf{x}})^2 \\ &= \sum_{c=1}^C N_c \mathbf{w}' (\bar{\mathbf{x}}_c - \bar{\mathbf{x}}) (\bar{\mathbf{x}}_c - \bar{\mathbf{x}})' \mathbf{w} \\ &= \mathbf{w}' \left[\sum_{c=1}^C N_c (\bar{\mathbf{x}}_c - \bar{\mathbf{x}}) (\bar{\mathbf{x}}_c - \bar{\mathbf{x}})' \right] \mathbf{w} \\ &= \mathbf{w}' \mathbf{S}_B \mathbf{w} \end{aligned} \quad (\text{A6})$$

where $\mathbf{S}_B = \sum_{c=1}^C N_c (\bar{\mathbf{x}}_c - \bar{\mathbf{x}}) (\bar{\mathbf{x}}_c - \bar{\mathbf{x}})'$. We can make this variation as large as we wish by multiplying a constant to \mathbf{w} . To obtain good separation, this variance should be large relative to the variation within each subset, which can be measured by

$$\begin{aligned} s_W^2 &= \sum_{c=1}^C \sum_{y_i \in S_c} (y_i - \bar{y}_c)^2 \\ &= \sum_{c=1}^C \sum_{\mathbf{x}_i \in S_c} (\mathbf{w}' \mathbf{x}_i - \mathbf{w}' \bar{\mathbf{x}}_c)^2 \\ &= \sum_{c=1}^C \sum_{\mathbf{x}_i \in S_c} \mathbf{w}' (\mathbf{x}_i - \bar{\mathbf{x}}_c) (\mathbf{x}_i - \bar{\mathbf{x}}_c)' \mathbf{w} \\ &= \mathbf{w}' \left[\sum_{c=1}^C \sum_{\mathbf{x}_i \in S_c} (\mathbf{x}_i - \bar{\mathbf{x}}_c) (\mathbf{x}_i - \bar{\mathbf{x}}_c)' \right] \mathbf{w} \\ &= \mathbf{w}' \mathbf{S}_W \mathbf{w} \end{aligned} \quad (\text{A7})$$

where $\mathbf{S}_W = \sum_{c=1}^C \sum_{\mathbf{x}_i \in S_c} (\mathbf{x}_i - \bar{\mathbf{x}}_c) (\mathbf{x}_i - \bar{\mathbf{x}}_c)'$. Fisher's linear discriminant analysis seeks to find \mathbf{w} that maximizes the following objective function

$$J(\mathbf{w}) = \frac{s_B^2}{s_W^2} = \frac{\mathbf{w}' \mathbf{S}_B \mathbf{w}}{\mathbf{w}' \mathbf{S}_W \mathbf{w}}. \quad (\text{A8})$$

Maximizing (A8) with respect to \mathbf{w} is equivalent to maximizing $\mathbf{w}'\mathbf{S}_{BW}\mathbf{w}$ subject to the constraint, $\mathbf{w}'\mathbf{S}_W\mathbf{w} = 1$, which can be obtained by maximizing the following objective function with respect to \mathbf{w} and a Lagrange multiplier λ

$$J(\mathbf{w}, \lambda) = \mathbf{w}'\mathbf{S}_{BW} - \lambda(\mathbf{w}'\mathbf{S}_W\mathbf{w} - 1). \quad (\text{A9})$$

Taking a derivative of (A9) with respect to \mathbf{w} and setting it to zero yields

$$\mathbf{S}_{BW}\mathbf{w} = \lambda\mathbf{S}_W\mathbf{w} \quad (\text{A10})$$

which is a generalized eigenvalue problem. Taking a derivative of (A9) with respect to λ and setting it to zero yields the constraint

$$\mathbf{w}'\mathbf{S}_W\mathbf{w} = 1. \quad (\text{A11})$$

If we premultiply (A10) by \mathbf{w}' on both sides, we can obtain the value of λ

$$\lambda = \frac{\mathbf{w}'\mathbf{S}_{BW}}{\mathbf{w}'\mathbf{S}_W\mathbf{w}} = \mathbf{w}'\mathbf{S}_{BW} \quad (\text{A12})$$

which is the maximum variance among the projected means. Therefore, the eigenvalue of the eigen-equation (A10) indicates the maximum variance among the projected means that can be achieved and the eigenvector indicates the direction, or weight, that maximally separates C subgroups.

If one dimension is not enough to separate C subgroups, we can consider more eigenvalue-eigenvector pairs that can be obtained by solving the generalized eigenvalue problem given by (A10). If there are C subgroups, $C - 1$ eigenvalue-eigenvector pairs are usually used to separate C subgroups. This is equivalent to constructing $C - 1$ linear combinations of variables given by

$$\mathbf{y}_i = \mathbf{W}\mathbf{x}_i \quad (\text{A13})$$

where $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_{C-1}]'$ and \mathbf{y} is the vector of $C - 1$ components.

After obtaining \mathbf{W} , we can allocate a new observation \mathbf{x}^{new} into one of C subgroups in the following way. First we calculate $\mathbf{y}^{\text{new}} = \mathbf{W}\mathbf{x}^{\text{new}}$ and calculate the distance between \mathbf{y}^{new} and the projected mean of each subgroup given by

$$\begin{aligned} \text{dist}_c &= (\mathbf{y}^{\text{new}} - \bar{\mathbf{y}}_c)'(\mathbf{y}^{\text{new}} - \bar{\mathbf{y}}_c) \\ &= (\mathbf{W}\mathbf{x}^{\text{new}} - \mathbf{W}\bar{\mathbf{x}}_c)'(\mathbf{W}\mathbf{x}^{\text{new}} - \mathbf{W}\bar{\mathbf{x}}_c) \\ &= (\mathbf{x}^{\text{new}} - \bar{\mathbf{x}}_c)' \mathbf{W}'\mathbf{W} (\mathbf{x}^{\text{new}} - \bar{\mathbf{x}}_c). \end{aligned} \quad (\text{A14})$$

Then the new observation is allocated to the subgroup that yields the smallest distance value.

FUNCTIONAL DISCRIMINANT ANALYSIS

Fisher's linear discriminant analysis can be readily extended to functional data. Suppose we have a set of N functions measured at T time points, denoted by $x_i(t)$, where $i = 1, \dots, N$ and $t = 1, \dots, T$, and those N observations belong to C subsets with N_1 in the subset S_1 , N_2 in the subset S_2, \dots, N_C in the subset S_C . Fisher's linear discriminant analysis aims to find linear combinations of

variables that separate C subsets as much as possible. For functional data, finding a linear combination of observed variables corresponds to finding a weighted integration of the functions given by

$$y_i = \int \xi(t)x_i(t)dt \quad (\text{A15})$$

where $\xi(t)$ indicates the weight function evaluated at time t and y_i is the component constructed by the weighted integration of the observed function $x_i(t)$.

In order to estimate the weight function $\xi(t)$, we will adopt a basis function expansion approach to approximate functions. Any function can be approximated up to some degree of approximation by a linear combination of suitable basis functions. Suppose that $\phi(t)$ is an M by 1 vector of M suitable basis functions evaluated at time t . Then the observed functions $x_i(t)$, and the weight function $\xi(t)$, can be represented by basis function expansions as follows

$$x_i(t) = \mathbf{v}_i'\phi(t) = \phi(t)'\mathbf{v}_i \quad (\text{A16})$$

$$\xi(t) = \mathbf{w}'\phi(t) = \phi(t)'\mathbf{w} \quad (\text{A17})$$

where \mathbf{v}_i is the M by 1 vector of the coefficients of basis functions for $x_i(t)$ and \mathbf{w} is the M by 1 vector of the coefficients of basis functions for $\xi(t)$. Then the weighted integration (A15) can be rewritten as

$$\begin{aligned} y_i &= \int \mathbf{w}'\phi(t)\phi(t)'\mathbf{v}_i dt \\ &= \mathbf{w}' \left(\int \phi(t)\phi(t)' dt \right) \mathbf{v}_i \\ &= \mathbf{w}'\mathbf{J}\mathbf{v}_i \end{aligned} \quad (\text{A18})$$

where $\mathbf{J} = \int \phi(t)\phi(t)' dt$ is the M by M symmetric matrix of inner products of the basis functions.

We want to find the weight function $\xi(t)$, or equivalently, the coefficients of its basis functions \mathbf{w} , that separate C subsets as much as possible. The measure of the separation between the projected data y can be measured by the variance of the means of the C subsets. The grand mean for the corresponding component is given by

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i = \frac{1}{N} \sum_{i=1}^N \mathbf{w}'\mathbf{J}\mathbf{v}_i = \mathbf{w}'\mathbf{J} \left(\frac{1}{N} \sum_{i=1}^N \mathbf{v}_i \right) = \mathbf{w}'\mathbf{J}\bar{\mathbf{v}}. \quad (\text{A19})$$

where $\bar{\mathbf{v}}$ is the mean vector of the coefficients of basis functions for $x_i(t)$. Likewise, the sample mean for the y_i 's in subset c is given by

$$\bar{y}_c = \frac{1}{N_c} \sum_{y_i \in S_c} y_i = \frac{1}{N_c} \sum_{\mathbf{v}_i \in S_c} \mathbf{w}'\mathbf{J}\mathbf{v}_i = \mathbf{w}'\mathbf{J} \left(\frac{1}{N_c} \sum_{\mathbf{v}_i \in S_c} \mathbf{v}_i \right) = \mathbf{w}'\mathbf{J}\bar{\mathbf{v}}_c. \quad (\text{A20})$$

where $\bar{\mathbf{v}}_c$ is the mean vector of the coefficients of basis functions for $x_i(t)$ in subset c . The variation among the projected means can be calculated by

$$\begin{aligned}
s_B^2 &= \sum_{c=1}^C N_c (\bar{y}_c - \bar{y})^2 \\
&= \sum_{c=1}^C N_c (\mathbf{w}' \mathbf{J} \bar{\mathbf{v}}_c - \mathbf{w}' \mathbf{J} \bar{\mathbf{v}})^2 \\
&= \sum_{c=1}^C N_c \mathbf{w}' \mathbf{J} (\bar{\mathbf{v}}_c - \bar{\mathbf{v}}) (\bar{\mathbf{v}}_c - \bar{\mathbf{v}})' \mathbf{J} \mathbf{w} \\
&= \mathbf{w}' \mathbf{J} \left[\sum_{c=1}^C N_c (\bar{\mathbf{v}}_c - \bar{\mathbf{v}}) (\bar{\mathbf{v}}_c - \bar{\mathbf{v}})' \right] \mathbf{J} \mathbf{w} \\
&= \mathbf{w}' \mathbf{J} \mathbf{S}_B \mathbf{J} \mathbf{w}
\end{aligned} \tag{A21}$$

where $\mathbf{S}_B = \sum_{c=1}^C N_c (\bar{\mathbf{v}}_c - \bar{\mathbf{v}}) (\bar{\mathbf{v}}_c - \bar{\mathbf{v}})'$. We want this variation as large as possible relative to the variation within each subset, which can be measured by

$$\begin{aligned}
s_W^2 &= \sum_{c=1}^C \sum_{y_i \in S_c} (y_i - \bar{y}_c)^2 \\
&= \sum_{c=1}^C \sum_{\mathbf{v}_i \in S_c} (\mathbf{w}' \mathbf{J} \mathbf{v}_i - \mathbf{w}' \mathbf{J} \bar{\mathbf{v}}_c)^2 \\
&= \sum_{c=1}^C \sum_{\mathbf{v}_i \in S_c} \mathbf{w}' \mathbf{J} (\mathbf{v}_i - \bar{\mathbf{v}}_c) (\mathbf{v}_i - \bar{\mathbf{v}}_c)' \mathbf{J} \mathbf{w} \\
&= \mathbf{w}' \mathbf{J} \left[\sum_{c=1}^C \sum_{\mathbf{v}_i \in S_c} (\mathbf{v}_i - \bar{\mathbf{v}}_c) (\mathbf{v}_i - \bar{\mathbf{v}}_c)' \right] \mathbf{J} \mathbf{w} \\
&= \mathbf{w}' \mathbf{J} \mathbf{S}_W \mathbf{J} \mathbf{w}
\end{aligned} \tag{A22}$$

where $\mathbf{S}_W = \sum_{c=1}^C \sum_{\mathbf{v}_i \in S_c} (\mathbf{v}_i - \bar{\mathbf{v}}_c) (\mathbf{v}_i - \bar{\mathbf{v}}_c)'$. Fisher's linear discriminant analysis seeks to find \mathbf{w} that maximizes the following objective function

$$J(\mathbf{w}) = \frac{s_B^2}{s_W^2} = \frac{\mathbf{w}' \mathbf{J} \mathbf{S}_B \mathbf{J} \mathbf{w}}{\mathbf{w}' \mathbf{J} \mathbf{S}_W \mathbf{J} \mathbf{w}} \tag{A23}$$

If we use enough number of basis functions, for example, the same number as the number of time points, the estimated weight function $\xi(t) = \mathbf{w}' \phi(t)$ could be jagged, which would make the interpretation of the weight function difficult. We would prefer a smoother version of the weight function, which can be obtained by maximizing the following regularized objective function

$$\begin{aligned}
J(\mathbf{w}, \rho) &= \frac{\mathbf{w}' \mathbf{J} \mathbf{S}_B \mathbf{J} \mathbf{w}}{\mathbf{w}' \mathbf{J} \mathbf{S}_W \mathbf{J} \mathbf{w} + \rho \int (D^2 \xi(t))^2 dt} \\
&= \frac{\mathbf{w}' \mathbf{J} \mathbf{S}_B \mathbf{J} \mathbf{w}}{\mathbf{w}' \mathbf{J} \mathbf{S}_W \mathbf{J} \mathbf{w} + \rho \int D^2 \mathbf{w}' \phi(t) D^2 \phi(t)' \mathbf{w} dt}
\end{aligned}$$

$$\begin{aligned}
&= \frac{\mathbf{w}' \mathbf{J} \mathbf{S}_B \mathbf{J} \mathbf{w}}{\mathbf{w}' \mathbf{J} \mathbf{S}_W \mathbf{J} \mathbf{w} + \rho \mathbf{w}' \left(\int D^2 \phi(t) D^2 \phi(t)' dt \right) \mathbf{w}} \\
&= \frac{\mathbf{w}' \mathbf{J} \mathbf{S}_B \mathbf{J} \mathbf{w}}{\mathbf{w}' \mathbf{J} \mathbf{S}_W \mathbf{J} \mathbf{w} + \rho \mathbf{w}' \mathbf{R} \mathbf{w}}
\end{aligned} \tag{A24}$$

where D^2 indicates the second derivative of a function and $\mathbf{R} = \int D^2 \phi(t) D^2 \phi(t)' dt$. The penalty term added to the objective function $\mathbf{w}' \mathbf{R} \mathbf{w}$ is the squared integration of the second derivative of the weight function, which measures the roughness or smoothness of the weight function. The penalty parameter $\rho \geq 0$ controls the importance of the penalty term. A larger value of penalty term will put more emphasis of the smoothness of the weight function and yield a smoother weight function. The optimal value of ρ can be obtained by a cross-validation method.

Maximizing (A24) with respect to \mathbf{w} is equivalent to maximizing $\mathbf{w}' \mathbf{J} \mathbf{S}_B \mathbf{J} \mathbf{w}$ subject to the constraint $\mathbf{w}' \mathbf{J} \mathbf{S}_W \mathbf{J} \mathbf{w} + \rho \mathbf{w}' \mathbf{R} \mathbf{w} = 1$, which can be obtained by maximizing the following objective function with respect to \mathbf{w} and a Lagrange multiplier λ

$$J(\mathbf{w}, \lambda) = \mathbf{w}' \mathbf{J} \mathbf{S}_B \mathbf{J} \mathbf{w} - \lambda (\mathbf{w}' \mathbf{J} \mathbf{S}_W \mathbf{J} \mathbf{w} + \rho \mathbf{w}' \mathbf{R} \mathbf{w}) \tag{A25}$$

Taking a derivative of (A25) with respect to \mathbf{w} and setting it to zero yields

$$\mathbf{J} \mathbf{S}_B \mathbf{J} \mathbf{w} = \lambda (\mathbf{J} \mathbf{S}_W \mathbf{J} + \rho \mathbf{R}) \mathbf{w} \tag{A26}$$

which is a generalized eigenvalue problem. Taking a derivative of (A25) with respect to λ and setting it to zero yields the constraint

$$\mathbf{w}' \mathbf{J} \mathbf{S}_W \mathbf{J} \mathbf{w} + \rho \mathbf{w}' \mathbf{R} \mathbf{w} = 1. \tag{A27}$$

If we premultiply (A26) by \mathbf{w}' on both sides, we can obtain the value of λ

$$\lambda = \frac{\mathbf{w}' \mathbf{J} \mathbf{S}_B \mathbf{J} \mathbf{w}}{\mathbf{w}' \mathbf{J} \mathbf{S}_W \mathbf{J} \mathbf{w} + \rho \mathbf{w}' \mathbf{R} \mathbf{w}} = \frac{\mathbf{w}' \mathbf{J} \mathbf{S}_B \mathbf{J} \mathbf{w}}{\mathbf{w}' \mathbf{J} \mathbf{S}_W \mathbf{J} \mathbf{w}} \tag{A28}$$

which is the maximum variance among the means of the component y_i . Therefore, the eigenvalue of the eigen-equation (A26) indicates the maximum variance among the means of the component that can be achieved and the eigenvector indicates the coefficients of the weight function that maximally separate C subgroups.

If one dimension is not enough to separate C subgroups, we can consider more eigenvalue-eigenvector pairs that can be obtained by solving the generalized eigenvalue problem given by (A10). If there are C subgroups, $C - 1$ eigenvalue-eigenvector pairs are usually used to separate C subgroups. This is equivalent to constructing $C - 1$ linear combinations of variables given by

$$\mathbf{y}_i = \mathbf{W} \mathbf{J} \mathbf{v}_i \tag{A29}$$

where $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_{C-1}]'$ and \mathbf{y} is the vector of $C - 1$ components.

After obtaining \mathbf{W} , we can allocate a new observation $x^{\text{new}}(t) = \phi(t)' \mathbf{v}^{\text{new}}$ into one of the C subgroups in the following

way. First we calculate $\mathbf{y}^{\text{new}} = \mathbf{WJv}^{\text{new}}$ and calculate the distance between \mathbf{y}^{new} and the projected mean of each subgroup given by

$$\begin{aligned}\text{dist}_c &= (\mathbf{y}^{\text{new}} - \bar{\mathbf{y}}_c)' (\mathbf{y}^{\text{new}} - \bar{\mathbf{y}}_c) \\ &= (\mathbf{WJv}^{\text{new}} - \mathbf{WJ}\bar{\mathbf{v}}_c)' (\mathbf{WJv}^{\text{new}} - \mathbf{WJ}\bar{\mathbf{v}}_c) \\ &= (\mathbf{v}^{\text{new}} - \bar{\mathbf{v}}_c)' \mathbf{W}'\mathbf{J}\mathbf{J}'\mathbf{W} (\mathbf{v}^{\text{new}} - \bar{\mathbf{v}}_c).\end{aligned}\quad (\text{A30})$$

Then the new observation is allocated to the subgroup that yields the smallest distance value.

CONCEALED INFORMATION TEST QUESTIONING PROTOCOL

Primary question 1: "What was the item you stole from the computer lab?"

Sub-question 1-1: "Was it a ring?"

Sub-question 1-2: "Was it a wallet?"

Sub-question 1-3: "Was it a pair of glasses?"

Sub-question 1-4: "Was it a notepad?"

Sub-question 1-5: "Was it a jewel?"

Primary question 2: "What did you do before you left the computer lab?"

Sub-question 2-1: "Did you clean the room?"

Sub-question 2-2: "Did you meet a friend?"

Sub-question 2-3: "Did you take a picture?"

Sub-question 2-4: "Did you turn off the webcam?"

Sub-question 2-5: "Did you have a drink?"

Primary question 3: "What was the item you stole from the computer lab?"

Sub-question 3-1: "Was it a jewel?"

Sub-question 3-2: "Was it a notepad?"

Sub-question 3-3: "Was it a watch?"

Sub-question 3-4: "Was it a wallet?"

Sub-question 3-5: "Was it a pair of glasses?"



Combining blink, pupil, and response time measures in a concealed knowledge test

Travis L. Seymour*, Christopher A. Baker and Joshua T. Gaunt

Cognitive Modeling Laboratory, Psychology Department, University of California Santa Cruz, Santa Cruz, CA, USA

Edited by:

Wolfgang Ambach, Institute for Frontier Areas of Psychology and Mental Health, Germany

Reviewed by:

Bruno Verschueren, Ghent University, Belgium

Andrea Webb, Draper Laboratory, USA

***Correspondence:**

Travis L. Seymour, Psychology Department, University of California Santa Cruz, 357 Social Sciences 2, Santa Cruz, CA 95064, USA.

e-mail: nogard@ucsc.edu

The response time (RT) based *Concealed Knowledge Test* (CKT) has been shown to accurately detect participants' knowledge of mock-crime-related information. Tests based on ocular measures such as pupil-size and blink-rate have sometimes resulted in poor classification, or lacked detailed classification analyses. The present study examines the fitness of multiple pupil and blink related responses in the CKT paradigm. To maximize classification efficiency, participants' concealed knowledge was assessed using both individual test measures and combinations of test measures. Results show that individual pupil-size, pupil-slope, and pre-response blink-rate measures produce efficient classifications. Combining pupil and blink measures yielded more accuracy classifications than individual ocular measures. Although RT-based tests proved efficient, combining RT with ocular measures had little incremental benefit. It is argued that covertly assessing ocular measures during RT-based tests may guard against effective countermeasure use in applied settings. A compound classification procedure was used to categorize individual participants and yielded high hit rates and low false-alarm rates without the need for adjustments between test paradigms and subject populations. We conclude that with appropriate test paradigms and classification analyses, ocular measures may prove as effective as other indices, though additional research is needed.

Keywords: deception, guilty knowledge, concealed information, lying, pupil, blinks, recognition

INTRODUCTION

COMBINING BLINK, PUPIL, AND RESPONSE TIME MEASURES IN A CONCEALED KNOWLEDGE TEST

Researchers have developed several paradigms to assess whether or not participants are concealing sensitive information (for reviews, see Ben-Shakhar and Furedy, 1990; Lykken, 1998; MacLaren, 2001; Ben-Shakhar and Elaad, 2003). This approach differs from the control questions "lie detector" test because it focuses on the ability of various dependent measures to indicate when participants recognize critical information as opposed to lying about it *per se*. A meta-analysis of concealed knowledge tests (CKT) revealed an average hit rate of 0.83 and a false-alarm rate of 0.04 (Ben-Shakhar and Elaad, 2003). In light of the dubious theoretical underpinnings and highly variable performance of the traditional "lie detector" test (National Research Council, 2003), many researchers have developed tests using indices of concealed knowledge, rather than indices of deception (c.f. Verschueren et al., 2011).

CONCEALED KNOWLEDGE DETECTION

Following previous work by Rosenfeld et al. (1988), Farwell and Donchin described a CKT paradigm in which responses to familiar crime-related *probes* could be compared to familiar *target* items not associated with the crime (Farwell and Donchin, 1991). Participants memorized a set of probe phrases (e.g., "White Shirt") and then used this information to enact a mock-crime scenario. Later, they memorized a set of target phrases (e.g., "Blue Coat") unrelated to the scenario. In a subsequent memory test, participants accurately indicated their recognition of target phrases, but

denied recognition of familiar-probe phrases. On trials containing novel *irrelevant* phrases, participants accurately indicated their lack of knowledge. The target stimuli in this paradigm are important because only they require an affirmative response. Without targets, one could respond "no" on each trial without considering the stimulus; a strategy that could attenuate the effectiveness of the test (for an alternate view, see Rosenfeld et al., 2006). Thus, targets force participants to process each stimulus (including crime-relevant probes). Using evoked-related brain potentials (ERP) to index stimulus familiarity in the brain's anterior cingulate cortex, Farwell and Donchin achieved a hit rate of 0.9 with no false-alarm rates.

Using a similar paradigm (but with a 1000 ms response deadline), Seymour et al. (2000) examined whether response time (RT) and accuracy were sufficient to detect concealed knowledge from a mock-crime. Results showed that "no" responses to crime-related probes were significantly slower and less accurate than to unfamiliar irrelevant items. A specialized individual classification procedure that compared participants' probe and irrelevant RT distributions led to a 0.93 hit rate with no false-alarms. Similar results have been reported in subsequent studies using related CKT test procedures and analyses (Seymour and Kerlin, 2008; Seymour and Fraynt, 2009; Verschueren et al., 2010; Visu-Petra et al., 2011).

Although the RT-based CKT can yield high detection rates, examinees may attempt to manipulate their responses to undermine a test's effectiveness. Studies have shown that a variety of physiological and neuropsychological-based tests are susceptible to strategic countermeasures that reduce detection

rates (Seymour and Kerlin, 2008; Seymour and Fraynt, 2009; Verschueren et al., 2010; Visu-Petra et al., 2011). For the RT measure in the CKT paradigm, results have been mixed. Some data suggest that attempting to appear unfamiliar with familiar-probes by equating probe and irrelevant RTs is generally ineffective (Seymour et al., 2000). However, effective countermeasures have been demonstrated using CKT paradigms without response deadlines (e.g., Rosenfeld et al., 2004), and emotional Stroop (Williams et al., 1996) based detection paradigms (Gronau et al., 2005; Degner, 2009).

One approach that may potentially lead to more accurate countermeasure-resistant paradigms involves simultaneously assessing multiple measures in a single paradigm (Gronau et al., 2005). Although previous work has examined in detail the anti-countermeasure benefits of combining various polygraph-based measures (respiratory rate, heart-rate, electrodermal response, etc.; c.f., Elaad, 2011), few have included RT and ocular measures. However, some studies have examined such measures in combined tests. Cutrow et al. (1972) reported that an amalgamation of respiratory rate, eye blink-rate, pulse, and electrodermal responses allowed differentiation between answers to mock-crime and irrelevant questions. However, classification analyses were omitted. Without individual classification rates (in particular false-alarm), this result cannot be properly evaluated. Allen et al. (1992) also analyzed a CKT using combined measures (ERP and RT) that yielded average hit rates of 0.98 and false-alarm rates of 0.03. Although the combined-measure false-alarm rate was only 0.02 greater than using ERP alone, the addition of RT reduced the miss rate by 0.04. Several studies have examined combinations of polygraph measures such as electrodermal response, heart-rate, and respiratory rate. Such combined tests often yield small but robust improvements over individual indicators (e.g., Elaad et al., 1992; Gamer et al., 2008). However, in other studies, such combinations have failed to outperform their individual counterparts (e.g., Bradley and Warfield, 1984; Verschueren et al., 2007). Although differences between studies may explain this disparity (c.f. Meijer et al., 2007), in the present study we examined the benefit of combining ocular and RT-based measures of concealed knowledge.

OCULAR MEASURES OF CONCEALED KNOWLEDGE

A potentially effective test may combine intentional motor responses such as RT with more autonomic ocular responses such as pupil-size and blinking rate; both of which can be assessed simultaneously without interference. Modern eye-trackers can be calibrated and used without participants' awareness, limiting opportunities for countermeasures. Even with conspicuous eye measurement, automatic responses such as blinking and pupil dilation may be difficult to control systematically in a covert fashion. Of course, the advantage of combining ocular and RT measures depends on the degree to which these measures are correlated with one another. One reason why consistently successful combined paradigms have been elusive is that the diagnostic accuracy of individual ocular measures remains uncertain (c.f. Gamer, 2011). For example, one potential ocular measure, internally cued (i.e., endogenous) blinking, is typically correlated with cognitive demand (unlike reflexive or voluntary blinks) (Drew,

1951; Holland and Tarlow, 1972; Bagley and Manelis, 1979; Stern et al., 1984; Bauer et al., 1987; Goldstein et al., 1992). Accordingly, they tend to be inhibited during the processing or anticipation of relevant stimuli and occur most frequently at junctures between processing. Peak blink-rate (maximum average blink-rate reached during each trial) tends to increase as a function of processing load whereas latency to peak rate (average time required on each trial to reach that trial's peak blink-rate) increases with processing duration (Stern et al., 1984; Bauer et al., 1987; Goldstein et al., 1992; Ichikawa and Ohira, 2004). Some studies have shown that overall blinking behavior is sensitive to concealed knowledge (Janisse and Bradley, 1980; Dionisio et al., 2001; Fukuda, 2001; Leal and Vrij, 2008). For example, Leal and Vrij (2010) examined blink activity during a paradigm in which participants made either truthful or deceptive statements about participation in a mock-crime. Results showed that liars displayed significantly fewer blinks for probe questions than for controls. Truth tellers showed no such difference. A discriminant analysis on probe-control differences for each participant yielded a 0.75 hit rate and a 0.23 false-alarm rate.

In addition to overall blink-rate, it has been suggested that temporal variations in blink activity can differentiate probe and irrelevant stimuli and perform significantly better than overall blink-rate (Stern et al., 1984; Fukuda, 2001; Ichikawa and Ohira, 2004; Leal and Vrij, 2008). For example, Fukuda (2001) measured the number of blinks participants produce on each trial during a concealed knowledge paradigm and plotted them as a function of trial duration. Analysis was done on the shape of the resulting temporal distribution of blinking (TDB) and assessed various characteristics such as average blink-rate, peak blink-rate, and time-to-peak. Results showed that responding to probe stimuli led to a higher average blink-rate that peaked earlier and higher than to irrelevant stimuli. Unfortunately, a detailed classification analysis was omitted making it difficult to assess the diagnosticity of the TDB measure. Nevertheless, a successful blinking measure might prove an important addition to a combined-measure CKT. Crucially, Goldstein et al. (1992) found that RT and blinking were uncorrelated and influenced by different task variables, suggesting that these measures may be ideal candidates for combined tests.

Similar to blinks, pupil-size has been shown to reliably index cognitive task demand (Beatty, 1982; Steinhauer and Hakerem, 1992; Karatekin et al., 2004), and has also been shown to index emotional arousal (Bradley et al., 2008). Because of such results, pupil-size has been explored as a measure of deception (Berrien and Huntington, 1942; Heilveil, 1976; Janisse and Bradley, 1980; Lubow and Fein, 1996; Dionisio et al., 2001; Webb et al., 2009a,b). Fluctuations in pupil-size can be highly reliable even when small in magnitude, with researchers reporting robust effects as small as 0.1 mm (Hakerem and Sutton, 1966) and 0.015 mm (Beatty, 1988). Lubow and Fein (1996) found greater pupil dilation following presentation of mock-crime-related probes than irrelevant items in a CKT paradigm. A classification analysis yielded hit rates of 0.50 and 0.70 with no false-alarms (overall detection accuracies of 75 and 85%). This was an improvement on an earlier pupil-based test reporting overall detection accuracies between 66 and 69% (Janisse and Bradley, 1980). A later study by Dionisio et al. (2001), in which participants made true and then false statements

about benign scenarios, reported greater average pupil-size during false than true statements for 92% of participants. Again, the necessary classification information (false-alarm rates in particular) was unavailable for this study, as well as the Janisse and Bradley studies. Cook et al. (2012) did report detailed classification results from a test consisting of true/false questions, e.g., “I took the \$20 from the secretary’s purse.” Both pupil-size and eye scan-patterns were recorded. Across two experiments, they found an average hit rate of 0.80, and an average false-alarm rate of 0.13. Kircher et al. (2010) reported results from tests using demographic and true/false questions. Deception was indexed using pupil-size, reading pattern, and RT measures, but average hit rate (0.80) and false-alarm rate (0.15) were similar to Cook and colleagues. Overall, pupil-based measures seem promising for the CKT paradigm, but more work is needed to find robust methods that increase hit rates and reduce false-alarm rates to levels comparable with other more established CKT measures.

A NEW TEST COMBINING BEHAVIORAL AND OCULAR MEASURES

The lack of detailed individual classification analyses limits the ability to assess ocular measures in some CKT studies. In the present study, this is remedied by examining both pupil-size and blink measures using an individual subject classification procedure for participants familiar with probes (to assess hit and miss rates) and participants unfamiliar with probes (to assess correct-rejection and false-alarm rates). Another question inconsistently answered in the literature is the fitness of combined-measure CKT paradigms. Although work exists showing successful combinations of polygraph measures (e.g., Gamer et al., 2008), consistent results are not available for combinations of ocular and RT measures. Although this disparity may be in part due to differences in test parameters or classification analyses, we argue that combinations of more disparate measures could be more diagnostic, and could potentially thwart the use of some countermeasures. To our knowledge, this is the first study to evaluate the combined diagnosticity of response-time, pupil-size, and blink measures. In addition to the standard mean pupil-size and peak blink-rate measures, we added Fukuda’s (2001) blink distribution measures and a new pupil-slope measure following observations by Lubow and Fein (1996). To require that participants process each stimulus, we used the 3-stimulus variant of the CKT (probe = “no,” target = “yes,” and irrelevant = “no”).

MATERIALS AND METHODS

PARTICIPANTS

Sixty undergraduate students (67% female) at the University of California Santa Cruz participated in the experiment for course credit. All participants had normal or corrected-to-normal vision.

MATERIALS AND APPARATUS

The stimuli were 66 luminance-matched color pictures of non-familiar human faces (half female) with neutral expressions taken from the Aberdeen Psychological Image Collection (Hancock, 2004). Pictures were presented on a 17" monitor with a refresh rate of 85 Hz and each subtended an area of $12.5 \times 16.2^\circ$ of visual angle at a viewing distance of 18". Stimulus presentation and randomization, as well as the recording of RT and accuracy were

managed using E-Prime presentation software (Schneider et al., 2002). RTs were entered on a Cedrus four-button response pad (Cedrus Corporation, San Pedro, CA, USA). An Arrington View-Point eye-tracker (Arrington Research, Inc. Scottsdale, AZ, USA) was used to record blinking and pupil-size at a sample rate of 60 Hz. Participants’ heads were stabilized using a chin rest. During calibration, the location and extent of participants’ right pupil and the location of their pupil glint were mapped. The best fitting ellipse was constantly computed to fit the pupil over time. Pupil-size is thus an online measure in millimeters of the transverse diameter of this ellipse. Blinks were also measured with respect to this geometry. When participants blink, their eyelid falls and the best fitting ellipse becomes increasingly flat before the pupil disappears altogether. This transition is used to detect blinks, but requires a threshold value. Pupil geometry is partially a function of viewing angle with respect to the display and the position of the eyes; thus, the exact height to width ratio of the ellipse that will indicate a blink must be determined separately for each participant. To achieve this, the range of aspect ratios noted during spatial calibration (participants cued to look at various points across the display) was recorded. Subsequently, a blink threshold was chosen for each participant to distinguish between real blinks and flattened ellipses that occurred naturally when eyes were moved toward the various edges of the display. The mean threshold ratio was 0.6.

DESIGN AND PROCEDURE

The experiment was comprised of a series of tasks to be completed in the following order: A probe-learning phase, a retention interval, a target-learning phase, and a picture recognition task. Each session lasted approximately 1 h.

Probe-learning phase

For each participant, a set of six probe faces was selected randomly from the entire pool of faces. The study procedure for probe faces was designed to ensure elaborative encoding of probe stimuli (c.f., Seymour and Kerlin, 2008). This is in contrast to mock-crime procedures during which individual variations in memory, motivation, and attention can lead to the encoding of some probe items but not others (Carmel et al., 2003). Such variations may increase potential external validity, but could lead to the confounding of mock-crime effectiveness and the diagnostic accuracy of the test (Seymour and Fraynt, 2009).

Participants studied each face for 45 s and were then shown one of six facial-feature questions (e.g., “did that person have facial hair?”). These questions were chosen randomly with replacement to prevent anticipation. After each feature judgment, the face was shown again for a mirror image judgment. Each image was either flipped on its vertical axis or not flipped at all. Participants pressed one button for “same” and another for “mirror” and were given immediate accuracy feedback. This cycle, in which face image study is followed by feature and mirror judgments, was repeated for each of the six probe faces. Once this cycle had been completed for all six probes, the order of faces was re-randomized and the study process was repeated until the entire set of probes was studied a total of three times. After this portion of the probe-learning phase was completed, participants were asked to rate each picture

for its perceived attractiveness (seven-point Likert scale), honesty (seven-point Likert scale), and age (open ended).

Retention interval

To prevent rehearsal of probe items during the 10 min retention interval, participants completed a set of difficult mathematical word problems (taken from Patalano and Seifert, 1994).

Target-learning phase

Following the retention interval, six additional faces were randomly selected to be target stimuli. Targets were studied in the same manner as probes. That is, faces were shown individually for study and followed by both feature and mirror judgments. However, for targets there were no attractiveness, honesty, or age ratings. This study difference affords participants a basis on which to distinguish probe and target faces in the subsequent recognition task (Seymour and Kerlin, 2008).

Picture recognition task

Before beginning the recognition task, participants' gaze coordinates were mapped to a standardized space via an eye-tracking calibration procedure. Following calibration, participants were shown a series of pictures and made speeded recognition judgments. On each trial, participants first saw a white visual mask with a black fixation-cross displayed at its center. After 1200 ms, a stimulus picture replaced the mask and remained on the screen until a response was made. Participants were asked to indicate on each trial their familiarity with the stimulus. For target faces they were to truthfully press a button marked "yes." Similarly, for irrelevant faces, participants were to truthfully respond "no." However, for probe faces participants were asked to deceptively respond "no," despite their actual familiarity with these stimuli. Note that although participants were told that they were completing a deception task and that success meant responding just as quickly and accurately to probe stimuli as they did to irrelevant stimuli, no specific countermeasure instructions or monetary incentives were offered. After each response a blank screen was shown before the next trial began for a random duration between 2000 and 2500 ms. A 3000 ms deadline was used; responses longer than the deadline were followed by an "ERROR: TOO SLOW" warning. Otherwise, no feedback was given during each block. In previous studies using this paradigm with two-word verbal phrases, deadlines of 1000 ms (Seymour et al., 2000) and 1500 ms were used (Seymour and Kerlin, 2008; Seymour and Fraynt, 2009). The use here of a 3000 ms deadline was necessary given the relative complexity and high feature overlap of face stimuli (Bruce, 1982).

Each trial block contained one presentation of each face picture in the stimulus set (six targets, six probes, and 24 irrelevants) in a new random-order, for a total of 36 trials. Participants were randomly assigned to either a *familiar-probe* condition (in which probes were previously studied faces) or an *unfamiliar-probe* condition (in which probes were new faces). To participants, unfamiliar-probes are essentially irrelevants; this condition is analogous to testing an unaware examinee and is used to estimate the test's false-alarm rate. Following each block, participants were shown a feedback screen including mean accuracy and the number of "Too Slow" errors for that block. In each condition, three blocks were completed for a total of 108 trials per participant.

ANALYSES AND PREDICTIONS

Individual and combined test measures

Prior to each individual measure's analysis, we calculated within-subject Z-scores to give a better indication of the effect size for each measure uncontaminated by individual differences in general responsiveness (c.f. Ben-Shakhar, 1985). In particular, for each participant we calculated the mean of all that participant's responses (regardless of stimulus type), and subtracted this value from each score prior to dividing this result by the SD of all of that participant's responses (regardless of stimulus type). Although all analyses and figures represent standardized data, **Table 1** lists the mean untransformed data for each measure. Although **Table 1** lists the mean and SD for each stimulus type, only probe and irrelevant stimuli were used for statistical analyses and classification. For the classification of individual participants' data, both individual and combined measures were used. Combined measures were simple sums of individual measures.

The Eta-squared statistic is included for each analysis as a measure of effect size. All *post hoc* t-tests were compared against a Tukey HSD corrected alpha level, and all t-tests were treated as *post hoc* unless otherwise noted. Lastly, all statistical tests were compared against a nominal alpha level of 0.05 unless otherwise noted.

Response time and accuracy

For RT and accuracy measures, we compared probe and irrelevant distributions as a function of the two probe-familiarity conditions. For the RT measure, only correct trials were included in the analysis. As in previous research using the present paradigm, we expected that Probes would be slower and less accurate in the familiar-probe condition compared to the unfamiliar-probe condition (e.g., Allen et al., 1992; Seymour and Kerlin, 2008; Seymour and Fraynt, 2009; Verschueren et al., 2010; Visu-Petra et al., 2011).

Blinking measures

Following Fukuda (2001), we analyzed endogenous blink-rate as a function of probe-familiarity condition and stimulus type for correct trials. The analysis window was divided into 25, 50 ms bins and a TDB was computed for each participant. Blink-rate was calculated for each bin by dividing the total number of blinks for that bin and stimulus type by the total number of trials for that stimulus type. The resulting value (i.e., blinks per 50 ms) was then multiplied by 20 for conversion into blinks-per-second (c.f. Fukuda, 2001) prior to being converted to Z-Scores. The resulting TDBs, averaged over participant, are plotted in **Figure 2** by condition. Fukuda reported significant inhibition throughout most of the time the stimulus was onscreen. However, in the period just prior to the response, a significant increase in blinking occurred on probe trials only. Thus, we predicted that pre-response blink-rate would be likewise diagnostic in the current study. To identify the appropriate region for analysis, we examined blinking behavior across each trial over all stimulus types. Similar to Fukuda, participants in the current study rarely blinked during stimulus presentation. Out of the 5616 available correct trials, only 151 (2.6%) contained blinking during the first 400 ms following stimulus onset. In contrast, during the period from 400 ms prior to stimulus offset (i.e., response initiation) to 800 ms after stimulus

Table 1 | Mean un-standardized data by stimulus type and condition for each measure.

Measure	Stimulus type			Effect
	Irrelevant	Probe	Target	
FAMILIAR-PROBE CONDITION				
Response time (ms)	740 (150)	1086 (269)	908 (142)	346
Accuracy (%)	98 (3)	69 (24)	84 (15)	29
Pupil-size (mm)	3.9 (0.51)	4.0 (0.52)	4.0 (0.51)	0.10
Pupil-slope ($\times 1000$ mm)	0.49 (0.11)	0.59 (0.14)	0.53 (0.13)	0.10
Peak blink-rate (b/s)	0.09 (0.23)	0.27 (0.49)	0.10 (0.19)	0.18
UNFAMILIAR-PROBE CONDITION				
Response time (ms)	792 (132)	757 (164)	828 (189)	-35
Accuracy (%)	97 (4)	98 (3)	87 (13)	-1.0
Pupil-size ($\times 1000$ mm)	4.1 (0.49)	4.1 (0.50)	4.1 (0.48)	0
Pupil-slope (mm)	0.46 (0.15)	0.46 (0.18)	0.48 (0.18)	0
Peak blink-rate (b/s)	0.30 (0.95)	0.37 (1.3)	0.40 (1.7)	0.07

SDs are indicated with parenthesis. Effect calculations involve subtracting irrelevant from probe responses, except for accuracy, which is irrelevant – probe.

offset we recorded 3108 trials with blinking (55%). This is typical for blinking behavior, which tends to occur between processing stages rather than during those stages. Thus, blinks were analyzed for this 1250 ms window relative to stimulus offset.

In addition to greater mean blink-rate for probes just prior to the response, and an even larger one afterward, Fukuda (2001) also found similar differences between familiar-probe and irrelevant items using peak blink-rate and time-to-peak blink-rate measures. Thus, we predicted that each of these four sub-measures of the TDB would also show greater blinks-per-second for probes than irrelevants in the familiar-probe condition only. If the TDB during the familiar-probe condition contains the numerous deviations predicted here, then we would also expect that the entire TDB function (binned blinks over time) for probes would differ significantly from the irrelevant TDB during the probe condition only. Thus, we also analyzed TDB as a function of condition. If diagnostic, classification on this function alone may be preferable to classification based on various individual components.

Pupil measures

Based on prior research described earlier, we predicted that mean pupil-size would be greater on probe than irrelevant trials in the familiar-probe condition. However, Lubow and Fein (1996) also observed increased pupil-slopes for familiar-probe stimuli. Although slope was not analyzed, this effect was visually apparent in their graphs. Thus, we predicted that pupil-size would not only be greater on average for familiar-probes than irrelevants, but would grow faster over time. Pupil-slope was computed by fitting a least-squares regression line through each trial's pupil data (stimulus onset to response) and then computing the change in pupil-size over time represented by this line. Both mean pupil-size and pupil-slope measures were computed over pupil data from the first 1500 ms of each trial following stimulus onset. **Figure 4** depicts the mean standardized pupil data as a function of stimulus type and time during this period. Because in the current paradigm stimulus offset is concomitant with the response, this visual representation is sub-optimal; although probes and fillers

are represented throughout this range, toward the end there is a greater proportion of probe than irrelevant responses (c.f., mean RT pattern in **Figure 1**; **Table 1**).

CLASSIFICATION RATIONALE AND PROCEDURE

Overall mean differences between probe and irrelevant responses are not sufficient conditions for successful diagnostic tests. Often, CKT classification procedures consider the range of test outcomes (e.g., differences between probe and irrelevant responses), choose a cutoff value that maximizes the differentiation between these responses in the studied sample (e.g., the median value), and then report the resulting classification results using this cutoff (e.g., Farwell and Donchin, 1991; Lubow and Fein, 1996). A popular alternative method is to derive the optimal cutoff based from a receiver operating characteristic (ROC) analyses (Green and Swets, 1966; Bamber, 1975; Hanley, 1982), which includes an analysis of the tradeoff between a test's hit and false-alarm rates over a series of cutoffs. A poor test (efficiency near 0.5) is one in which hits and false-alarms are perfectly related so that a cutoff change that achieves a 1% increase in the hit rate results in the same increase in the false-alarm rate. An efficient test (efficiency near 1) allows the maximization of hit rate with minimum increases in false-alarm rate. Thus, ROC analysis offers a better understanding of the fitness of the test under investigation across a variety of cutoffs. To classify a group of responses from a CKT procedure, the cutoff that maximizes hit rate and minimizes false-alarm rate can be chosen and applied to the data.

Other classification approaches for CKT data that may involve determining cutoff points include maximum rank analysis (e.g., Lykken, 1959; Bradley and Warfield, 1984), discriminant-function analysis (Nose et al., 2009), and logistic regression analysis (Gamer et al., 2006; Gamer, 2011). The primary advantage of such techniques is their ability to model the relationship between the predictor variables and test outcomes (e.g., guilty vs. innocent). The resulting discriminant-function is then used to calculate hit and false-alarm rates for the sample. This allows researchers to understand the discriminability of the sample under investigation,

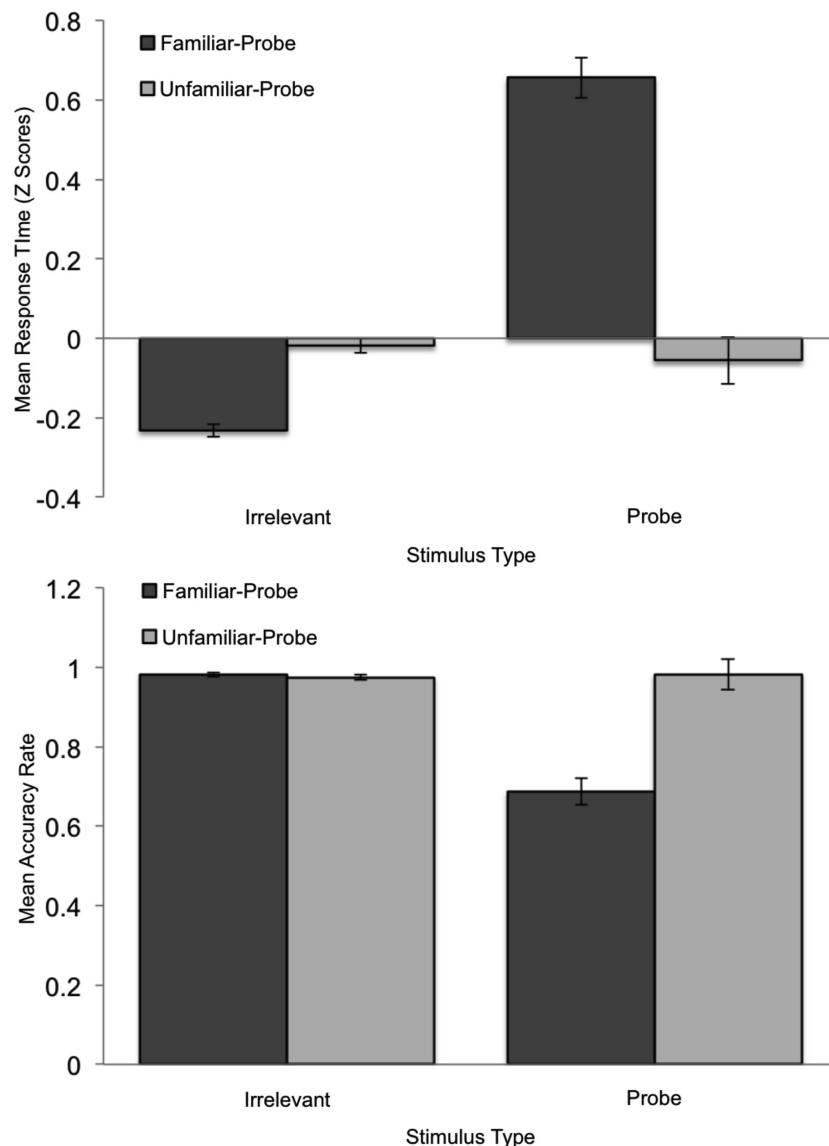


FIGURE 1 | Mean standardized RT (top graph) and Accuracy (bottom graph) plotted as a function of Stimulus Type and Condition. Error bars indicate standard error of the mean.

but may not give as clear a view of how well the discriminant function will classify data from future tests. This is not a flaw in these methods, but requires that researchers either generate the classification model on a subset of available data and use it to predict the remaining data, or use the entire dataset and use the same function for classification in subsequent tests (e.g., Bradley et al., 1996). The latter is particularly difficult to do successfully if subject demographics or test parameters change from test to test (e.g., stimulus modality, response deadline, response stimulus interval, etc.). Regardless of whether one classifies using median cutoffs, ranks, or one of the various methods of producing discriminant-functions, functions developed using existing participant data may need to be updated for successful classification of future participants. This is especially probable if subsequent

participants or test paradigms differ significantly from those used to develop the classification function.

In the present study, we avoid this particular concern by not basing classification on observed differences between probe and irrelevant responses in the current dataset and paradigm, but on theoretical ways in which any two distributions of responses may vary when produced by different psychological processes. In this way, the classification remains constant across changes to subjects, test parameters, or diagnostic measures.

Following Seymour et al. (2000) we used a *compound classification procedure* (CCP) in which each participant's distribution of probe RTs was compared to their irrelevant RT distribution. Seymour and colleagues used three separate statistical tests that evaluated whether response distributions differed with respect to

(a) the number of response errors (Fisher's exact test), (b) their shape or skew (Kolmogorov-Smirnov test), (c) and their variation of scores (variance-ratio test). It was assumed that relative to a distribution of unfamiliar irrelevant responses, a distribution of familiar-probe responses would contain more errors, would be less positively skewed, or would have a greater variance. It was further assumed that differences might emerge on all three tests, or some subset. Thus, a statistical difference on either test would lead to the conclusion that participants were familiar with probes (if accurate, a hit is recorded, otherwise it is a false-alarm). No statistical difference on any test indicated that participants were unfamiliar with probes (if accurate, a correct-rejection, otherwise a miss). Using the three-test CCP, Seymour et al. achieved hit rates of 0.98 and 0.93, and false-alarm rates of 0.02 and 0 using test alphas of 0.05 and 0.01, respectively. This analysis technique has no free parameters and allows data produced by any continuous measure to be evaluated. The nominal alpha level required for each test's significance is technically variable, however, it would be difficult to justify altering it beyond the standard 0.05 level. Due to the prohibitive nature of false-alarms in forensic contexts, it may be reasonable in some cases to reduce the level below 0.05 to make the test more conservative, but there is no more justification for increasing the alpha level above 0.05 than there would be for other statistical analyses in psychological research. Although Seymour and colleagues' initial report used a verbal phrase based CKT, similar hit rates (0.91) and false-alarm rates (0.03) were achieved in a subsequent test using face pictures as stimuli (Seymour and Kerlin, 2008).

As in previous studies (Seymour et al., 2000; Seymour and Kerlin, 2008; Seymour and Fraynt, 2009), response accuracy in the present study successfully discriminates between probe and irrelevant responses in the familiar-probe condition. Despite this, we chose not to include accuracy in classification analyses because in previous studies where incentives were promised (Seymour et al., 2000; Seymour and Fraynt, 2009), the accuracy effect was significantly attenuated. Such attenuation has also been noticed in paradigms that offered no explicit incentive (e.g., Rosenfeld et al., 2004). Thus, although the diagnosticity of combined measures that include accuracy would likely be enhanced here, it is not believed that such benefits would extend to future studies using incentives, or applied contexts involving natural incentives. Thus each individual and combined measure was evaluated on the basis of distribution variance and shape, but the Fisher exact test for number of errors was not used.

In Seymour et al. (2000) each participant completed both familiar-probe and unfamiliar-probe tests thus serving as their own control for the classification analysis. In the present study, probe-familiarity was manipulated between subjects; data from participants in the familiar-probe condition were used to analyze hit and miss rates, and data from the unfamiliar-probe condition were used to assess false-alarms and correct-rejections. For each participant, probe and irrelevant response distributions were compared using each individual and combination of measures. Each comparison involved two statistical analyses; a variance-ratio test, and a Kolmogorov-Smirnov test. Thus, each participant's probe and irrelevant response distributions were subject to 22 statistical comparisons (i.e., two statistical analyses for 11 individual and

combined measures). For each participant's statistical comparisons, a nominal alpha of 0.05 was assumed and Bonferroni corrected to 0.025.

Classification of each participant began with a variance-ratio test (also called the *F*-test for variances) to evaluate the one-tailed hypothesis that probe and irrelevant response distributions have different spreads. Subsequently, data were converted to overlapping cumulative distribution functions (normalized by sample size), and a Kolmogorov-Smirnov test (for review, see Kotz et al., 1983) was used to evaluate the one-tailed hypothesis that the cumulative probability at the maximum vertical deviation between the two curves, *D*, would be greater for probe than irrelevant distributions. The *D*-statistic ranges from 0 (no deviation) to 1 (maximal deviation). For sample sizes n_1 (probe = 18) and n_2 (irrelevant = 72), the corresponding *p*-value was determined by entering $D/S(n)$ into a *D*-statistic table, where $s(n) = \sqrt{n_1 + n_2/n_1 n_2}$. Values of 1.36 and 1.63 correspond to typical alpha levels of 0.05 and 0.01 and would require maximal deviations between distributions of 36 and 39% respectively. This statistic is particularly useful for comparing the shape of two response distributions because it is non-parametric. Also, unlike Student's *t*-test, it does not make assumptions about the underlying distribution and is not influenced by changes in scale.

In the CCP, a "hit" results (probe knowledge indicated) if any 1 of the constituent tests' null hypotheses is rejected. Lack of familiarity with probes is concluded only if neither test reaches statistical significance. A conservative threshold for significance balances the liberal nature of this rule. Bonferroni corrected alpha levels are used for each of the underlying statistical tests, so that a nominal alpha of 0.05 requires an actual difference between distributions at the $p < 0.025$ level. Additional care is warranted when comparing distributions that differ significantly in size, as is the case with each participant's probe ($n_1 = 18$) and irrelevant ($n_2 = 72$) distributions. For example, if probe and irrelevant distributions each contained 15 very slow RTs, this might suggest that such RTs are not diagnostic and the fact that mean probe RT is greater than mean irrelevant RT is an artifact of the small probe sample. This spurious difference may also manifest itself in the variance-ratio and K-S statistics, leading to an increased false-alarm rate. To address this issue, a Fisher randomization procedure (Fisher, 1935) is used to verify any significant differences that result from K-S or variance-ratio tests. First a participant's observed probe and irrelevant scores are pooled into one distribution of size $n_1 + n_2$. Then two new samples of sizes n_1 and n_2 are drawn without replacement and compared using the statistic of interest (K-S or variance-ratio, two tailed). After 1000 repetitions, if more than five statistical differences are found between these sampled distributions that equals or exceeds the original statistic for the observed distributions, the null hypothesis is accepted. The effect of this procedure is to essentially test how many probe-like responses are present in the observed irrelevant distribution. The more probe-like responses there are in the irrelevant distribution, the greater the chance of sampling a new probe distribution that is significantly different than a sampled irrelevant distribution using the statistic under investigation. If such a difference occurs more than 5 times out of 1000, the original statistical difference between the observed probe and irrelevant distributions is

considered spurious and recorded as having been non-significant. Thus, although either of the constituent tests in the CCP may be used to determine probe-familiarity, the standard of proof is relatively high. One result of this conservatism is that the default classification is an unfamiliar-probe one.

The CCP is related to the parallel testing method (Appendix K, National Research Council, 2003) in that a set of predictors is assessed and a critical result on either test indicates the presence of some target condition (e.g., disease, guilty, etc.), and only non-significant results on all measures indicates the absence of the target condition. One difference is that in the parallel testing method, independent methods are ideally sought so that the inclusion of additional tests incrementally increases the hit rate of the method. Alternatively, the CCP was designed to assess various aspects of the same characteristic – the shape of the response distribution – achieved using variance-ratio and K-S tests. The goal of this overlap is to address complete or partial tradeoffs in participants' responses to familiar-probe stimuli; they tend to be either more variable than irrelevants, more skewed than irrelevants, or both. A third test, Fisher's exact, was previously used to address the final tradeoff observed whereby participants would trade speed for accuracy more in familiar-probe than irrelevant responses (c.f. Seymour et al., 2000). Although multiple correlated measures are not generally ideal when trying to minimize misses and false-alarm, the corrected alpha level required for each additional test in the CCP, and possibly the need to pass the Fisher randomization procedure, may counteract this concern. Indeed, it is possible that the combination of these constraints causes the test to be overly cautious. As a result, if the measure under investigation is not sufficiently diagnostic, both false-alarm rates and hit rates may be lowered. Ultimately, the true impact of the CCP on a test's sensitivity and specificity would need to be modeled with statistical simulations. However, the low false-alarm rate and high hit rate previously reported using the CCP gives some indication that the low false-alarm rate does not come at the cost of an extreme number of misses.

RESULTS

Successful eye-tracking calibration of eight (13%) participants in the unfamiliar-probe condition was not possible. Thick eyeglasses, shifting contact lenses, and heavy applications of eyeliner make-up were among the most common obstacles. Thus, data from 52 participants (30 familiar-probe, 22 unfamiliar-probe) were included in the analysis.

OMNIBUS TESTS

Response time and accuracy analysis

Response time data were submitted to a 2 Condition (familiar-probe vs. unfamiliar-probe) \times 2 Stimulus Type (irrelevant, probe) mixed-model ANOVA with Stimulus Type as the within-subjects variable (see **Figure 1**, top graph). This analysis revealed main effects of Stimulus Type, $F(1, 50) = 23.36, p < 0.001, \eta^2 = 0.15$, and Condition, $F(1, 50) = 5.4, p = 0.02, \eta^2 = 0.06$, as well as a Condition \times Stimulus Type interaction, $F(1, 50) = 25.40, p < 0.001, \eta^2 = 0.16$. Participants in the familiar-probe condition took an average of 346 ms ($SD = 218$ ms) longer to respond "no" to familiar-probes than irrelevants, $t(29) = 8.68, p < 0.001$. In the

unfamiliar-probe condition, participants could not distinguish probes from irrelevants and no differences emerged.

A similar analysis was performed on accuracy data and is also plotted in **Figure 1** (bottom graph). This analysis revealed main effects of Stimulus Type, $F(1, 50) = 20.62, p < 0.001, \eta^2 = 0.16$, and Condition, $F(1, 50) = 8.51, p < 0.005, \eta^2 = 0.08$, as well as a Condition \times Stimulus Type interaction, $F(1, 50) = 52.73, p < 0.001, \eta^2 = 0.33$. Participants in the familiar-probe condition produced 29% ($SD = 23\%$) more response errors to familiar-probe faces than irrelevant faces, $t(29) = 6.92, p < 0.001$. No such difference emerged in the unfamiliar-probe condition.

Blinking analysis

To assess the overall TDB by condition, A 2 Condition (familiar-probe vs. unfamiliar-probe) \times 2 Stimulus Type (probe vs. irrelevant) \times 25 Time (50 ms bins) mixed-model ANOVA was performed on TDB data with Stimulus Type and Time as within-subjects variables. There was a significant main effect of Time due to the increase in blinking 200–400 ms after the manual response, $F(12.79, 639.60) = 28, p < 0.001, \eta^2 = 0.29$. Mauchly's test indicated that the assumption of sphericity had been violated for this effect ($\epsilon = 0.35$). Thus, degrees of freedom were corrected using Greenhouse-Geisser estimates. No other main effects or interactions were observed despite the large number of degrees of freedom available for this analysis.

To examine the predicted effects of peak blink-rate, time to reach peak blink-rate, and average blink-rate for the period 200–400 ms post-response, a set of 2 Condition (familiar-probe vs. unfamiliar-probe) \times 2 Stimulus Type (probe vs. irrelevant) mixed-model ANOVAs were performed on these measures, but each failed to yield significant main effects or interactions, $Fs < 1$. To examine the predicted effect of pre-response blink-rate, we analyzed differences between probe and irrelevant data that can be seen in **Figure 2** (top graph) for familiar-probes only, −400 to −100 ms relative to stimulus offset. Mean standardized blink-rates for bins during this period are plotted in **Figure 3** as a function of Condition and Stimulus Type. A 2 Condition (familiar-probe vs. unfamiliar-probe) \times 2 Stimulus Type (probe vs. irrelevant) mixed-model ANOVA was performed that yielded a main effect of Stimulus Type, $F(1, 50) = 5.60, p = 0.02, \eta^2 = 0.02$, and a Condition \times Stimulus Type interaction, $F(1, 50) = 3.78, p = 0.06, \eta^2 = 0.01$, approaching significance. A *post hoc* comparison revealed that in the familiar-probe condition, mean blink-rate during this period was 0.18 ($SD = 0.38$) b/s higher on probe than irrelevant trials, $t(29) = 2.63, p < 0.05$.

Pupil analysis

Figure 4 shows standardized pupil data over time as a function of Stimulus Type and Condition, and allows one to assess the sources of mean pupil and pupil-slope effects. These effects are summarized in **Figure 5** which depicts Z-Scores for the mean pupil-size data averaged over time as a function of Stimulus Type and Condition (top graph), as well as a similar plot of the pupil-slope data (bottom graph). The goal of the following analysis was to test the prediction that familiar-probe faces would lead to a greater mean pupil-size, and a greater pupil-slope compared to irrelevant faces.

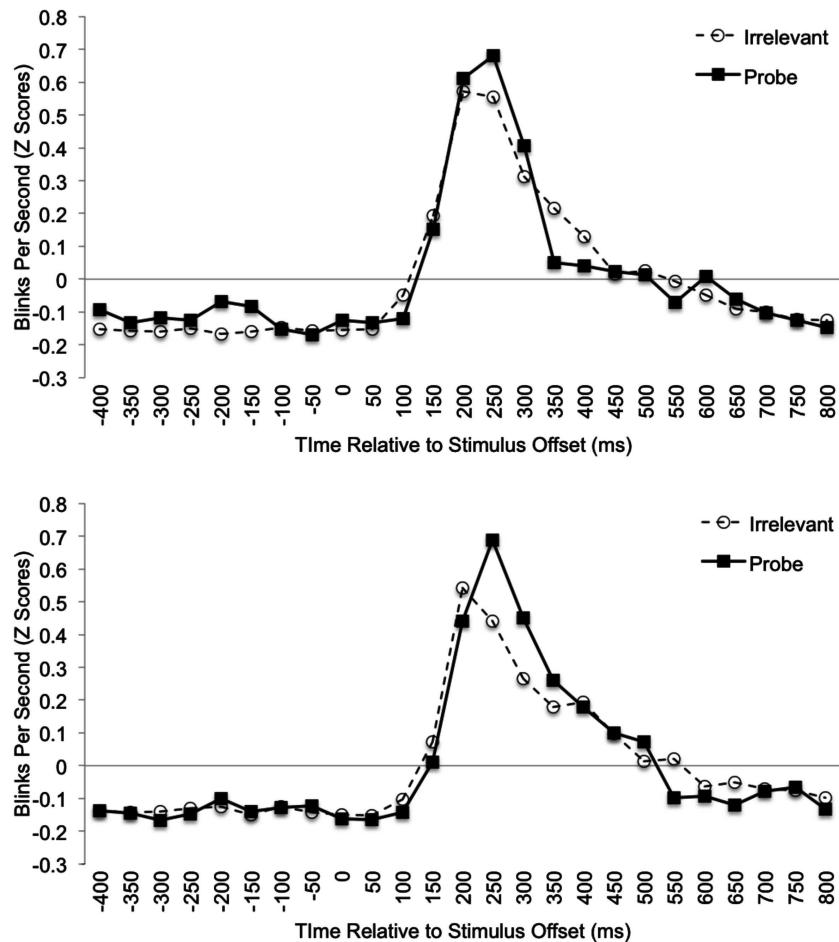


FIGURE 2 | Mean standardized blink data plotted as a function of Time (50 ms bins relative to stimulus offset) and Stimulus Type for the familiar-probe (top graph) and unfamiliar-probe (bottom graph) conditions.

A Condition (familiar-probe vs. unfamiliar-probe) \times 2 Stimulus Type (probe vs. irrelevant) mixed-model ANOVA was performed on mean pupil-size with Stimulus Type as the within-subjects variable and revealed a main effect of Stimulus Type, $F(1,50) = 27.28$, $p < 0.001$, $\eta^2 = 0.01$, as well as a Condition \times Stimulus Type interaction, $F(1,50) = 27.93$, $p < 0.001$, $\eta^2 = 0.01$. These results verify that mean pupil-size was 0.10 mm ($SD = 0.08$) greater on probe trials than irrelevant trials, $t(29) = 6.86$, $p < 0.001$, but only when probes were familiar. A similar analysis performed on pupil-slope revealed a main effect of Stimulus Type, $F(1,50) = 7.73$, $p < 0.01$, $\eta^2 = 0.02$, as well as a Condition \times Stimulus Type interaction, $F(1,50) = 23.1$, $p < 0.001$, $\eta^2 = 0.07$. This pattern of results is similar to the average pupil result and indicates that pupil-size grew 8% faster when viewing probe than irrelevant faces, $t(29) = 6.14$, $p < 0.001$, but only in the familiar-probe condition.

Classification analysis

The results of the classification analysis for the present data are listed in **Table 2** and show that RT led to more accurate classifications than pupil-size, $Z = 1.77$, $p < 0.05$, and slope, $Z = 2.71$,

$p < 0.01$. This was not true for RT vs. pre-response blink-rate, $Z = 1.43$, $p = 0.08$. Although combinations of RT and ocular measures produced higher classification rates than tests based on individual ocular measures, all $p < 0.05$, this was likely driven by significant differences between individual RT and pupil measures. Similarly, combining ocular measures did not significantly improve overall classification accuracy compared to pupil-size alone. However the hit rate achieved by combining pupil and blink measures was higher than pupil-size alone, $Z = 1.81$, $p < 0.05$. Bivariate correlations were calculated between RT and various ocular measures; we found that only the RT and pupil-size measures were significantly correlated, $r(30) = 0.65$, $p < 0.05$.

DISCUSSION

The primary goal of the present study was to examine whether RT and eye-based measures could be successfully used to detect concealed knowledge either alone or in combination. Although several studies have previously reported successful RT-based tests, previous ocular-based paradigms have less consistent successes and have yielded a wider range of false-alarm and miss rates. Because multiple aspects of the eyes' response to a stimulus can be assessed

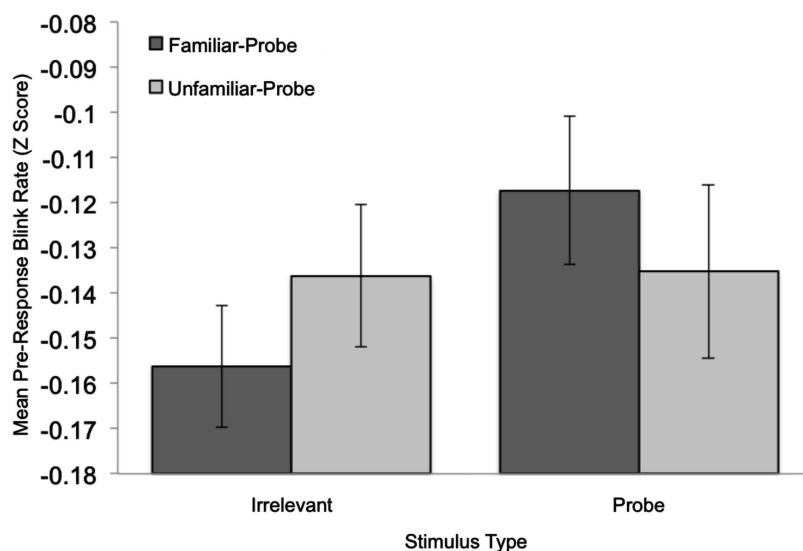


FIGURE 3 | Mean standardized pre-response (−100 to −400 ms relative to the response) blink-rate data plotted as a function of Stimulus Type and Condition. Error bars indicate standard error of the mean.

simultaneously using modern eye-trackers, we analyzed pupil-size, pupil-slope, average blink-rate, peak blink-rate, and overall temporal distribution of blinks. To our knowledge, no previous study has simultaneously examined RT and this array of ocular measures in a CKT paradigm.

Participants in this study learned sets of probe and target face pictures and were later asked to respond “yes” to indicate familiarity of target faces and “no” to indicate lack of familiarity with novel irrelevant faces. Participants also responded to probe faces and were asked to respond “no” regardless of whether the probes were the ones previously studied (familiar-probe condition), or whether the probes were novel faces (unfamiliar-probe condition). With this paradigm, we examined the individual and combined diagnosticity of RT, accuracy, and multiple indices of pupil and blink responding. For individual measures we predicted that responsiveness would be greater on probe than irrelevant trials, but only in the familiar-probe condition.

PERFORMANCE OF INDIVIDUAL MEASURES

Consistent with predictions, participants were significantly slower and less accurate when responding “no” to familiar-probe faces compared to irrelevants. This pattern of results for RT and accuracy measures is similar to ones previously reported with the CKT paradigm (e.g., Allen et al., 1992; Seymour et al., 2000; Seymour and Kerlin, 2008; Seymour and Fraynt, 2009; Verschueren et al., 2010; Visu-Petra et al., 2011). Based on work by Lubow and Fein (1996), we also predicted that average pupil-size and mean pupil-slope would be greater when responding to probes compared to irrelevants in the familiar-probe condition. Although Lubow and Fein reported a successful test based on mean-pupil size, they only commented on apparent differences in pupil-slope. The present results show that pupil-size grows faster and achieves a greater final size on familiar-probe trial than irrelevant trials. For blinking behavior, numerous predictions were made following Fukuda

(2001)’s successful demonstration that the way blinking is distributed over the course of test trials (especially the period before and after the overt response) can discriminate between those with and without concealed knowledge. Unfortunately, an analysis of the overall function relating blinking to time (temporal distribution of blinks) compared across conditions did not reach statistical significance. This was also true for predicted increases in related peak blink-rate and time-to-peak blink-rate measures; these showed no sensitivity to probe-familiarity. Fukuda reported greater increases in blink-rate just prior to the overt response, and also just after the response. In the present data, a similar prediction for the post-response blink-rate was not supported; significant increased blinking was noticed, but this increase was not greater for familiar-probe stimuli. Our final prediction for blinking was based on Fukuda’s pre-response blink-rate finding. Here we did find a small, but statistically significant increase in blinking for familiar-probe trials compared to irrelevants in the period −400 to −100 ms before to the overt response. Interestingly, this increase in blink-rate was most prominently observed in the averaged data for the period between 250 and 100 ms prior to response onset (see Figure 3). The lack of effect during the final 100 ms of this period suggests that pre-response blinks may be indexing a single, late, processing stage associated with concealed knowledge responding and is consistent with a recently proposed response-conflict based model (Seymour and Schumacher, 2009; Schumacher et al., 2010). In their Parallel Task-Set model, Seymour and colleagues offers an account of both the timing of response-conflict in the CKT paradigm and the additional variance in processing observed for familiar-probe trials. Overall, we found that RT, accuracy, pre-response blink-rate, pupil-size, and pupil-slope measures each differentiated responses in the familiar and unfamiliar-probe conditions. To our knowledge, this is the first demonstration of a CKT paradigm simultaneously assessing these measures.

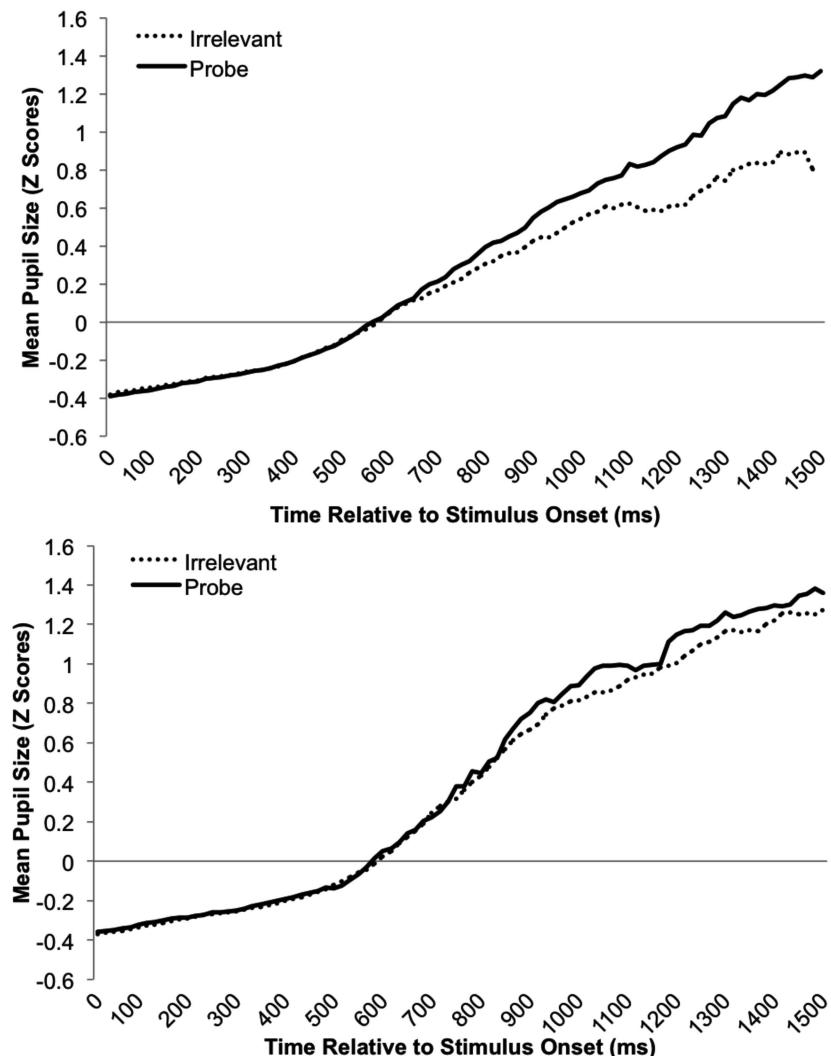


FIGURE 4 | Mean standardized pupil-size data plotted as a function of time (from stimulus onset to 1500 ms afterward) and stimulus type for the familiar-probe (top graph) and unfamiliar-probe (bottom graph) conditions.

COMPOUND CLASSIFICATION PROCEDURE

A CCP comparing probe and irrelevant distributions on shape and variance was used. Significant differences between probe and irrelevant distributions on the basis of shape or variance indicated familiarity with the probe faces. Although this procedure has been used in previous studies (Seymour et al., 2000; Seymour and Kerlin, 2008; Seymour and Fraynt, 2009), the present study is the first to describe this procedure in detail, and the first demonstration of its fitness for data other than RT and Accuracy. The 0.98 classification rate observed with the RT measure was comparable to the 0.92–0.97 rates previously reported using this paradigm (Seymour et al., 2000; Seymour and Kerlin, 2008; Seymour and Fraynt, 2009). Similarly, the pupil-size measure yielded a higher overall classification rate (0.92) here than the 0.66–0.88 rates typically reported (Janisse and Bradley, 1980; Lubow and Fein, 1996). Although tests based on combined measures yielded high classification rates, they were not overall more accurate than using RT in isolation. We

note that the failure of compound measures to outperform singular ones was not due to correlations between various measures, as only RT and pupil-size were correlated.

For the pupil-slope measure, it is less clear how to interpret previous studies. Although slope changes were noted previously in Lubow and Fein's (1996) pupil-size based paradigm, classification accuracy using pupil-slope was not provided. Overall, the performance of the present slope-based analysis was less impressive than those using pupil-size and blink measures. This is more likely to be a result of the relatively low discriminability of the slope measure rather than limitations of the CCP. Although no participants in the unfamiliar-probe condition showed slope differences between probe and irrelevant stimuli, 30% of participants in the familiar-probe condition also failed to show such differences, resulting in a relatively high miss rate. However, the overall 85% classification accuracy provided by the slope measure was equally high as Lubow and Fein's pupil-size based test.

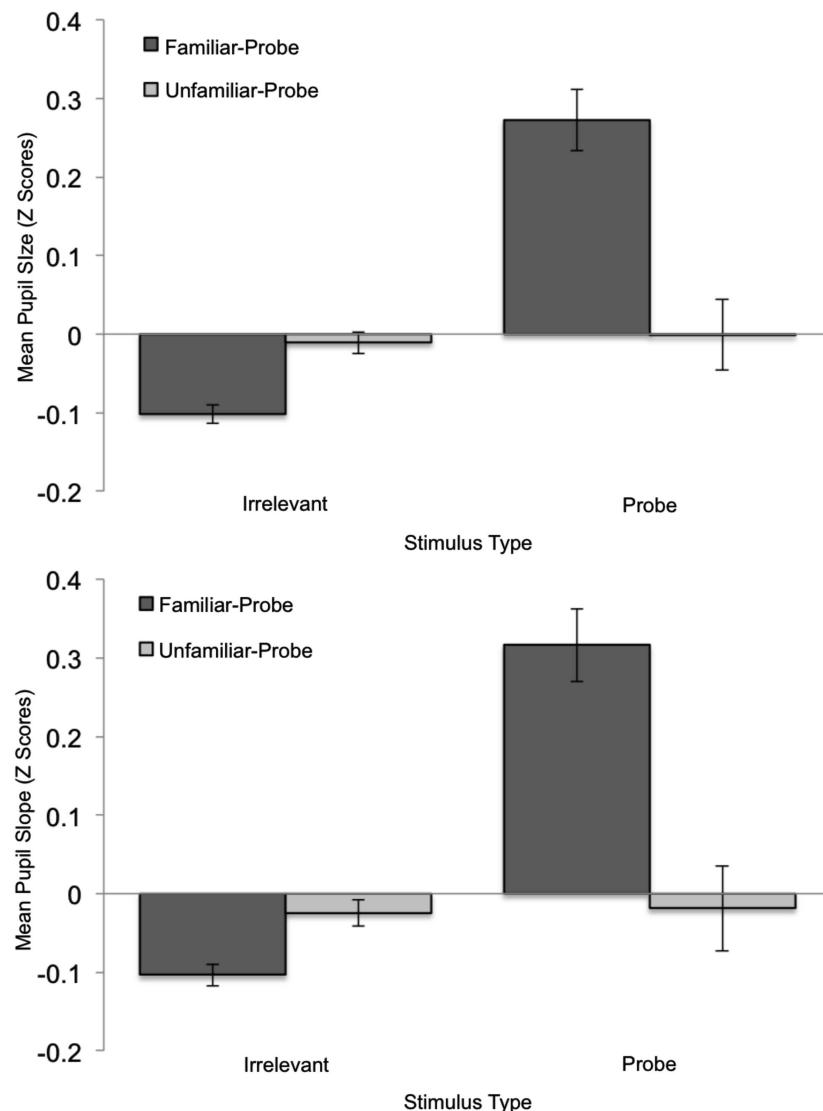


FIGURE 5 | Mean standardized pupil-size (top graph) and pupil-slope (bottom graph) plotted as a function of Stimulus Type and Condition. Error bars indicate standard error of the mean.

Although in the present study the overall temporal distribution of blinks did not discriminate familiar-probe and irrelevant trials as in Fukuda's (2001) study, we did find the predicted difference in mean blink-rate just prior to the response. When analyzed using the CCP, blink-rate yielded an overall classification rate of 0.93, comparable to performance of the pupil-size measure (0.85), and not statistically different than the classification rate using RT (0.98). This was surprising for a mean difference of less than one-quarter blink per second. This result highlights an important advantage of the CCP's focus on the shape and variance of response distributions instead of a single cutoff value: it is less affected by the distribution overlap if the distributions have different shapes (e.g., Farwell and Donchin, 1991). Leal and Vrij (2010) also examined blink responses during a CKT and found hit and

false-alarm rates (0.75 and 0.23, respectively) lower than in the present study (0.90 and 0.045 respectively). One possible source of this difference is the nature of their analysis window; it was only reported that blinks were analyzed "during an arbitrarily defined 10 s window" between stimulus onset and the vocal response. This window would be a super-set of the one analyzed in the present study in which only a small subset proved diagnostic (i.e., the 400 ms just prior to response onset). Thus, it is possible that Leal and Vrij averaged over a relatively small amount of diagnostic and a large amount of non-diagnostic blink data, artificially limiting the accuracy of their classification. If this is the case, then results from the present study and previous studies may yet indicate that blink analysis of concealed knowledge is more promising than previously thought.

Table 2 | Results from the compound classification procedure (variance-ratio and Kolmogorov-Smirnov tests only) for individual and combined measures.

Predictor	Hit rate	False-alarm rate	Classification accuracy
INDIVIDUAL MEASURES			
Manual			
Response time (RT)	0.97	0	98%
Ocular			
Pupil-size	0.83	0	92%
Pupil-slope	0.70	0	85%
Pre-resp. blink-rate	0.90	0.045	93%
COMBINED MEASURES			
Manual and ocular			
RT + pupil	1.0	0	100%
RT + slope	0.97	0	98%
RT + blink	1.0	0.045	98%
All (RT + ocular)	1.0	0.045	98%
Ocular			
Pupil + slope	0.90	0	94%
Pupil + blink	0.97	0.045	96%
Pupil + slope + blink	0.97	0.045	96%

COMBINED VS. INDIVIDUAL MEASURES

Classification analyses were performed to examine the prediction that tests based on multiple measures would outperform those using individual measures. Although most combined measures led to higher detection rates than individual measures, few improvements were statistically significant; one notable exception was found using ocular measures. Although combining measures did not change the false-alarm rate, combining pupil-size and blink-rate measures led to a significantly greater classification than using pupil-size alone. Otherwise, combined tests appeared to offer only minor improvements; most likely due to the strong performance produced by the individual measures (RT in particular). We found a correlation between RT and pupil-size measures, but not between pupil-size and blink measures. This may explain why the RT + pupil combined-measure failed to improve upon RT alone, whereas the pupil + blink measure did. Thus, it appears that the high individual classification accuracy of some individual measures may have constrained the improvement offered by combinations. Similarly, high correlations between ocular and electrodermal measures (e.g., Bradley et al., 2008) suggest that other limitations may exist for combinations involving ocular measures.

SUGGESTIONS FOR FUTURE RESEARCH

One limitation of the present study is in its ability to consider pupil-size independent of RT. This is due to the fact that trials ended immediately following the overt manual response. Because collection of pupil data also ended on each trial concomitant with the response, it is possible that the pupil-size based concealed knowledge effect is solely an indication of the larger RTs on familiar-probe trials relative to irrelevant trials. The significant correlation between RT and pupil-size supports this alternative explanation. Follow-up studies that lack an overt response, or in

which the collection of pupil data continues for some time following the response, would be informative. However, Lubow and Fein's (1996) report of successful pupil-size based CKTs without response-terminated pupil recording tempers this interpretation somewhat. Furthermore, the presence of pupil-slope effects here (which were not correlated with RT) and in Lubow and Fein's study suggests that average pupil-size differences between probe and irrelevant trials are not merely a result of the passage of time. Despite these caveats, further investigation is warranted. One interesting alternative for avoiding dependence on overt responses is to use more complex stimuli (e.g., sentences or picture arrays) and examine ocular scan-patterns during the CKT (e.g., Webb et al., 2009a,b; Kircher et al., 2010; Cook et al., 2012).

Another issue for further study involves a detailed comparison of the CCP with the diverse range of previously reported CKT classification procedures. For example, the implications of using correlated measures of underlying response distribution morphology, the effect of the corrected alphas, and the influence of the Fishers randomization procedure would need to be modeled statistically to properly distinguish the CCP from related techniques such as the Independent Parallel Testing procedure (IPT; National Research Council, 2003), discriminant-function analysis (e.g., Nose et al., 2009), and logistic regression analysis (e.g., Gamer et al., 2006; Gamer, 2011). Of particular interest is how exactly hit rates and false-alarm rates are affected by each additional CCP sub-test. It is also unknown whether the CCP offers a significant advance over straightforward modifications to established approaches such as ROC analysis (Green and Swets, 1966; Bamber, 1975; Hanley, 1982). Such comparisons with the CCP would need to consider its primary design feature; reliance on generic differences between response distributions. This focus on only ways in which two distributions may vary in CKT-related paradigms (deviation, skewness, and in some cases number of observations; c.f., Seymour and Schumacher, 2009) means that there is no need to vary classification parameters between tests, even if test parameters or subject demographics change. Unlike some discriminant-function based procedures, its fitness is not based on a limited set of previously observed data. Thus, the only parameter that can change is the nominal alpha for the constituent statistical tests, and this would only be justifiable if the test were made more strict, but not less. Such a change would be in service of an even lower tolerance of false-alarms than offered by the standard alpha level of 0.05, and not the nature of the underlying test.

The closest alternative to the CCP is the IPT approach, however, the constituent tests can be anything, and the cutoffs used for these tests may vary from one use to the next. For example, Meijer et al. (2007) reported such a procedure for successfully combining performance on a skin-conductance based CKT with performance on a test of malingering. Although each test used standard task-specific cutoffs to classify subjects prior to the combined classification using IPT, such classifications could have been decided using a number of potential decision policies; each having a potential impact on this test's sensitivity and specificity (National Research Council, 2003). In contrast, the constituent tests for the CCP are always statistical hypothesis tests; combining measures occurs prior to classification and results in two distributions of combined

scores (one for probes and one for irrelevants) that are then compared statistically. Thus, it may be useful to investigate whether or not an IPT modified to accept the raw score from individual or combined CKT measures would be effectively similar to the CCP.

The present work has implications for applied work in forensic settings. However, an important next step in this research is to examine combined efficiency of ocular and RT measures in paradigms using mock-crimes that facilitate variable probe encoding, and longer retention intervals that would allow for forgetting or interference (e.g., Carmel et al., 2003). Such manipulations have previously been shown to modulate the effectiveness of the RT measure and may provide more room for the contribution of simultaneous ocular measures along with RT (Seymour and Fraynt, 2009). Such research may also employ explicit countermeasure instructions to manipulate the motivation to “beat the test.” Countermeasure manipulations are sometimes sufficient to attenuate the RT-based concealed knowledge effect (e.g.,

Rosenfeld et al., 2004), but not always (Seymour et al., 2000; Seymour and Fraynt, 2009). Even with such manipulations, it may be possible that conducting CKT research using undergraduate populations who lack the intrinsic motivation to deceive found in applied contexts limits the generalizability of our results. However, despite larger effect-sizes on average for laboratory settings compared to tests in the field, such differences do not always affect classification accuracy. For example, a study by Pollina et al. (2004) showed that classification accuracy was similar in laboratory and field-tests, despite differences in effect-sizes. Similarly, in a meta-analysis of CKT studies reported by Ben-Shakhar and Elaad (2003), it was shown that significant differences in test effect-sizes resulted when highly motivated participants ($d = 1.76$) were compared to those with low motivation ($d = 1.34$). However, this disparity failed to result in differences in respective test efficiencies ($a = 0.82$ and 0.80 respectively, for high and low motivation participants).

REFERENCES

- Allen, J. J., Iacono, W. G., and Danielson, K. D. (1992). The identification of concealed memories using the event-related potential and implicit behavioral measures: a methodology for prediction in the face of individual differences. *Psychophysiology* 29, 504–522.
- Bagley, J., and Manelis, L. (1979). Effect of awareness on an indicator of cognitive load. *Percept. Mot. Skills* 49, 591–594.
- Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *J. Math. Psychol.* 12, 387–415.
- Bauer, L. O., Goldstein, R., and Stern, J. A. (1987). Effects of information-processing demands on physiological response patterns. *Hum. Factors* 29, 213–234.
- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychol. Bull.* 91, 276–292.
- Beatty, J. (1988). “Pupillometric signs of selective attention in man,” in *Neurophysiology and Psychophysiology: Experimental and Clinical Applications*, eds G. C. Galbraith, M. L. Kietzman, and E. Donchin (Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.), 138–143.
- Ben-Shakhar, G. (1985). Standardization within individuals: a simple method to neutralize individual differences in skin conductance. *Psychophysiology* 22, 292–299.
- Ben-Shakhar, G., and Elaad, E. (2003). The validity of psychophysiological detection of information with the guilty knowledge test: a meta-analytic review. *J. Appl. Psychol.* 88, 131–151.
- Ben-Shakhar, G., and Furedy, J. J. (1990). *Theories and Applications in the Detection of Deception: A Psychophysiological and International Perspective*. New York, NY: Springer Verlag Publishing, 169.
- Berrien, F. K., and Huntington, G. H. (1942). An exploratory study of pupillary responses during deception. *J. Exp. Psychol.* 32, 443–449.
- Bradley, M. M., Miccoli, L., Escrig, M. A., and Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology* 45, 602–607.
- Bradley, M. T., MacLaren, V. V., and Carle, S. B. (1996). Deception and nondeception in guilty knowledge and guilty actions polygraph tests. *J. Appl. Psychol.* 81, 153–160.
- Bradley, M. T., and Warfield, J. F. (1984). Innocence, information, and the guilty knowledge test in the detection of deception. *Psychophysiology* 21, 683–689.
- Bruce, V. (1982). Changing faces: visual and non-visual coding processes in face recognition. *Br. J. Psychol.* 73, 105–116.
- Carmel, D., Dayan, E., Naveh, A., Raveh, O., and Ben-Shakhar, G. (2003). Estimating the validity of the guilty knowledge test from simulated experiments: the external validity of mock crime studies. *J. Exp. Psychol. Appl.* 9, 261–269.
- Cook, A. E., Hacker, D. J., Webb, A. K., Osher, D., Kristjansson, S. D., Woltz, D. J., et al. (2012). Lyin’eyes: ocular-motor measures of reading reveal deception. *J. Exp. Psychol. Appl.* 18, 301.
- Cutrow, R. J., Parks, A., Lucas, N., and Thomas, K. (1972). The objective use of multiple physiological indices in the detection of deception. *Int. J. Psychophysiol.* 40, 239–245.
- Gamer, M. (2011). “Detecting concealed information using autonomic measures,” in *Memory Detection: Theory and Application of the Concealed Information Test*, eds B. Verschueren, G. Ben-Shakhar, and E. Meijer (Cambridge University Press), 27–45.
- Gamer, M., Rill, H. G., Vossel, G., and Gödert, H. W. (2006). Psychophysiological and vocal measures in the detection of guilty knowledge. *Int. J. Psychophysiol.* 60, 76–87.
- Gamer, M., Verschueren, B., Crombez, G., and Vossel, G. (2008). Combining physiological measures in the detection of concealed information. *Physiol. Behav.* 95, 333–340.
- Goldstein, R., Bauer, L. O., and Stern, J. A. (1992). Effect of task difficulty and interstimulus interval on blink parameters. *Int. J. Psychophysiol.* 13, 111–117.
- Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York, NY: John Wiley.
- Gronau, N., Ben-Shakhar, G., and Cohen, A. (2005). Behavioral and physiological measures in the detection of concealed information. *J. Appl. Psychol.* 90, 147–158.
- Hakerem, G. A. D., and Sutton, S. (1966). Pupillary response at visual threshold. *Nature* 212, 485–486.
- Hancock, P. (2004). *Psychological Image Collection at Stirling*. Available at: <http://pics.psych.stir.ac.uk/>
- Hanley, J. A. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 743, 29–36.
- Heilveil, I. (1976). Deception and pupil size. *J. Clin. Psychol.* 32, 675–676.
- Holland, M., and Tarlow, G. (1972). Blinking and mental load. *Psychol. Rep.* 31, 119–127.

- Ichikawa, N., and Ohira, H. (2004). Eyeblink activity as an index of cognitive processing: temporal distribution of eyeblinks as an indicator of expectancy in semantic priming. *Percept. Mot. Skills* 98, 131–140.
- Janisse, M. P., and Bradley, M. T. (1980). Deception, information and the pupillary response. *Percept. Mot. Skills* 50, 748–750.
- Karatekin, C., Couperus, J. W., and Marcus, D. J. (2004). Attention allocation in the dual-task paradigm as measured through behavioral and psychophysiological responses. *Psychophysiology* 41, 175–185.
- Kircher, J. C., Cook, A. E., and Hacker, D. J. (2010). *Deception Detection using Oculomotor Movements*. Available at: <http://www.google.com/patents?hl=en&lr=&vid=USPATAPP12820488&id=4i3wAAAAEBAJ&oi=fnd&dq=kircher+cook+pupil&printsec=abstract>
- Kotz, S., Johnson, N. L., and Read, C. B. (1983). “Kolmogorov-Smirnov statistics,” in *Encyclopedia of Statistical Sciences*, Vol. 4, eds S. Kotz, N. L. Johnson, and C. B. Read (New York, NY: John Wiley & Sons), 393–396.
- Leal, S., and Vrij, A. (2008). Blinking during and after lying. *J. Nonverbal Behav.* 32, 187–194.
- Leal, S., and Vrij, A. (2010). The occurrence of eye blinks during a guilty knowledge test. *Psychol. Crime Law* 16, 349–357.
- Lubow, R. E., and Fein, O. (1996). Pupillary size in response to a visual guilty knowledge test: new technique for the detection of deception. *J. Exp. Psychol. Appl.* 2, 164–177.
- Lykken, D. T. (1959). The GSR in the detection of guilt. *J. Appl. Psychol.* 43, 385–388.
- Lykken, D. T. (1998). *A Tremor in the Blood: Uses and Abuses of the Lie Detector*. New York, NY: Plenum Press.
- MacLaren, V. V. (2001). A qualitative review of the guilty knowledge test. *J. Exp. Psychol. Appl.* 86, 674–683.
- Meijer, E. H., Smulders, F. T. Y., Johnston, J. E., and Merckelbach, H. L. G. J. (2007). Combining skin conductance and forced choice in the detection of concealed information. *Psychophysiology* 44, 814–822.
- National Research Council. (2003). *The Polygraph and Lie Detection*. Washington, DC: National Academies Press.
- Nose, I., Murai, J., and Taira, M. (2009). Disclosing concealed information on the basis of cortical activations. *Neuroimage* 44, 1380–1386.
- Patalano, A. L., and Seifert, C. M. (1994). Memory for impasses during problem solving. *Mem. Cognit.* 22, 234–242.
- Pollina, D. A., Dollins, A. B., Senter, S. M., Krapohl, D. J., and Ryan, A. H. (2004). Comparison of polygraph data obtained from individuals involved in mock crimes and actual criminal investigations. *J. Appl. Psychol.* 89, 1099.
- Rosenfeld, J. P., Biroshak, J. R., and Furedy, J. J. (2006). P300-based detection of concealed autobiographical versus incidentally acquired information in target and non-target paradigms. *Int. J. Psychophysiol.* 60, 251–259.
- Rosenfeld, J. P., Cantwell, B., Nasman, V. T., Wojdac, V., Ivanov, S., and Mazzeri, L. (1988). A modified, event-related potential-based guilty knowledge test. *Int. J. Neurosci.* 24, 157–161.
- Rosenfeld, J. P., Soskins, M., Bosh, G., and Ryan, A. (2004). Simple, effective countermeasures to P300-based tests of detection of concealed information. *Psychophysiology* 41, 205–219.
- Schneider, W., Eschman, A., and Zuccolotto, A. (2002). *E-Prime: User's Guide*. Sharpsburg: Psychology Software Incorporated.
- Schumacher, E. H., Seymour, T. L., and Schwarb, H. (2010). Brain activation evidence for response conflict in the exclude recognition task. *Brain Res.* 1329, 113–123.
- Seymour, T. L., and Fraynt, B. R. (2009). Time and encoding effects in the concealed knowledge test. *Appl. Psychophysiol. Biofeedback* 34, 177–187.
- Seymour, T. L., and Kerlin, J. R. (2008). Successful detection of verbal and visual concealed knowledge using an RT-based paradigm. *Appl. Cogn. Psychol.* 22, 475–490.
- Seymour, T. L., and Schumacher, E. H. (2009). Electromyographic evidence for response conflict in the exclude recognition task. *Cogn. Affect. Behav. Neurosci.* 9, 71–82.
- Seymour, T. L., Seifert, C. M., Shafto, M. G., and Mosmann, A. L. (2000). Using response time measures to assess “guilty knowledge.” *J. Appl. Psychol.* 85, 30–37.
- Steinhauer, S. R., and Hakerem, G. (1992). “The pupillary response in cognitive psychophysiology and schizophrenia,” in *Psychophysiology and Experimental Psychopathology: A Tribute to Samuel Sutton*, eds D. Friedman and G. E. Bruder (New York, NY: New York Academy of Sciences), 182–204.
- Stern, J., Walrath, L. C., and Goldstein, R. (1984). The endogenous eyeblink. *Psychophysiology* 21, 22–33.
- Verschueren, B., Ben-Shakhar, G., and Meijer, E. (2011). *Memory Detection: Theory and Application of the Concealed Information Test*, 1st Edn. New York: Cambridge University Press.
- Verschueren, B., Crombez, G., Degrootte, T., and Rosseel, Y. (2010). Detecting concealed information with reaction times: validity and comparison with the polygraph. *Appl. Cogn. Psychol.* 24, 991–1002.
- Verschueren, B., Crombez, G., Koster, E. H., Van Bockstaele, B., and De Clercq, A. (2007). Startling secrets: startle eye blink modulation by concealed crime information. *Biol. Psychol.* 76, 52–60.
- Visu-Petra, G., Bus, I., and Miclea, M. (2011). Detecting concealed information from a mock crime scenario by using psychophysiological and RT-based measures. *Cogn. Brain Behav.* 15, 19–37.
- Webb, A., Hacker, D., Osher, D., Cook, A., Woltz, D., Kristjansson, S., et al. (2009a). Eye movements and pupil size reveal deception in computer administered questionnaires. Foundations of augmented cognition. *Neuroergon. Oper. Neurosci.* 553–562.
- Webb, A. K., Honts, C. R., Kircher, J. C., Bernhardt, P., and Cook, A. E. (2009b). Effectiveness of pupil diameter in a probable lie comparison question test for deception. *Legal Criminol. Psychol.* 14, 279–292.
- Williams, J. M. G., Mathews, A., and MacLeod, C. (1996). The emotional stroop task and psychopathology. *Psychol. Bull.* 120, 3.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 28 September 2012; **accepted:** 21 December 2012; **published online:** 04 February 2013.

Citation: Seymour TL, Baker CA and Gaunt JT (2013) Combining blink, pupil, and response time measures in a concealed knowledge test. *Front. Psychology* 3:614. doi: 10.3389/fpsyg.2012.00614

This article was submitted to Frontiers in Cognitive Science, a specialty of Frontiers in Psychology.

Copyright © 2013 Seymour, Baker and Gaunt. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.



Detecting false intent using eye blink measures

Frank M. Marchak*

Veridical Research and Design Corporation, Bozeman, MT, USA

Edited by:

Wolfgang Ambach, Institute for Frontier Areas of Psychology and Mental Health (IGPP), Germany

Reviewed by:

Chris N. H. Street, University College London, UK
Kyosuke Fukuda, Fukuoka Prefectural University, Japan

***Correspondence:**

Frank M. Marchak, Veridical Research and Design Corporation, 211 West Main Street – Lower Level, PO Box 6503, Bozeman, MT 59771, USA
e-mail: fmarchak@vradc.com

Eye blink measures have been shown to be diagnostic in detecting deception regarding past acts. Here we examined—across two experiments with increasing degrees of ecological validity—whether changes in eye blinking can be used to determine false intent regarding future actions. In both experiments, half of the participants engaged in a mock crime and then transported an explosive device with the intent of delivering it to a “contact” that would use it to cause a disturbance. Eye blinking was measured for all participants when presented with three types of questions: relevant to intent to transport an explosive device, relevant to intent to engage in an unrelated illegal act, and neutral questions. Experiment 1 involved standing participants watching a video interviewer with audio presented ambiently. Experiment 2 involved standing participants questioned by a live interviewer. Across both experiments, changes in blink count during and immediately following individual questions, total number of blinks, and maximum blink time length differentiated those with false intent from truthful intent participants. In response to questions relevant to intent to deliver an explosive device vs. questions relevant to intent to deliver illegal drugs, those with false intent showed a suppression of blinking during the questions when compared to the 10 s period after the end of the questions, a lower number of blinks, and shorter maximum blink duration. The results are discussed in relation to detecting deception about past activities as well as to the similarities and differences to detecting false intent as described by prospective memory and arousal.

Keywords: credibility assessment, false intent, blink, oculometrics, deception detection

INTRODUCTION

Interest in determining veracity has a long history, with documented methodologies extending back to at least 900 BC [for a review, see Trovillo (1939a,b)]. The majority of these efforts have focused on the detection of deception, a term which has a variety of characterizations (Masip et al., 2004). A widely accepted definition of deception is provided by Vrij (2008) as “a successful or unsuccessful deliberate attempt, without forewarning, to create in another a belief which the communicator considers to be untrue” (p. 15). While the temporal period during which this untruth is committed is unspecified, the majority of research has focused on detecting indicators of concealed past behavior.

The question arises concerning the ability of determining whether an individual is being deceptive regarding future intentions. In contrast to knowledge of past activities, intent involves a goal or plan of action for the future, in which both the execution of the action and its outcome are uncertain. As such, intent may be defined as “a person’s mental representations of his/her planned future actions” (Vrij et al., 2011a), and by extension, false intent involves misleading others regarding upcoming but not yet realized actions.

There have been efforts recently to determine if an individual is misleading others regarding the true purpose, or intent, of their future actions. Vrij and colleagues (2011a); Vrij and his colleagues (2011b) have shown differences in verbal responses to questions between truthful intent participants and those with false intent in terms of number of words, level of details, plausibility,

contradictions, and corrections. Meixner and Rosenfeld (2011), using a P300-based concealed information test, were able to detect with a high level of accuracy individuals who planned a mock terrorism attack. Aikins et al. (2010) examined changes in respiratory sinus arrhythmia (RSA) in participants that were either truthful or were to respond deceptively about a future mock crime. The data showed greater reductions in RSA during testing of participants that were deceptive regarding an upcoming task compared to the truthful intent participants.

Since being untruthful regarding both past and future acts includes the attribute of a desire to mislead, it has been hypothesized that cues indicative of false intent arise from analogous emotional, cognitive, and behavioral processes involved in deception (Martin et al., 2007). Traditional deception detection techniques focus on changes in measures of autonomic functions—such as respiration, cardiac activity, and electrodermal activity—as indicators of deceptive responses regarding previous activities (Abrams, 1989). However, other measures, such as changes in ocular parameters including eye movements, pupil diameter, and blink rate, have been examined as possible alternative markers of deception (e.g., Marchak et al., 2011; Cook et al., 2012).

One of the first examinations of these indices was conducted by Cutrow et al. (1972). Evaluating multiple physiological measures, including eye blinks, they concluded that eye blink rate decreases under circumstances of lying. Fukuda (2001) investigated the temporal distribution of eye blinks while subjects performed a guilty knowledge test (GKT) with playing cards. It

was found that more eye blinks occurred before responses following presentation of the relevant card while more eye blinks occurred after responses for presentation of irrelevant cards. DePaulo et al. (2003) reviewed the literature on deception detection and identified over 100 cues to deception. Of relevance here, they found that blinking was more prevalent when lies involved transgressions, discovery of which could have serious consequences. In another study employing emotionally arousing stimuli in a GKT, Thonney et al. (2005) found a difference in GKT eye blinking scores for emotional stimuli over neutral stimuli.

Leal and Vrij (2008) examined blink rates in liars and truth tellers during and after verbal recall of events and found that liars showed a decrease in blink rate during deception as compared with a baseline period and an increase in blink rate in the period following the telling of the lie. In a study of the GKT with the same subjects (Leal and Vrij, 2010) they found that liars exhibited a lower blink rate in response to key items as compared to control items, but there was no difference for truth tellers.

Taken together, these findings suggest that blink parameters are diagnostic in determination of deception regarding past actions. These differences can be explained both by theories of cognitive load (e.g., Fogarty and Stern, 1989; Fukuda et al., 2005; Irwin and Thomas, 2010) as well as arousal-based theories (e.g., Stern, 1992). While it is difficult to isolate the specific cause of the differences in blink behavior between liars and truth tellers, it appears that the findings are reliable and repeatable.

The question arises if these same measures can be used to determine whether an individual is being deceptive regarding future intentions as opposed to past actions. The present work attempted to determine if changes in eye blink parameters could be used to detect false intent. Based on the findings of differences in blink parameters between truthful individuals and those deceptive about prior actions, the primary hypothesis is that individuals with false intent will exhibit a suppression of blink rate during intent relevant questions, accompanied by a rebound in blink rate in the period following the question end, as well as a lower overall number of blinks and shorter blink durations when compared to those with truthful intentions. This effect was examined in two experiments that manipulated ecological validity between a controlled, standardized prerecorded video presentation of questions and questions presented by a live interviewer.

EXPERIMENT 1

METHODS

Participants

Participants ($N = 54$) were recruited through advertisements in a local newspaper and through an online classified advertising site. A total of 25 (9 female/16 male; average age = 27.76, $SD = 8.83$) participated in the false intent condition and 29 (12 female/17 male; average age = 28.05, $SD = 9.05$) in the truthful intent condition. The experimental design and data collection procedures were reviewed and approved by the Montana State University Human Subjects Committee and informed consent was obtained from all subjects.

Apparatus

Pupil diameter, blink, and eye movement data were collected using a Smart Eye Pro version 5.4 remote eye tracker. The system has a 60 Hz sampling frequency and is capable of achieving pupil measurement accuracy to .01 mm. Voice responses were collected using a Cedrus SV-1 Voice Key.

Pre-recorded auditory instructions and questioning information were presented as sound files using a TDT System 3 Psychoacoustic Workstation through Altec Lansing VS2120 amplified speakers. Video was presented on a LaCie 324 LCD monitor, a 24-inch widescreen display located 60 cm from the participant. Data from all sources were time-stamped and synchronized through a Cambridge Research Systems ViSaGe, a multiple I/O interface, timing and visual stimulus generation device.

Procedure

Participants responding to the recruitment advertisements were directed to call a local phone number and were presented with a voicemail message instructing them to leave a name and number where they could be contacted.

Four offices in three local buildings were used. Participants first reported to an intake office in a facility that houses professional businesses where they completed an Institutional Review Board (IRB) consent form and were provided directions to the location of another building, within walking distance, where the participant was to receive further instructions.

The test site was located in a building approximately $\frac{1}{2}$ block away from the intake office. Participants walked independently to the building and entered the instruction room. Each participant was randomly assigned to a false intent or truthful intent condition. In both conditions, participants listened to taped instructions through headphones and were told that in addition to the \$25 USD payment for participation they could earn an additional \$25 USD for successfully passing a credibility assessment test. Participants in the false intent condition heard instructions in which they were to commit a mock crime by taking a "fuse lighter" from a downstairs office in the building and providing it to a "contact" after completing a credibility assessment test at another location. Documentation taken from Pickett (1999) was provided illustrating the fuse-lighter as was a photograph of the fictitious "contact." In the truthful intent condition, participants heard instructions in which they were to remove a note from the door of a downstairs office and were not provided with the supplementary materials.

In both conditions, participants exited the instruction room, walked around the block, and entered the building through a side door. They then proceeded downstairs to a basement office. Those in the truthful intent condition simply removed a sticker containing numbers from the door. Those in the false intent condition were required to enter the office and find and remove the fuse lighter. The office containing the fuse-lighter was furnished to resemble a working facility.

In both conditions, the participants exited the building through a third door and proceeded approximately 2 blocks to our laboratory to take a credibility assessment examination.

Details of the examination and data collection are provided in the next section.

DATA COLLECTION

After arriving at the laboratory, both false intent and truthful intent participants were questioned about the event using a Relevant-Relevant comparison test format (RRT). Based on a paradigm developed by Kircher and Raskin (2006), the RRT consists of two sets of relevant questions as well as neutral questions (N). One set of relevant questions addressed matters related to plans to transport illegal drugs for delivery to another person after the credibility assessment test (Drug Intent) while the other set of relevant questions was related to plans to deliver a fuse lighter to a contact after the credibility assessment test (Explosives Intent). Both sets of relevant planned actions were to occur after the questioning and the questions asked about future actions, but only the Explosive Intent was carried out by those in the false intent condition. **Table 1** provides a list of the questions employed and question durations in milliseconds.

In contrast to the Comparison Question Test (CQT) format (Abrams, 1989), in which responses to relevant questions are compared to responses regarding general undesirable acts (e.g., “Before the age of 18, did you ever take something that didn’t belong to you?”), the RRT compares responses to two types of relevant questions—one directed toward the area of interest and the other toward an equally relevant area but one in which the examinee has not engaged. In both test formats, neutral questions are interspersed between the questions of interest to serve as buffers and permit the examinee’s physiological reactions to return to baseline levels. In the RRT, those in the truthful intent condition should have similar responses to both of the relevant question

types, while those in the false intent condition should have different reactions to the explosives intent questions compared to the drug intent questions.

In both truthful intent and false intent conditions, participants stood facing a video monitor and questions were presented aurally through recordings over speakers simultaneously with a prerecorded video of an interviewer asking the questions. Ocular-based parameters were measured in an ambient illumination-controlled room (15.1 lux above). Participants responded “yes” or “no” verbally into a microphone and response time was recorded, but not used in the current analyses due to poor reliability of the data collection device. Question start and question end was marked by a time stamp in the video synchronized with the ocular data. Each question was followed by a 15 s interval before the next question was presented.

After completing the data collection process, participants were debriefed. Those in the false intent condition were told that they did not need to meet a contact, the experiment was completed, and they were asked to return the fuse lighter. Participants in both conditions were told that they would receive the \$25 USD bonus. They were then paid and thanked for their participation.

DATA PROCESSING

For each participant, ocular data were time-stamped and synchronized with the video and audio presentation of the questions. Blinks were identified from the eye tracker data as intervals of 60–1000 ms where the pupil diameter was equal to zero. Three measures were calculated from these data. Blink Count Difference was determined as the number of blinks in the period from the end of a question to 10 s after question end minus the number of blinks during the question presentation. This serves as a measure of blink suppression during question presentation. The use of the 10 s time period was suggested by Stern (pers. Commun., November 20, 2008) based on his experience and was verified through pilot testing that examined time intervals from 5 to 20 seconds ($N = 8$). The average duration of all questions was 2912.92 ms, while the average durations for the Drug Intent, Explosives Intent, and Neutral questions were 2730 ms, 3000 ms, and 2960.8 ms, respectively. Number of blinks was the total number of blinks during the question and the 10 s period following the question end. Maximum blink duration was the length in milliseconds of the longest blink time during the analysis period for each question.

RESULTS

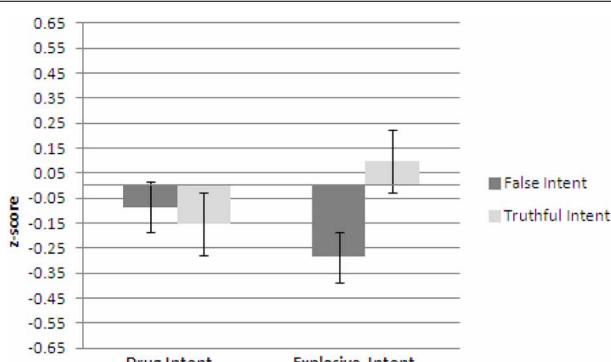
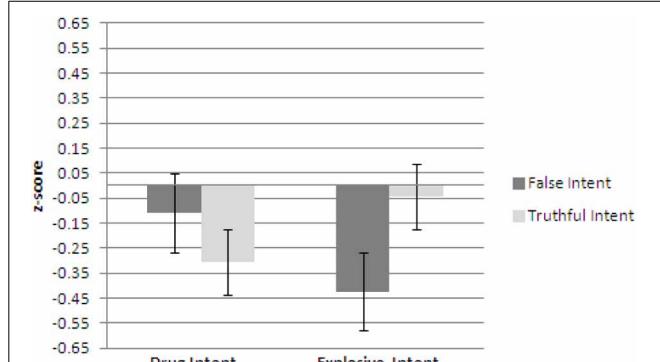
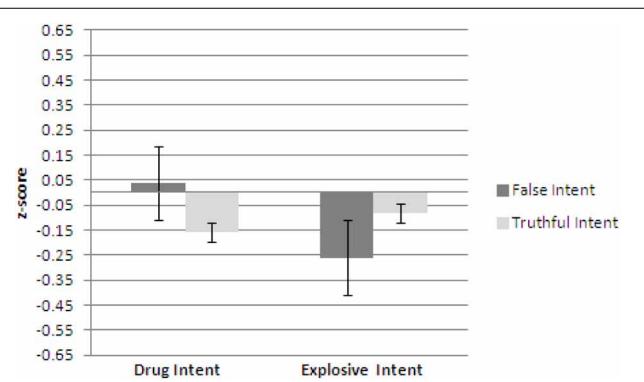
All participants verbally responded “yes” or “no” to all questions and none of the responses were eliminated from analysis. **Table 2** presents the raw means and standard deviations of the blink count difference, number of blinks, and maximum blink duration for Drug Intent, Explosives Intent, and Neutral questions for participants in both the false intent and truthful intent conditions. The ocular-based data were normalized by calculating z -scores and submitted to a repeated measures multivariate analysis of variance (RMANOVA). For Drug Intent vs. Explosives Intent questions, there were significant within-subject multivariate effects for Relevance \times Intent condition, $F_{(3, 50)} = 3.908$, $p = 0.014$, $\eta^2_p = 0.190$. There were significant within-subjects

Table 1 | Test questions and question durations.

Do you live in Bozeman or a surrounding community? (Neutral)	3005 ms
Do you intend to transport illegal drugs today? (Drug Intent)	2610 ms
Were you born between nineteen twenty and nineteen ninety? (Neutral)	3020 ms
Do you plan to help cause a disturbance today? (Explosives Intent)	2420 ms
Did you attend elementary school in the United States? (Neutral)	2890 ms
Are you attempting to transport part of an explosive device? (Explosives Intent)	3520 ms
Is today's date Sunday, February eighth two thousand and nine? (Neutral)	3610 ms
Are you attempting to smuggle illegal drugs today? (Drug Intent)	2670 ms
Are you in Bozeman, Montana at this moment? (Neutral)	2280 ms
Do you plan to provide a fuse lighter to someone today? (Explosives Intent)	3060 ms
Did you use public transportation to arrive here today? (Neutral)	2960 ms
Do you plan to provide illegal drugs to another person? (Drug Intent)	2910 ms

Table 2 | Raw means and standard deviations for blink count difference, number of blinks, and maximum blink duration (ms) by question type—Experiment 1.

Measure	Drug intent		Explosives intent		Neutral	
	Mean	SD	Mean	SD	Mean	SD
BLINK COUNT DIFFERENCE						
False intent	0.183	0.182	0.134	0.193	0.244	0.172
Truthful intent	0.232	0.205	0.279	0.224	0.266	0.204
NUMBER OF BLINKS						
False intent	5.107	3.661	4.467	2.693	5.882	3.319
Truthful intent	5.805	3.999	6.368	4.438	6.823	4.303
MAXIMUM BLINK DURATION						
False intent	210.000	149.505	177.760	109.814	221.700	136.474
Truthful intent	201.741	122.883	212.126	119.379	231.398	123.879

**FIGURE 1 | Experiment 1—Blink count difference as the number of blinks in the period from the end of a question to 10 s after question end minus the number of blinks during the question presentation for false intent and truthful Intent participants.** Error bars show Standard Error.**FIGURE 2 | Experiment 1—Number of blinks for false intent and truthful intent participants.** Error bars show Standard Error.**FIGURE 3 | Experiment 1—Maximum blink duration for false intent and truthful intent participants.** Error bars show Standard Error.

effects of Blink Count Difference, $F_{(1, 52)} = 6.213$, $p = 0.016$, $\eta_p^2 = 0.107$, and Number of Blinks $F_{(1, 52)} = 7.096$, $p = 0.010$, $\eta_p^2 = 0.120$. Maximum Blink Duration approached significance $F_{(1, 52)} = 3.526$, $p = 0.066$, $\eta_p^2 = 0.064$.

Figures 1–3 show plots of normalized Blink Count Difference, Number of Blinks, and Maximum Blink Duration, respectively, for Drug Intent and Explosives Intent questions for participants in both the false intent and truthful intent conditions. Participants in the false intent condition showed a lower blink count difference, fewer numbers of blinks, and shorter but not significantly different maximum blink duration for the Explosives Intent vs. the Drug Intent questions.

In order to examine the relationship between the neutral and relevant questions, the normalized data for the Neutral, Drug Intent, and Explosives Intent questions were submitted to a RMANOVA. There were significant within-subject multivariate effects for Relevance \times Intent condition, $F_{(6, 47)} = 2.585$, $p = 0.030$, $\eta_p^2 = 0.248$. There were significant within-subjects effects of Blink Count Difference, $F_{(2, 104)} = 5.328$, $p = 0.006$, $\eta_p^2 = 0.093$, and Number of Blinks $F_{(2, 104)} = 4.540$, $p = 0.014$, $\eta_p^2 =$

0.080. Maximum Blink Duration was not significant $F_{(2, 104)} = 1.846$, $p = 0.163$.

To examine the effect of age and gender, the data were analyzed in the same manner as above but with age and gender as between subject variables. For Drug Intent vs. Explosives Intent questions, there were no significant effects for Gender by Age, $F_{(21, 45)} =$

$1.140, p = 0.346$, Age by Intent, $F_{(12, 45)} = 1.678, p = 0.104$, or Gender by Intent, $F_{(3, 13)} = 1.727, p = 0.211$.

Tables 3, 4 show the results of a discriminant analysis using all data and a leave-one-out procedure, respectively, to investigate how well the three blink parameters successfully classify false intent and truthful intent individuals. Using all the data, 72.4% of the truthful intent participants and 64.0% of those with false intent were correctly classified, with an overall correct classification rate of 68.5%. Results of the leave-one-out analysis found 72.4% of the truthful intent participants and 60.0% of those with false intent correctly classified, with an overall correct classification rate of 67.6%.

DISCUSSION

Participants with false intent showed a significantly lower blink count difference and lower number of blinks for the Explosive Intent questions as compared to the Drug Intent questions, with the difference in maximum blink duration approaching significance, as compared to participants with truthful intent. There were no significant differences due to age or gender.

EXPERIMENT 2

METHODS

Participants

Participants ($N = 57$) were recruited through the same avenues used in Experiment 1. A total of 29 (11 female/18 male; average age = 26.06, $SD = 9.77$) participated in the false intent condition and 28 (9 female/19 male; average age = 33.40, $SD = 13.18$) participated in the truthful intent condition. The experimental design and data collection procedures were reviewed and approved by the Montana State University Human Subjects Committee and informed consent was obtained from all subjects.

Apparatus

The same equipment and set up used in Experiment 1 was employed with the following changes. The video monitor was replaced with a podium located in front of the participant and a live interviewer located 60 cm from the participant. A Mimo IMO touch screen input display was used to present the question text

Table 3 | Discriminant analysis classification results—Experiment 1.

Actual	Predicted		Overall (%)
	Truthful intent (%)	False intent (%)	
Truthful intent	72.4	27.6	
False intent	36.0	64.0	68.5

Table 4 | Discriminant analysis leave-one-out classification results—Experiment 1.

Actual	Predicted		Overall (%)
	Truthful intent (%)	False intent (%)	
Truthful intent	72.4	27.6	
False intent	40.0	60.0	67.6

to the interviewer. A push button was used by the interviewer to time stamp the question start and question end and synchronize with the ocular data. The voice key was not used.

Procedure

The procedure and instructions were the same as in Experiment 1.

DATA COLLECTION

The data collection process was similar to that in Experiment 1. After arriving at the laboratory, participants stood before the podium and a research assistant performed a short calibration procedure for the eye tracker that involved having the participant fixate on five spots located in front of them. When calibration was complete, the interviewer entered and sat behind the podium. Questions were presented to the interviewer on a Mimo IMO touch screen display, which also indicated the inter-question interval time before the displaying the next question. Each question was read aloud. When the interviewer was ready, he pushed a hand-held button to code question start time into the data stream. After reading the question, question end time was also marked by the interviewer pressing a button. When the 15 s inter-trial interval had passed, the display presented the next question to the interviewer.

DATA PROCESSING

Data processing was identical to Experiment 1.

RESULTS

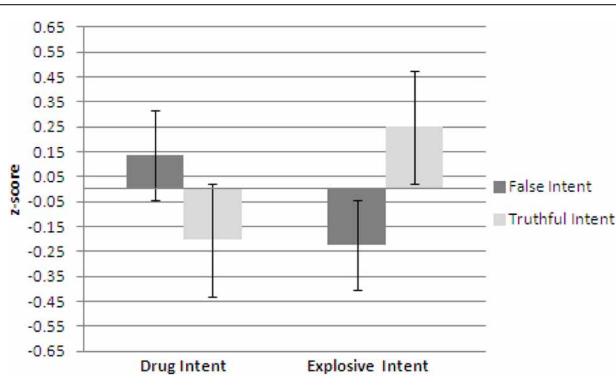
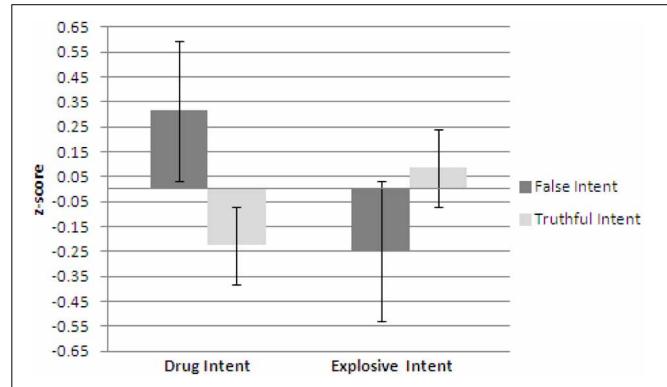
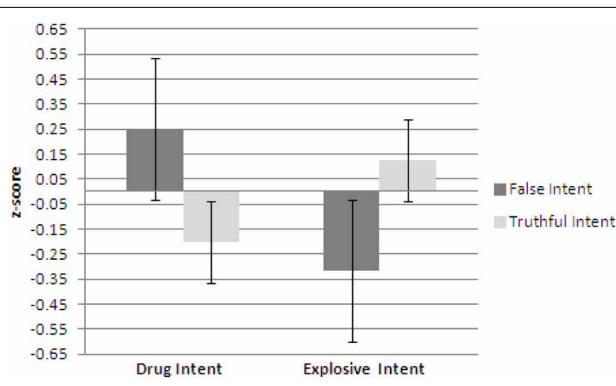
All participants verbally responded “yes” or “no” to all questions and none of the responses were eliminated from analysis. **Table 5** presents the raw means and standard deviations of the blink count difference, number of blinks, and maximum blink duration for Drug Intent, Explosives Intent, and Neutral questions for participants in both the false intent and truthful intent conditions. The ocular-based data were normalized by calculating z-scores and submitted to a repeated measures multivariate analysis of variance (RMANOVA). For Drug Intent vs. Explosives Intent questions, there were significant within-subject multivariate effects for Relevance \times Intent Condition, $F_{(3, 53)} = 10.362, p = 0.000, \eta_p^2 = 0.370$. There were significant within-subjects effects of Blink Count Difference, $F_{(1, 55)} = 12.983, p = 0.001, \eta_p^2 = 0.191$, Number of Blanks $F_{(1, 55)} = 20.156, p = 0.000, \eta_p^2 = 0.268$, and Maximum Blink Duration $(1, 55) = 18.179, p = 0.000, \eta_p^2 = 0.248$.

Figures 4–6 show plots of normalized Blink Count Difference, Number of Blinks, and Maximum Blink Duration, respectively, for Drug Intent and Explosives Intent questions for participants in both the false intent and truthful intent conditions. Participants in the false intent condition showed a lower blink count difference, fewer numbers of blinks, and shorter maximum blink duration for the Explosives Intent vs. the Drug Intent questions when compared to participants in the truthful intent condition.

In order to examine the relationship between the neutral and relevant questions, the normalized data for the Neutral, Drug Intent, and Explosives Intent questions were submitted to a RMANOVA. There were significant within-subject multivariate effects for Relevance \times Intent condition, $F_{(6, 53)} = 5.696,$

Table 5 | Raw means and standard deviations for blink count difference, number of blinks, and maximum blink duration (ms) by question type—Experiment 2.

Measure	Drug intent		Explosives intent		Neutral	
	Mean	SD	Mean	SD	Mean	SD
BLINK COUNT DIFFERENCE						
False intent	0.251	0.276	0.302	0.258	0.288	0.269
Truthful intent	0.321	0.224	0.229	0.221	0.256	0.192
NUMBER OF BLINKS						
False intent	4.219	3.785	5.130	4.349	4.850	4.520
Truthful intent	4.633	3.201	3.933	3.460	4.634	3.661
MAXIMUM BLINK DURATION						
False intent	239.958	113.008	285.505	125.791	245.495	97.888
Truthful intent	217.567	87.608	194.067	71.835	234.250	79.868

**FIGURE 4 | Experiment 2—Blink count difference as the number of blinks in the period from the end of a question to 10 s after question end minus the number of blinks during the question presentation for false intent and truthful intent participants.** Error bars show Standard Error.**FIGURE 6 | Experiment 2—Maximum blink duration for false intent and truthful intent participants.** Error bars show Standard Error.**FIGURE 5 | Experiment 2—Number of blinks for false intent and truthful intent participants.** Error bars show Standard Error.

$p = 0.000$, $\eta_p^2 = 0.406$. There were significant within-subjects effects of Blink Count Difference, $F_{(2, 110)} = 7.448$, $p = 0.001$, $\eta_p^2 = 0.119$, and Number of Blinks $F_{(2, 110)} = 9.186$, $p = 0.000$,

$\eta_p^2 = 0.143$. Maximum Blink Duration was significant $F_{(2, 110)} = 10.13$, $p = 0.000$, $\eta_p^2 = 0.156$.

To examine the effect of age and gender, the data were analyzed in the same manner as above but with age and gender as between subject variables. For Drug Intent vs. Explosives Intent questions, there were no significant effects for Gender by Age, $F_{(18, 51)} = 0.763$, $p = 0.731$, Age by Intent, $F_{(15, 51)} = 1.012$, $p = 0.459$, or Gender by Intent, $F_{(3, 15)} = 0.131$, $p = 0.940$.

Tables 6, 7 show the results of a discriminant analysis using all data and a leave-one-out procedure, respectively, to investigate how well the three blink parameters successfully classify false intent and truthful intent individuals. Using all the data, 72.0% of the truthful intent participants and 78.1% of those with false intent were correctly classified, with an overall correct classification rate of 75.4%. Results of the leave-one-out analysis found 68.0% of the truthful intent participants and 78.1% of those with false intent correctly classified, with an overall correct classification rate of 73.7%.

DISCUSSION

Similar to Experiment 1, participants with false intent showed a significantly lower blink count difference, lower number of blinks, and lower maximum blink duration for the Explosives Intent

Table 6 | Discriminant analysis classification results—Experiment 2.

Actual	Predicted		Overall (%)
	Truthful intent (%)	False intent (%)	
Truthful intent	72.0	28.0	
False intent	21.9	78.1	75.4

Table 7 | Discriminant analysis leave-one-out classification results—Experiment 2.

Actual	Predicted		Overall (%)
	Truthful intent (%)	False intent (%)	
Truthful intent	68.0	32.0	
False intent	21.9	78.1	73.7

questions as compared to Drug Intent questions. However, there were differences in the responses to the Drug Intent questions as compared to those found in Experiment 1. One possible explanation for these differences could be related to the difference of being questioned by a videotaped interviewer and interacting with a live interviewer. While the questions and procedures were the same, participants in Experiment 1 did not have direct interactions with a live person. Riby et al. (2012) found a difference in skin conductance level (SCL) and increased arousal to live faces compared to video-mediated faces. This increase in arousal as a result of interacting with a live interviewer could contribute to the differences found in blink parameters between the two experiments.

GENERAL DISCUSSION

The goal of this effort was to determine if variations in blink measures could differentiate between those with false intent and truthful intent individuals. Two experiments with differing degrees of ecological validity were conducted using either a pre-recorded interviewer presented on a computer monitor or a live interviewer.

To date, the majority of research on determining false intent has employed verbal or non-verbal cues. Vrij and colleagues (2011a); Vrij and his colleagues (2011b) successfully detected false intent employing a structured interview and analysis of the resulting transcripts, as well as based on speech cues and participant willingness to be photographed. Similarly, Clemens et al. (2011) demonstrated that strategic interviewing elicits reliable cues to detecting false intent.

Only one study on detecting false intent has examined the use of physiological cues. Aikins et al. (2010) detected false intent by examining respiratory sinus arrhythmia (RSA)—an indicator of autonomic response—showing that individuals with false intent displayed decreased RSA compared to individuals with true intentions.

In both experiments reported here, it was found that for questions relevant to the harmful act to be committed, those with false intent showed a lower blink count difference, fewer numbers of blinks, and shorter maximum blink duration for questions related

to their intent compared to questions related to another act for which they had no intent. These findings are consistent with previous findings in the literature that used blink measures to determine deception regarding past activities. While these analyses focused on factors related to blink counts and time, it would be possible to examine additional measures such as blink waveforms (Stern et al., 1984).

Two factors could contribute to the findings of differences in blink parameters of those with false intent: cognitive load and arousal. Both theories of cognitive load (e.g., Fogarty and Stern, 1989; Fukuda et al., 2005; Irwin and Thomas, 2010) as well as arousal-based theories (e.g., Stern, 1992) have been implicated in the context of deception detection.

As noted by Vrij et al. (2011a,b), differences between liars and truth tellers in both deception about past activities and future intentions are potentially affected by the increased cognitive load brought on in the untruthful situation. The effect of cognitive load on deception detection has been documented (e.g., Vrij et al., 2006; Leal et al., 2008). The effect of cognitive load on intention can be examined from two perspectives of memory about future events: episodic future thought (EFT) and prospective memory.

Granhaug and Knieps (2011) have proposed that EFT is the central mental process involved in forming an intention and have used this framework to propose that the activation of a mental image in the pre-experiencing of an intention will be stronger for a true vs. a false intention. The current study did not explicitly test this construct so no comment can be made on its applicability based on the available data.

Prospective memory is defined as the process that permits remembering to engage in an intended action at some particular point in the future (Kvavilashvili and Ellis, 1996). Kliegel et al. (2000) describe prospective memory as consisting of three processes: developing a plan, remembering the plan, and remembering to execute the plan at some future time. In the current study, those with false intent had a plan to meet a contact and deliver the fuse lighter after taking the credibility assessment test, so were engaged in the first two processes but were stopped before the opportunity to begin the third process.

No studies were found that examined physiological measures of prospective memory with the exception of Hartwig et al. (2013), who examined gaze behavior to determine different approaches employed in solving prospective memory tasks. Hartwig et al. (2013) used the skewness of Voronoi cell distributions of fixation densities to quantify viewing strategies (Velichkovsky, 1999). Over et al. (2006) demonstrated that different visual tasks can be differentiated by skewness differences in the Voronoi cell sizes, and that tasks involving the same behavior would have similar skewness. Hartwig et al. (2013) found that when a prospective memory task was missed, participants exhibited gaze behavior similar to that seen in free viewing, including differential attention to details over only a few areas of interest. This viewing behavior resulted in a few large Voronoi cells and multiple small cells. If the prospective memory task was solved successfully, gaze behavior took on characteristics somewhat between those seen in free viewing and those seen in visual search, which was characterized by a large number of fixations across the entire display and many small Voronoi cells. These

findings seem to imply that different approaches and levels of cognitive effort are involved in carrying out a prospective memory task and that the different processes are reflected in the ocular measures.

The effect of arousal on eye blink behavior has been investigated by Tanaka (1999) who examined the changes in blink rate, amplitude, and duration as a function of arousal level and found differences between a high arousal vigilance task and a low arousal counting task. Thonney et al. (2005) used experimentally aroused emotions of remorse and guilt and examined the effect of eye blinking and electrodermal response on Guilty Knowledge Test accuracy. They found that eye blinking was diagnostic for only the treatment group but not as accurate as electrodermal measures.

These findings have implications for further research on both blink measures and determination of false intent. One issue involves the contribution of cognitive workload and arousal to the changes in blink behavior. Both have been shown to affect blink rate, and while recent research has suggested that such findings are due primarily to cognitive load, neither this work nor previous efforts (Fukuda, 2001; Leal and Vrij, 2008, 2010) have explicitly addressed this question. Thus, no definitive conclusions may be drawn regarding the specific contributions of arousal and cognitive load to the findings.

Another factor of interest is distinguishing false intent about future actions (i.e., plan to deliver object to contact) from lying about past actions (i.e., the mock crime). In a standard polygraph examination employing, for example, the Comparison Question Technique, participants are asked a series of questions, a subset of which are relevant to the past act, such as the mock crime. Here, participants were asked questions relevant to their upcoming actions—delivering an explosive device to a contact that would use it to cause a disturbance—to be completed in the future

REFERENCES

- Abrams, S. (1989). *The Complete Polygraph Handbook*. Lexington, MA: Lexington Books.
- Aikins, D. E., Martin, D. J., and Morgan, C. A. I. (2010). Decreased respiratory sinus arrhythmia in individuals with deceptive intent. *Psychophysiology* 47, 633–636. doi: 10.1111/j.1469-8986.2010.00976.x
- Clemens, F., Granhag, P. A., and Strömwall, L. A. (2011). Eliciting cues to false intent: a new application of strategic interviewing. *Law Hum. Behav.* 35, 512–522. doi: 10.1007/s10979-010-9258-9
- Cook, A. E., Hacker, D. J., Webb, A., Osher, D., Kristjansson, S. D., Woltz, D. J., et al. (2012). Lyin' eyes: ocular-motor measures of reading reveal deception. *J. Exp. Psychol. Appl.* 18, 201–213. doi: 10.1037/a0028307
- Cutrow, R. J., Parks, A., Lucas, N., and Thomas, K. (1972). The objective use of multiple physiological indices in the detection of deception. *Psychophysiology* 9, 578–587. doi: 10.1111/j.1469-8986.1972.tb00767.x
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., and Cooper, H. (2003). Cues to deception. *Psychol. Bull.* 129, 74–118. doi: 10.1037/0033-2909.129.1.74
- Fogarty, C., and Stern, J. A. (1989). Eye movements and blinks: their relationship to higher cognitive processes. *Int. J. Psychophysiol.* 8, 35–42. doi: 10.1016/0167-8760(89)90017-2
- Fukuda, K. (2001). Eye blinks: new indices for the detection of deception. *Int. J. Psychophysiol.* 40, 239–245. doi: 10.1016/S0167-8760(00)00192-6
- Fukuda, K., Stern, J. A., Brown, T. B., and Russo, M. B. (2005). Cognition, blinks, eye-movements, and pupillary movements during performance of a running memory task. *Aviat. Space. Environ. Med.* 76, C75–C85.
- Granhag, P. A., and Knieps, M. (2011). Episodic future thought: illuminating the trademarks of forming true and false intentions. *Appl. Cogn. Psychol.* 25, 274–280. doi: 10.1002/acp.1674
- Hartwig, M., Schnitzspahn, K. M., Kliegel, M., Velichkovsky, B. M., and Helmert, J. R. (2013). I see you remembering: what eye movements can reveal about process characteristics of prospective memory. *Int. J. Psychophysiol.* 88, 193–199. doi: 10.1016/j.ijpsycho.2013.03.020
- Irwin, D. E., and Thomas, L. E. (2010). “Eyeblinks and cognition,” in *Tutorials in Visual Cognition*, ed V. Coltheart (New York, NY: Psychology Press), 121–141.
- Kircher, J. C., and Raskin, D. C. (2006). *Computerized Screening System (CSS). Proposal*. Salt Lake City: Scientific Assessment Technologies, Inc.
- Kliegel, M., McDaniel, M. A., and Einstein, G. O. (2000). Plan formation, retention, and execution in prospective memory: a new approach and age-related effects. *Mem. Cogn.* 28, 1041–1049. doi: 10.3758/BF03209352
- Kvavilashvili, L., and Ellis, J. (1996). “Varieties of intention: some distinctions and classifications,” in *Prospective Memory Theory and Applications*, eds M. Brandimonte, G. O. Einstein, and M. A. McDaniel (Mahwah, NJ: Lawrence Erlbaum Associates), 23–51.
- Leal, S., and Vrij, A. (2008). Blinking during and after lying. *J. Nonverbal Behav.* 32, 187–194. doi: 10.1007/s10919-008-0051-0
- Leal, S., and Vrij, A. (2010). The occurrence of eye blinks during a guilty knowledge test. *Psychol. Crime Law* 16, 349–357. doi: 10.1080/10683160902776843
- Leal, S., Vrij, A., Fisher, R. P., and van Hooff, H. (2008). The after the questioning. None of the questions referred to activities previously performed.
- Vrij et al. (2011b) explicitly compared differences in verbal cues and detection accuracy between individuals lying about past activities and future intentions, and found a higher accuracy rate in determining false intent, although this may have been attributed to differences in how the observers scored the transcripts. One way to examine this effect explicitly using the current approach would be to present truthful intent participants with information about the mock crime without actual participation in it or possessing information regarding the future actions and compare the responses with those who committed the mock crime.
- It might also be possible to add a third condition in which participants complete the mock crime in terms of obtaining the fuse lighter but are not told that it is to be delivered to a contact. Martin et al. (2011) found physiological differences in several parameters, including pupil diameter, between individuals who planned to participate in a malicious event after passing through screening, without first committing a mock crime or attempting to smuggle an illegal device, and innocent participants. Comparisons of the blink behaviors of this group with the group that intends to meet a contact could provide a direct comparison between lying about past actions and false intent. The findings presented here serve as an initial step toward determining the ability of using physiological measures to determine false intent as opposed to lying about previous acts.

ACKNOWLEDGMENTS

This research was supported by the US Department of Homeland Security. The author wishes to thank Tanner Keil, Jennifer McMillan, and Pamela Westphal for their assistance with data collection and analysis, as well as the reviewers for their thoughtful and directed feedback.

- time of the crime: cognitively induced tonic arousal suppression when lying in a free recall context. *Acta Psychol.* 129, 1–7. doi: 10.1016/j.actpsy.2008.03.015
- Marchak, F. M., Keil, T. L., McMillan, J. E., and Westphal, P. S. (2011). Ocular-based measures of malintent. *Psychophysiology* 48, S10.
- Martin, D. J., Martin, J. Z., and Coskren, W. D. (2007). *A Theory of Malintent and Survey of Sensing Abilities for the Future Attribute Screening Technologies (FAST) Program*, Washington, DC: U.S. Department of Homeland Security.
- Martin, D. J., Martin, J. Z., Webb, A. K., and Horgan, A. J. (2011). Malintent: theory and detection in primary screening. *Psychophysiology* 48(Suppl. 1), S10.
- Masip, J., Garrido, E., and Herrero, C. (2004). Defining deception. *An. Psicol.* 20, 147–171.
- Meixner, J. B., and Rosenfeld, J. P. (2011). A mock terrorism application of the P300-based concealed information test. *Psychophysiology* 48, 149–154. doi: 10.1111/j.1469-8986.2010.01050.x
- Over, E. A. B., Hooge, I. T. C., and Erkelens, C. J. (2006). A quantitative measure for the uniformity of fixation density: the Voronoi method. *Behav. Res. Methods* 38, 251–261. doi: 10.3758/BF03192777
- Pickett, M. (1999). *Explosives Identification Guide*. Albany, NY: Delmar Publishing.
- Riby, D. M., Whittle, L., and Doherty-Sneddon, G. (2012). Physiological reactivity to faces via live and video-mediated communication in typical and atypical development. *J. Clin. Exp. Neuropsychol.* 34, 385–395. doi: 10.1080/13803395.2011.645019
- Stern, J. A. (1992). “The eye blink: affective and cognitive influences,” in *Anxiety: Recent Developments in Cognitive, Psychophysiological, and Health Research*, eds D. G. Forgays, T. Sosnowski and K. Wrzeniewski (Washington, DC: Hemisphere Publishing Corporation), 109–128.
- Stern, J. A., Walrath, L. C., and Goldstein, R. (1984). The endogenous eyeblink. *Psychophysiology* 21, 22–33. doi: 10.1111/j.1469-8986.1984.tb02312.x
- Tanaka, Y. (1999). Arousal level and blink activity. *Jpn. J. Psychol.* 70, 1–8. doi: 10.4992/jjpsy.70.1
- Thonney, J., Kanachi, M., Sasaki, H., and Hatayama, T. (2005). Eye blinking as a lie-detection index in an emotionally arousing context. *Tohoku Psychol. Folia* 64, 58–67.
- Trovillo, P. V. (1939a). A history of lie detection. *J. Crim. Law Criminol.* 29, 848–881. doi: 10.2307/1136489
- Trovillo, P. V. (1939b). A history of lie detection (Concluded). *J. Crim. Law Criminol.* 30, 104–119. doi: 10.2307/1136392
- Velichkovsky, B. M. (1999). “From levels of processing to stratification of cognition: converging evidence from three domains of research,” in *Stratification in Cognition and Consciousness*, eds B. H. Challis and B. M. Velichkovsky (Amsterdam/Philadelphia: John Benjamins), 203–266.
- Vrij, A. (2008). *Detecting Lies and Deceit: The Psychology of Lying and the Implications for Professional Practice*, 2nd Edn. Chichester: Wiley.
- Vrij, A., Fisher, R., Mann, S., and Leal, S. (2006). Detecting deception by manipulating cognitive load. *Trends Cogn. Sci.* 10, 141–142. doi: 10.1016/j.tics.2006.02.003
- Vrij, A., Granhag, P. A., Mann, S. A., and Leal, S. (2011a). Lying about flying: the first experiment to detect false intent. *Psychol. Crime Law* 17, 611–620. doi: 10.1080/10683160903418213
- Vrij, A., Leal, S., Mann, S. A., and Granhag, P. A. (2011b). A comparison between lying about intentions and past activities: verbal cues and detection accuracy. *Appl. Cogn. Psychol.* 25, 212–218. doi: 10.1002/acp.1665
- Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Received:** 15 September 2012; **accepted:** 23 September 2013; **published online:** 11 October 2013.
- Citation:** Marchak FM (2013) Detecting false intent using eye blink measures. *Front. Psychol.* 4:736. doi: 10.3389/fpsyg.2013.00736
- This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology.*
- Copyright © 2013 Marchak. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*



The current and future status of the concealed information test for field use

Izumi Matsuda^{1*}, Hiroshi Nittono² and John J. B. Allen³

¹ National Research Institute of Police Science, Chiba, Japan

² Graduate School of Integrated Arts and Sciences, Hiroshima University, Higashi-Hiroshima, Japan

³ Department of Psychology, University of Arizona, Tucson, AZ, USA

Edited by:

Wolfgang Ambach, Institute for Frontier Areas of Psychology and Mental Health, Germany

Reviewed by:

Frank M. Marchak, Veridical Research and Design Corporation, USA
Gershon Ben-Shakhar, The Hebrew University of Jerusalem, Israel
Donald Krapohl, National Center for Credibility Assessment, USA

*Correspondence:

Izumi Matsuda, National Research Institute of Police Science, 6-3-1 Kashiwanoha, Kashiwa, Chiba 227-0882, Japan.
e-mail: izumi@nrips.go.jp

OVERVIEW

The Concealed Information Test (CIT) assesses an examinee's crime-relevant memory on the basis of differences in physiological responses between crime-relevant and crime-irrelevant items (Lykken, 1959). Although many studies have supported the validity of the CIT, it has not been widely used in field situations. There appear two reasons for its unpopularity. First, some examiners appear to prefer an alternative method termed the Control Question Test (CQT), even though the validity of the CQT has been seriously questioned (Ben-Shakhar, 2002). Second, the CIT is believed to be difficult to apply in non-laboratory field settings. In Japan, however, the autonomic-based CIT is routinely applied successfully in criminal investigations. Even so, CIT results have not been widely influential in court settings.

In this paper, we review the current status of the CIT in the field and laboratory studies, with the goal of outlining steps that can contribute to an increased probative value of the CIT in court. First, we review how Japanese examiners have tried to overcome the difficulties of the CIT for field application. Second, we review statistical methods that can be used to support judgments in field CIT applications, and investigate new measures that can be added to the current CIT implementations.

Throughout this paper, we will emphasize viewpoints relevant to field applications. In the field, an examinee is often not willing to take the test and does not comply with instructions. Therefore, in Japan, a classic autonomic-based CIT has been used, which simply consists of one crime-relevant item and several crime-irrelevant items and does not require an overt behavioral response. This paper will focus on how this existing field CIT can be expanded, but it will not review other alternative approaches. For example, other memory detection or lie detection tests that are still in the

The Concealed Information Test (CIT) is a psychophysiological technique for examining whether a person has knowledge of crime-relevant information. Many laboratory studies have shown that the CIT has good scientific validity. However, the CIT has seldom been used for actual criminal investigations. One successful exception is its use by the Japanese police. In Japan, the CIT has been widely used for criminal investigations, although its probative force in court is not strong. In this paper, we first review the current use of the field CIT in Japan. Then, we discuss two possible approaches to increase its probative force: sophisticated statistical judgment methods and combining new psychophysiological measures with classic autonomic measures. On the basis of these considerations, we propose several suggestions for future practice and research involving the field CIT.

Keywords: concealed information test, field application, probative force, statistical judgment, combination of measures

laboratory stage, such as the autobiographic implicit association test (Sartori et al., 2008), show promise but are outside of the scope of this paper.

CURRENT STATUS OF FIELD CIT

WHAT IS THE CIT?

The CIT, also known as the guilty knowledge test (GKT; Lykken, 1959), is used in criminal investigations to examine whether a person recognizes crime-relevant information that innocent people would not know. In the CIT, an examiner presents several items to an examinee, one of which is a crime-relevant item. The items are selected such that innocent examinees would not be able to distinguish the crime-relevant (critical) item from the crime-irrelevant (non-critical) items. Each item is presented once in a block and this block is repeated several times in different presentation orders. During the CIT, the examiner records physiological responses to the items. In the case that the responses do not differ between the critical and non-critical items, the examiner would infer that the examinee does not recognize the critical item. On the other hand, in the case that the responses differ between the critical and non-critical items, the examiner would infer that the examinee recognizes the critical item. Thus, the CIT can provide important forensic information for the police and the justice system, identifying individuals with key information about the crime. Such individuals may be guilty of committing the crime, or have other useful information about the crime if they were not the perpetrator.

The CIT is considered to have a solid scientific foundation, as many laboratory studies have demonstrated its effectiveness (for a review, see Ben-Shakhar and Elaad, 2003). Although published field data are relatively scarce (Elaad, 1990; Elaad et al., 1992; Hira-

and Furumitsu, 2002; Osugi, 2010), the response pattern of the various physiological measures in field CITs are similar to those observed in laboratory CITs (i.e., skin conductance increase, heart rate decrease, respiration suppression, and finger pulse volume decrease for critical items as compared to non-critical items; Elaad, 1990; Elaad et al., 1992; Osugi, 2010; Verschueren et al., 2011).

POTENTIAL PROBLEMS IN THE FIELD APPLICATION OF THE CIT

To date, the CIT has not been widely used in field settings. This may reflect, in part, the belief that the CIT is difficult to apply in field settings for a variety of reasons (Krapohl, 2011). First, the CIT can produce false positive cases. Critical items that only a guilty person knows are sometimes difficult to find. Some innocent examinees may know the details of the crime through any number of means, including media reports and rumors (i.e., informed innocent examinees; for a review, see Bradley et al., 2011). Other innocent examinees may, via repeated interrogations or repetitions of crime details, come to have false recollections for crime-relevant items (Allen and Mertens, 2008). If these innocent examinees take the CIT, they would show different responses for critical and non-critical items, resulting in false positive outcomes. Second, the CIT is vulnerable to false negative outcomes. If critical items are selected that are not memorable to the perpetrator of the crime, it is unlikely to be recognized, thus producing a false negative outcome. Even if examinees do have crime-relevant memories and recognize the crime-relevant item, physiological differences sometimes might not be observed. For example, although skin conductance is typically measured in the CIT, one study reported that approximately one out of four people were electrodermal non-responders to orienting stimuli (Venables and Mitchell, 1996). Third, some studies have shown that the CIT is vulnerable to physical countermeasures (e.g., pressing the toes against the floor when non-critical items are presented) as well as mental countermeasures (e.g., counting numbers each time a non-critical item appears; for a review, see Ben-Shakhar, 2011). In the next section, we will introduce how Japanese CIT examiners have attempted to overcome these three problems.

CURRENT FIELD USE OF THE CIT IN JAPAN

In spite of the three problems outlined above, the CIT has been officially and systematically used in Japan for the last 50 years. About 100 trained examiners perform about 5,000 CITs per year (Osugi, 2011). All examiners (who are not investigators) belong to a forensic science laboratory of a prefectural police headquarter. The CQT (Reid, 1947) is no longer used. The results of the CIT have been accepted as evidence in court since the 1960s. Although Japan's successful application of the CIT in the field has attracted attention from foreign researchers and examiners, not much has been written about how the potential problems for field use of the CIT have been addressed in Japan. Therefore, potential solutions are reviewed briefly below, and more details are available from Osugi (2011).

Prevention of false positive cases

Japanese CIT examiners make every effort to prevent false positive cases through every step in the process, from pre-exam preparation to the actual administration of the CIT. On a routine basis,

an examiner advises criminal investigators to conduct the CIT at an early stage of the investigation in order to make it less likely that crime-relevant items become known to a wider audience over time. When an examiner is requested to conduct the CIT, he/she first consults with investigators. An examiner also checks media reports related to the crime and to the record of investigation. Furthermore, before conducting each CIT, an examiner presents all the items in the CIT to an examinee, and asks the examinee if there are items that he/she recognizes or feels different from the others. If the examinee points out the crime-relevant item, the examiner would not administer the CIT question about that item.

Prevention of false negative cases

Japanese CIT examiners strive to select critical items that a guilty person should remember. They try to avoid using peripheral features of the crime, and instead use central features as critical items (Carmel et al., 2003; Gamer et al., 2010; Nahari and Ben-Shakhar, 2011). In addition, before each CIT, an examiner explains the meaning of each question item to an examinee, in order that the examinee will understand what the examiners are asking.

However, even when an examinee might recognize a critical item, he/she sometimes may not show a different physiological response between the critical and non-critical items. One of the strategies to avoid this type of false negative case is the simultaneous measurement of multiple validated responses. In Japan, a new polygraph system has been used since 2003, which simultaneously records skin conductance, heart rate, pulse volume, and respiration. These measures are thought to reflect the different aspects of a physiological response. Laboratory studies show that combining these multiple measures could reduce false negative rates while maintaining low false positive rates (e.g., Gamer et al., 2008a).

Counter-countermeasures

To guard against physical countermeasures, an examiner monitors an examinee's behavior and his/her physiological responses carefully during the CIT. When the examiner thinks that the examinee is intentionally applying countermeasures (e.g., frequent body movements, sighs, or sniffing), he or she would instruct the examinee to refrain from such activities (Osugi, 2011). Although specific sensors to detect physical countermeasures have not been applied in Japan yet, it may be useful to introduce, for example, pressure-based sensors incorporated in the test chair and floor pads, which have been used in some other countries.

Previous studies have suggested that mental countermeasures affect skin conductance, but do not affect respiration (Ben-Shakhar and Dolev, 1996; Honts et al., 1996). In Japan, an examiner measures multiple autonomic indices including respiration, which can serve to lessen the chance that countermeasures will change the outcome of the CIT. To measure an examinee's physiological response from various response channels can thus contribute to reducing the effect of unobservable mental countermeasures.

Other attempts

Examiners in Japan also use other procedures to get more accurate and/or informative results. First, examiners always conduct a pretest before asking about crime-relevant information. In the pretest, an examiner asks an examinee to memorize a number

on a card in private and then presents several numbers including the memorized number. The pretest not only helps the examinee to understand the CIT paradigm, but also helps the examiner to know the physiological response pattern of the examinee when he or she recognizes an item. Considering the response pattern, the examiner conducts the subsequent CITs. For example, if the examinee showed high reactivity in skin conductance response in the pretest, the examiner judges the responses of subsequent CITs paying more attention to the skin conductance response.

Second, an examiner sometimes uses a *searching CIT*. The searching CIT is different from the typical CIT in that an examiner does not know which item is crime-relevant in advance. For example, if a weapon has been missing, an examiner can ask an examinee about the place where he/she abandoned a weapon, such as "Was a weapon abandoned in area A, area B, . . . , or area E?" Indeed, the judgment is more difficult for a searching CIT than for a usual CIT with known solutions, because the examiner has to judge not only whether the examinee has recognition but also which item the examinee recognizes. Additionally, in the case that the question items do not cover all possibilities, the finding of no physiological differences between items cannot support an examiner's conclusion "the examinee does not recognize the crime-relevant item;" instead, this finding can only support the conclusion that "the examinee does not recognize any items in this question set." But if an examiner develops an appropriate question set, the searching CIT can suggest potential new crime-relevant information of which even investigators have no knowledge. In the above example, if the responses differ between area A and other areas, the investigators will focus investigation on area A and consequently may find the missing weapon.

Third, in Japan, an examiner only decides on whether an examinee recognizes each crime-relevant item and never integrates the results of multiple CIT questions to judge whether the examinee is guilty or innocent. It is the investigators' task, rather than the examiner's task, to integrate the results across the CIT questions and evaluate the examinee's likelihood of guilt. Some authors, however, have argued that examiners should integrate results across multiple CIT questions in order to obtain more statistically reliable and robust results (Ben-Shakhar and Elaad, 2002). However, Japanese examiners have maintained the approach of only adopting a judgment for each CIT question. One of the justifications for conducting the test in this manner is that it allows the examiner to clarify which items the examinee recognizes and which items the examinee does not. For example, in the case of a theft that was conducted by a group of perpetrators, information indicating whether the examinee knows each crime-relevant item may become a clue to reveal what role he/she played in the crime (e.g., a major culprit or just a lookout). Thus, treating results from each CIT question separately can facilitate investigations of cases involving multiple suspects, and provide details to guide and facilitate the investigators' continuing inquiries for any type of case. Additionally, as described above, Japanese examiners sometimes use searching CITs; in such cases where an examiner does not know with certainty which alternative is the critical item for a given CIT question, it is difficult to integrate CIT results across questions.

Validity of the field CIT in Japan

One article has reported on field CIT datasets using the current polygraph system in Japan. Kobayashi et al. (2009) analyzed the data of 113 CIT questions obtained from 38 examinees (33 men and 5 women, mean age = 36.4, $SD = 12.5$). Subsequent investigations confirmed that all of these examinees recognized the critical items of these CIT questions. For each CIT question, the responses were compared between critical and non-critical items with a *t* test. If the *p* value did not exceed 0.10, the examinee was judged as recognizing the critical item. The correct detection rates were 52.5% for the skin conductance response, 49.5% for heart rate (average in 16–20 s after the item onset), 38.1% for respiration line length (average in 0–15 s), and 26.2% for normalized pulse volume (average in 6–10 s). It should be noted that these values are correct detection rates (i.e., sensitivities) for individual CIT questions using a single measure. Although Kobayashi et al. did not report the data, combining the various physiological measures should increase the overall detection rate. In the actual field CIT, examiners arrive at a conclusion by combining all of the available measures. In addition, to address the specificity of the CIT (i.e., how well each measure correctly indicates non-recognition of critical items when examinees do not have recognition), a larger dataset including both guilty and innocent subjects would be required.

IMPROVING THE PROBATIVE FORCE OF THE CIT IN COURT

Although the CIT has been widely used for criminal investigations and its results have been sometimes accepted as evidence in court in Japan, the CIT results are not considered sufficiently strong that they typically directly affect the outcomes in court. To improve the probative force of the CIT, we believe the following two approaches are most promising.

The first approach is to use statistical methods to interpret the results. In field use of the CIT in Japan, CIT results are mainly derived through the examiners' visual inspections (Osugi, 2011). If the judgment is underpinned by statistical methods, the CIT results would become more convincing for judges. Moreover, such an approach is well-justified in the literature: statistical actuarial judgment has greater reliability and validity than judgments based on visual impressions (Dawes, 1979). In laboratory studies, Lykken's scoring and *z*-score averaging have been commonly used for decision-making (Meijer et al., 2011). Lykken, 1959 scoring is based on the rank of the critical item among all items in descending order of the response values. *Z*-score averaging uses the average standardized response value across blocks and measures (Ben-Shakhar, 1985). Although these two methods are simple and clear, they do have drawbacks. We will review these two methods critically and compare them with other proposed methods below.

The second approach is to add new measures to current field CIT to increase its accuracy. In the current field CIT, heart rate, skin conductance, respiration, and pulse volume are recorded. New measures can be introduced either by improving quantification methods of currently recorded responses or by recording new response channels, such as reaction time, facial responses, activations using functional magnetic resonance imaging (fMRI), and features of the electroencephalogram (EEG) and event-related

potential (ERP). We will review these new measures and evaluate these from the viewpoint of field application.

STATISTICAL EVALUATION METHODS

Here, we review statistical methods that have been used in previous studies. First, we review standard statistical methods such as Lykken's scoring and z-score averaging. We then review five other proposed methods: logistic regression discrimination, latent class discrimination, Bayesian classification, multivariate normal distribution discrimination, and dynamic mixture distribution discrimination. Finally, we outline recommendations for their use.

Standard statistical methods

Lykken's scoring method. This is a traditional discrimination method proposed by Lykken (1959; **Figure 1**). This method assigns a score of 2 if the critical item elicited the largest response, a score of 1 if the critical item elicited the second largest response, and a score of 0 otherwise in each block. If the average of the scores across blocks exceeds a threshold, it is judged that the examinee recognizes the critical item.

Lykken's scoring method has several advantages. First, this method is very practical. It can be used without quantification and parameter estimations. Second, because responses are ranked within each block, correction is not required even if physiological levels change between blocks as a result of habituation.

However, Lykken's scoring method has its drawback: this method does not take into account quantitative differences between responses to critical and non-critical items (Meijer et al.,

2011). For example, even when the response to the critical item might be three times as large as the next largest response, the score would be the same as when it is only slightly larger.

Z-score averaging. Z-score averaging is widely used in laboratory studies to capture quantitative differences between items (Ben-Shakhar, 1985; **Figure 1**). In this method, a response to each item is first standardized using the mean and SD of each measure within a block. The aim of the standardization is (1) to cancel out the differences in physiological levels among blocks and (2) to treat multiple measures that have different units in the same dimension. If a measure typically decreases to a critical item (e.g., heart rate, respiration, or pulse volume), its z-score is multiplied by -1 . The scores for the critical item are then averaged across all blocks and all measures. We then judge whether the averaged z-score is significantly high enough to exceed typical cut points using the standard normal distribution. This method needs no parameter estimation *a priori* and thus is easy to apply to field CIT.

However, this method has two disadvantages. First, this method assumes that for every subject, all measures respond in the normative expected direction. It thus does not consider individual differences in response patterns. The physiological measures that respond distinctively between critical and non-critical items are sometimes different between examinees (Matsuda et al., 2006). For example, Osugi (2011) reported results from field data in which a guilty examinee showed constant distinctive responses only in respiration. In such a case, with an increasing number of measures, the average z-score will become smaller and thus might lead to a false negative. Second, this method does not consider the differences in general accuracies among measures. For example, in laboratory studies, accuracy is usually higher for skin conductance than for other measures (i.e., heart rate, respiration, and pulse volume; e.g., Ben-Shakhar and Elaad, 2003; Gamer et al., 2008b). However, with z-score averaging, all measures are weighted equally. It might be preferable if each measure were weighted according to its accuracy.

Proposed statistical methods

To overcome the disadvantages of z-score averaging, other statistical methods have been proposed: logistic regression discrimination, latent class discrimination, Bayesian classification, multivariate normal distribution discrimination, and dynamic mixture distribution discrimination. We will explain these methods below and in **Figure 2**, and evaluate these methods from the viewpoint of field application. In particular, we will focus on whether a new method overcomes the limitations of z-score averaging.

Logistic regression discrimination. This method considers the differences in accuracy among measures by allocating a weight to the z-score of each measure (Gamer et al., 2006, 2008b; **Figure 2A**). The weights are acquired from the CIT datasets of previous examinees, where ground truth has already been established. Each weight reflects the effectiveness of the measure for estimating recognition. If these weights are all 1, the result will be the same as the one of z-score averaging.

This method is practical and widely used in various research domains. If the sample size is large, the weight parameters will be estimated quite stably.

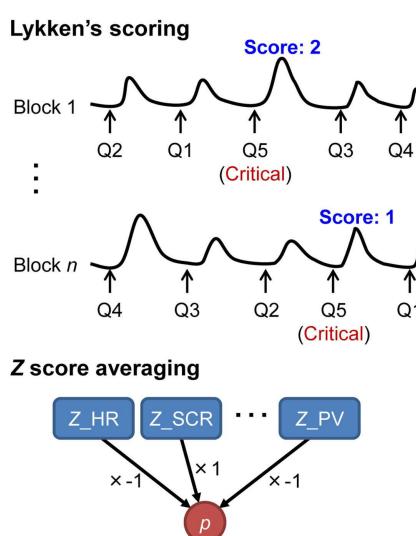
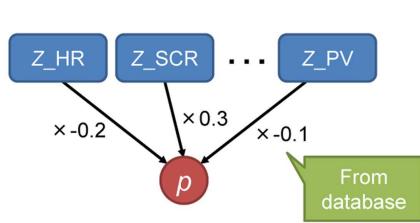
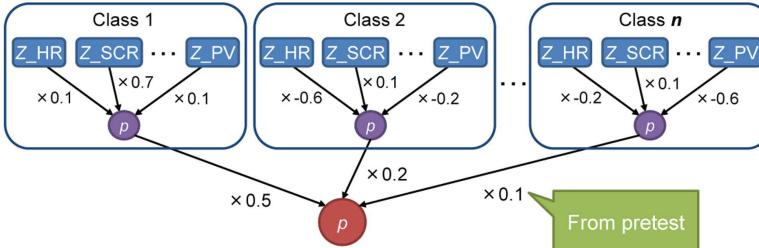


FIGURE 1 | Illustrations of the standard statistical methods: Lykken's scoring and z-score averaging. Z_{HR} , a z-score for heart rate; Z_{SCR} , a z-score for skin conductance response; Z_{PV} , a z-score for pulse volume; p , probability. Lykken's scoring assigns a score of 2 if the critical item elicited the largest response, a score of 1 if the critical item elicited the second largest response, and a score of 0 otherwise in each block. In z-score averaging, z-scores are simply averaged across blocks and measures. Z-scores may be multiplied by -1 if a smaller response is characteristic of recognition.

A Logistic regression method**B Latent class discrimination method****C Bayesian classification method**

	Z score	HR	SCR	...	PV
If recognized	Exceed a threshold (Sensitivity)	0.8	0.9	...	0.7
	Not-exceed a threshold	0.2	0.1	...	0.3
If not recognized	Exceed a threshold	0.3	0.2	...	0.4
	Not-exceed a threshold (Specificity)	0.7	0.8	...	0.6

Probability of recognition

$$\frac{0.2 \times 0.8 \times 0.9 \times \dots \times 0.3}{0.2 \times 0.8 \times 0.9 \times \dots \times 0.3 + 0.8 \times 0.3 \times 0.2 \times \dots \times 0.6} = 0.6$$

Prior probability of recognition
Prior probability of unrecognition

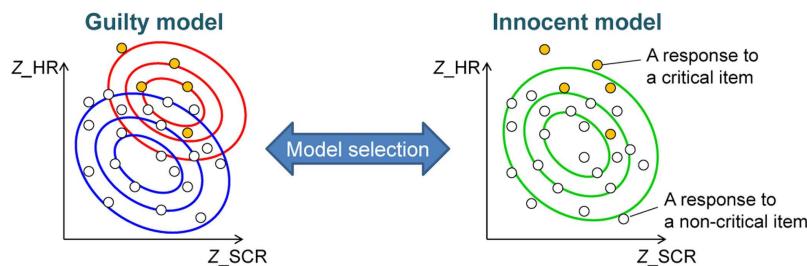
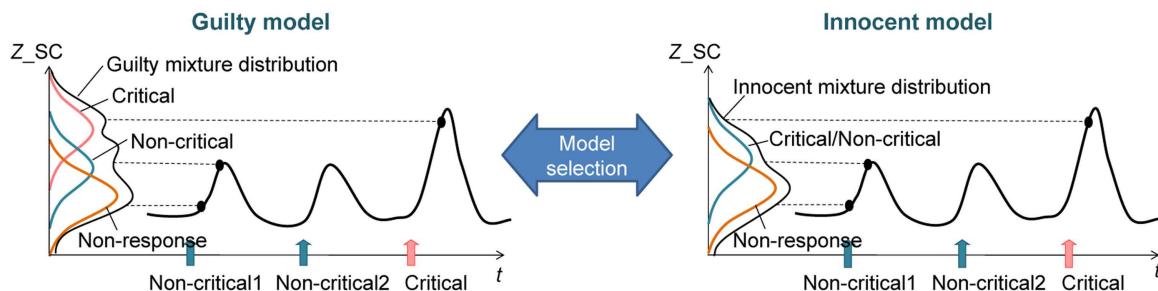
D Multivariate normal distribution method**E Dynamic mixture distribution method**

FIGURE 2 | Illustrations of the five proposed statistical discrimination methods. Z_{HR} , a z-score of heart rate; Z_{SCR} , a z-score for skin conductance response; Z_{PV} , a z-score for pulse volume; p , probability. **(A)** The logistic regression method is similar to z-score averaging, but each z-score is weighted according to the accuracy of the measure estimated from previous datasets. **(B)** The latent class discrimination method is a two-layer model of the logistic regression method. There is an appropriate regression formula for each class, and the result of the regression formula is summed across classes with a weight of the likelihood of an examinee belonging to a class according to his/her pretest result. **(C)** The Bayesian classification method calculates the probability of recognition by multiplying prior probabilities and the probabilities that a standardized response value of each measure exceeds/does not exceed a threshold in

the recognition condition. Here is the case that a participant's heart rate change and skin conductance response exceeded the threshold, while his/her pulse volume did not exceed the threshold. **(D)** In the multivariate normal distribution method, a guilty model (two-distribution model) and an innocent model (one-distribution model) are applied to the obtained responses in a CIT (each small circle represents a response to a critical (yellow) or a non-critical (white) item). The better fitted model will be selected. **(E)** The dynamic mixture distribution method uses time series and is an extended version of the multivariate normal distribution method. In this method, a guilty model (representing time series with a mixture of three distributions) and an innocent model (representing time series with a mixture of two distributions) are applied to the obtained time series in a CIT. The model that fits the time series best is selected.

On the other hand, this method does not sufficiently consider individual differences in response patterns. This is because the parameters are calculated to be fitted to the normative response pattern. Similar to z-score averaging, if a guilty examinee shows distinctive responses only in a small number of measures, this method might produce a false negative. Additionally, the logistic regression method may underperform the z-score averaging if the sample size is not large enough to reliably estimate the parameters (c.f., Dawes, 1979).

Latent class discrimination. This method is an extended version of the logistic regression model that considers individual differences in response patterns. As mentioned before, in the field CIT, an examiner conducts a pretest using cards to capture the response pattern of an examinee. However, the results of the pretest are not considered in most statistical methods. Therefore, Matsuda et al. (2006) proposed the latent class discrimination method (**Figure 2B**). In this method, previously obtained examinees are grouped into several classes, for each of which a discriminant formula (e.g., logistic regression formula) is calculated and fit to the response pattern of the examinees belonging to that class. It is then estimated if a given examinee recognizes a critical item using the following process. First, the probability that the examinee would recognize the critical item is computed by applying the discriminant formula of each class to his/her standardized response values. Second, the probability that the examinee belongs to a class is computed by using his/her pretest data. Finally, the recognition probability is calculated by summarizing each class's recognition probability across all classes with a weight of the probability for the class that the examinee belongs to. In this manner, each examinee can be distinguished through his/her response pattern.

This method considers several response patterns as latent classes. In addition, the accuracies of the measures have been reflected as parameters of a discriminant formula in each class. Moreover, these parameters can be estimated stably with a large dataset of previous examinees.

However, factoring in the pretest data can also become a drawback in practical applications. In Japan, about 5–6 CITs are typically conducted after the pretest. It takes about 2 or 3 h to finish all the CITs (Osugi, 2011). Therefore, a response pattern may change from the pretest to the last CIT for an examinee. In addition, this method is based on a more complex, hierarchical model, and consequently needs to estimate more parameters than the logistic regression method. This implies that the latent class discrimination method requires a larger dataset than the logistic regression method for parameter estimation.

Bayesian classification. This method combines multiple measures by using computations based on Bayes' theorem (Allen et al., 1992; **Figure 2C**). This approach calculates the probability that an examinee recognizes an item using (1) the sensitivity/specificity of each measure (i.e., the probability that a response value exceeds (or does not exceed) a threshold in the condition that an examinee recognizes (or does not recognize) the item) and (2) a prior probability (i.e., the probability that the examinee shows the distinctive response by chance to each item, which is determined by the number of items in the test). This method also uses a within-subjects

standardization, so that large individual differences in response magnitude are eliminated, and the pattern of responses across critical and non-critical items is retained. First, for each standardized measure, the sensitivity, specificity, and threshold are calculated from a previously obtained dataset. The standardized response value of a given examinee is then compared to the threshold. If the response value exceeds (or does not exceed) the threshold, the sensitivity (or 1 – sensitivity) is entered into Bayes' formula to calculate recognition probability. Similarly, the specificity or 1 – specificity can be entered into Bayes' formula to calculate the probability of a failure to recognize crime-relevant items.

As this method treats responses as binary data – that is, whether a response exceeds the threshold or not – quantitative differences between items are not fully captured with this method. On the other hand, thanks to dealing with binary values, this method is not excessively affected by outliers. Controlling the influence of factors that will produce outliers is difficult in the field situation as compared with the laboratory situation. For this reason, for field CIT applications, the Bayesian classification may be preferred to the other statistical methods.

Multivariate normal distribution discrimination. In contrast to logistic regression, latent class discrimination, and Bayesian classification, which require previously obtained data to estimate their parameters, the multivariate normal distribution method requires only the CIT results of the current examinee (Adachi, 1995; **Figure 2D**). If the examinee recognizes a critical item, the distribution of the responses should differ between critical and non-critical items (i.e., *guilty model*). In contrast, if the examinee does not recognize the critical item, the distribution should not differ between critical and non-critical items (i.e., *innocent model*). Both the guilty model and the innocent model are applied to the given responses in the CIT. If the guilty model better fits the responses than the innocent model, the examinee is judged as recognizing the critical item.

This method only requires that responses to critical and non-critical items differ, and does not require a previous dataset. In addition, this method has no assumptions of typical response patterns. Therefore, it can deal with various response patterns, even if the response pattern is very different from the typical normative pattern.

However, with this method, we can estimate model parameters (i.e., mean and SD of distributions) only from the given data. The sample size is thus the number of repetitions; for example, if each item is repeated five times, the sample size is five, which is too small to be used to estimate stable parameters. In addition, although the accuracy of each measure can be calculated based on previous datasets, this method does not use previous datasets. Therefore, the differences in accuracy between measures cannot be taken into account.

Dynamic mixture distribution discrimination. In order to estimate stable model parameters by using only the given data, the extended version of the multivariate normal distribution method – the dynamic mixture distribution method – was developed (Matsuda et al., 2009a; **Figure 2E**). Similar to the multivariate normal distribution method, this method prepares a guilty model and an

innocent model, but applies these models to time series data. The guilty model represents the response time series using three distributions: a non-response distribution corresponding to the base level, a critical response distribution corresponding to responses to the critical item, and a non-critical response distribution corresponding to responses to the non-critical items. In contrast, the innocent model represents the response time series using two distributions: a non-response distribution and a pooled critical/non-critical response distribution corresponding to responses to both critical and non-critical items. The guilty and innocent models are applied to the time series of the CIT data. If the time series is more compatible with the guilty model than with the innocent model, the examinee is judged as recognizing the critical item.

Similar to the multivariate normal distribution, this method requires no previous dataset and no assumption of typical response patterns. Therefore, this method is very flexible and can easily accommodate individual differences in response patterns, even if an individual's response pattern is very different from the typical normative response pattern. Additionally, because time series data are used, stable model parameters may be estimated with the typical number or repetitions in the CIT.

However, since this method does not depend on previous datasets, the accuracy of each measure cannot be taken into account. Furthermore, this method requires complex calculations for parameter estimations (i.e., Gibbs sampler). Given current technology, it takes at least about 10 min to finish the calculation of the parameters. If the calculation algorithm is improved, this method might be ideally suited to field CIT use.

Summary of statistical methods

Table 1 summarizes the advantages and disadvantages of the various statistical methods. As the table shows, a perfect statistical method does not exist. More studies are required to continue to improve existing methods.

However, the most promising method at present would appear to be the latent class discrimination method or the dynamic mixture discrimination method. **Table 1** shows the methodological advantages of the latent class and dynamic mixture distribution methods as compared to the other methods, recognizing that their parameter calculations are complex. Furthermore, superiority of these two methods in terms of discrimination performance was

demonstrated empirically (Matsuda et al., 2009a). In this study, 19 guilty participants were discriminated from 15 innocent participants by using the logistic regression, latent class, multivariate normal distribution, and dynamic mixture distribution methods. The discrimination performance was higher for the latent class and for the dynamic mixture distribution methods than for the logistic regression and the multivariate normal distribution methods. Of course, this result should be verified by using larger number of field CIT datasets. In addition, their discrimination performance should be also compared with that of the Bayesian classification method, which is expected to be robust in the face of outliers.

Methods requiring previously obtained datasets may have limited utility for filed CIT applications. Such methods (i.e., the logistic regression, latent class, and Bayesian discrimination methods) require the parameters to be estimated from the field CIT data for which valid ground truth data are available for each examinee. However, the exact confirmation of this knowledge is very difficult to obtain in the field situation, since it is difficult to know with absolute certainty who is guilty and who is innocent in a field case. It may take a rather long time to collect a sufficient number of appropriate field datasets for parameter estimation. If the parameters are estimated from an insufficient number of field samples, these methods may underperform the simple z-score averaging (Dawes, 1979). In contrast, methods that require only the current dataset (i.e., the multivariate normal distribution and dynamic mixture distribution method) have a strong advantage for field use since they do not require a previously obtained dataset. But this also indicates that the latter methods may be more influenced by missing values and measurement artifacts than the former methods. Even when adopting the latter methods, evaluating their generalizability will require using a field dataset.

ADDITIONAL MEASURES

In order to improve the probative force of the CIT in court, it would be also promising to use additional measures that can potentially increase the accuracy of the CIT. The current field CIT, that is based on measures of autonomic responses (i.e., skin conductance, heart rate, respiration, pulse volume), has been working well so far in Japan. Therefore, it would be more promising to add new measures

Table 1 | Comparison of statistical methods in terms of features that are important for field application.

Statistical method	Flexibility for individual differences	Consideration of accuracy differences among measures	Need of previous dataset for parameter estimation	Stability of parameter estimation	Complexity of model
Standard: Z-score averaging	Low	No	No	No parameters	Low
(A) Logistic regression	Low	Yes	Yes	Stable	Medium
(B) Latent class discrimination	High (assume subgroups having different response patterns)	Yes	Yes	Stable	High
(C) Bayesian classification	Medium	Yes	Yes	Stable	Medium
(D) Multivariate normal distribution	High (no assumption of a typical response pattern)	No	No	Unstable	Medium
(E) Dynamic mixture distribution	High (no assumption of a typical response pattern)	No	No	Stable	High

to the autonomic-based CIT instead of altering the current field CIT completely to use alternative measures. In this section, we will review additional CIT measures that can be obtained by using two approaches. The first approach is to refine the quantification of the classic autonomic responses. The second approach is to implement new physiological measures to augment the autonomic responses used currently.

Quantification of new/refining aspects of autonomic responses

The Improvement of current quantification methods is a simple way to increase accuracy of the current test. Here, we will review some examples of how quantification might be refined.

Respiration. Respiration has been operationalized as respiration line length in almost all CIT studies (for a review, see Gamer, 2011a). The respiration line length is defined as the sum of the moving distances of the respiration curve in a specified time interval. The respiration line length decreases when respiration is suppressed (i.e., shorter respiratory time and smaller amplitude), and thus is a good measure for the CIT. However, the line length is biased by how the parts of the respiratory cycles are included in the time interval. To account for this bias, Elaad et al. (1992) shifted the starting point of the time interval slightly, calculated the line length for each shift, and then averaged the line lengths for all shifts. However, even this method cannot remove the bias completely (Figure 2 in Matsuda and Ogawa, 2011).

To fully resolve this bias problem, a new quantification method – a weighted average respiration line length – has been recently proposed (Matsuda and Ogawa, 2011). This method calculates the respiration line length per cycle, weights it with the proportion that the cycle occupies in the time interval, and then averages the weighted line lengths across all cycles involved in the time interval. The discrimination performance was significantly better for the weighted average respiration line length than for the traditional respiration line length.

Moreover, there is an undeniable possibility that changes in respiratory rate and amplitude are elicited independently in the CIT. To extract more precise information from respiration, respiratory rate, and amplitude could be measured separately. In order to quantify these, the use of the weighted average method would be preferable (e.g., Matsuda et al., 2009a).

Pulse volume. Recently, pulse volume has been quantified as finger pulse waveform length in a way similar to that of respiration line length (Elaad and Ben-Shakhar, 2006; Vandenbosch et al., 2009). The finger pulse waveform length can reflect both pulse rate and amplitude information. As mentioned above, the line length is affected by which proportion of a cycle is included. However, the effect of this bias is much smaller for pulse volume than for respiration, because the cycle time of a pulse is much shorter. On the other hand, since heart rate is computed with an electrocardiogram in Japan, the measurement of finger pulse volume length is redundant.

In Japan, normalized pulse volume has been applied to the field CIT to evaluate vascular tone more accurately. The normalized pulse volume is computed per pulse cycle by dividing the

amplitude of the cycle by the average voltage during the cycle. The normalized pulse volume is advocated as a more valid measure for the assessment of vascular tone than the usual pulse volume (Sawada et al., 2001). The validity of the normalized pulse volume has also been confirmed in a CIT study (Matsuda et al., 2009a).

Adding new measures

New physiological or behavioral measures can be recorded in addition to autonomic responses in the field, particularly if the recording is easy and stable. Here, we will review reaction time, facial features, fMRI activations, and EEG/ERP features.

Reaction time. One possible measure that has been considered is reaction time after item onset (for a review, see Verschuere and De Houwer, 2011). Some studies reported high accuracy of individual classification using reaction time. For example, Allen et al. (1992) reported a sensitivity of 0.950 and the specificity of 1.000.

However, in the current situation in the field, there may be problems with using reaction time. First, reaction time can be controlled intentionally. It might therefore be easier to use countermeasures that affect reaction time than those that affect autonomic responses. In fact, some studies use the response time as a measure of countermeasures (Rosenfeld et al., 2008; Winograd and Rosenfeld, 2011). Second, it is uncertain whether examinees would follow the instructions, such as “respond as quickly and accurately as possible.” Unlike the autonomic-based CIT, a reaction-time task requires examinees to respond actively. Even when examinees are innocent, however, they may not take the test willingly and thus may not cooperate. In addition, attributes of field examinees are more diverse than those of participants in laboratory studies. For example, elder examinees have slower and more variable reaction-times, which might render this measure less useful in some populations.

Despite these limitations, research might profit from further examination of reaction time in the CIT. It is an easily obtained measure, and individual differences in response times might not be of concern if quantified using within-subject metrics (*z*-scores). Moreover, it might be possible to identify reaction-time response patterns that would suggest when reaction time can, and when it cannot, provide useful information.

Facial features. Facial expressions have potential as a measure in current field CIT examinations. Because a face is usually not covered, it is easy to record the information without attaching special electrodes (i.e., with a remote-sensing technique).

It is well-known that lie detection can make use of facial muscle activity (Ekman, 2001). However, as far as we know, no study has reported the use of facial muscle changes in the CIT, but automated Facial action coding system (FACS; Littlewort et al., 2011) might make this an easy possibility to explore further. On the other hand, facial skin surface temperature has been measured in the CIT (Pollina et al., 2006). In this study, the temperature increased for critical items compared to non-critical items in a region below the eyes. Its individual classification result was a sensitivity of 0.917 and a specificity of 0.917.

Information related to the eyes has also been applied to the CIT. Startle eye blinks reduced more for critical items than for

non-critical items (Verschuere et al., 2007). Temporal distributions of blinks differed between critical and non-critical items (Fukuda, 2001). Pupil sizes increased more for critical items than for non-critical items (Bradley and Janisse, 1981; Lubow and Fein, 1996). Lubow and Fein (1996) reported a sensitivity of 0.50–0.70 and a specificity of 1.00 using pupil sizes.

Thus a variety of facial measures show some promise for use in the CIT, but none have been extensively researched. Therefore, future research should determine if use of these facial measures can increase the validity of the current autonomic-based CIT.

fMRI. Recent research has utilized fMRI in CIT-like experiments (for a review, see Gamer, 2011b). Nose et al. (2009) reported the accuracy of fMRI in the CIT: the sensitivity was 0.84 and the specificity was 0.84. However, the use of fMRI in the field would be difficult at the present time. First, the equipment for fMRI is expensive and not portable. Second, examinees must be extremely cooperative as they are not able to move during the fMRI scanning and would have to tolerate the noise during the test. Third, some examinees could not be tested if they have metal in their bodies that would make fMRI unsafe. Although technical improvement of recordings and analyses are expected in future research, fMRI measures may inherently carry no more or no less weight than other measures used in the CIT.

EEG/ERPs. Many laboratory studies have measured EEG during the CIT and reported significant differences in ERP components between critical and non-critical items, especially P3 amplitudes (Rosenfeld et al., 1988; Farwell and Donchin, 1991; Allen et al., 1992; Rosenfeld, 2011). A recent meta-analysis showed that the P3 measure is more effective than the traditional autonomic measures in detecting participants' concealed knowledge: Cohen's d was 2.55 for the P3 amplitude and 1.72 for skin conductance response (Ben-Shakhar and Meijer, 2012). This result is similar to that of Allen and Iacono (1997), in which they compared the area under ROC curve from their ERP data to published skin conductance data. The increase of the P3 amplitude is thought to reflect the significance of the critical item for the examinees (Rosenfeld, 2011), which is often embedded within an oddball paradigm. In addition, recent studies with rather long inter-stimulus intervals (>7 s) reported the increase of the N2 (Matsuda et al., 2009b, 2012; Gamer and Berti, 2010) and the late positive potential (Matsuda et al., 2009b, 2012) for the critical item.

Due to the progress of recording and analysis techniques it has become easier to measure EEG in field situations. In fact, an EEG can be recorded with a polygraph system currently used in field CIT in Japan, although the stimulus presentation/control system for it has not been equipped yet. A recent study measured ERPs under the standard protocol of the autonomic-based field CIT (Matsuda et al., 2011). This study showed that late positive potential significantly differed between critical and non-critical items, even when each item was presented only five times. Importantly, including the late positive potential improved the discrimination performance of the standard autonomic-based CIT. Furthermore, Rosenfeld (2011) have

proposed a new protocol of the ERP-based CIT in order to make the test resistant to countermeasures ("complex trial-based CIT"), and have reported high accuracies. Collectively these studies indicate that features of the ERP would be promising additions to the field CIT.

Moreover, although most studies quantified EEG in the time domain, some recent studies focused on information in the frequency domain (Abootalebi et al., 2006, 2009; Zhao et al., 2011). These studies show that differences in wavelet features can reflect the differences between critical and non-critical items. Furthermore, the frontal asymmetry of left and right EEG alpha power may have promise as a new measure. Frontal EEG asymmetry is an index of the basic emotional dimension of approach versus withdrawal (Coan and Allen, 2004). In the CIT, relative right frontal alpha activity was significantly lower for critical items than for non-critical items (Matsuda et al., submitted). This result suggests that the critical item would elicit withdrawal-oriented motivation and emotion, which may be an additional indicator of recognition of the critical item.

SUMMARY

In the present paper, we reviewed how the CIT has been used for field criminal investigations in Japan, and suggest that with appropriate training and institutional support, the CIT can frequently be used in field applications. We also reviewed various statistical methods and potential new measures, which may contribute to improved validity and increased probative value of the CIT. We suggested that more studies of these various statistical methods are required before applying the statistical methods in the field. We also highlighted the promise of adding new quantification of existing measures and adding new measures such as EEG/ERP indices to the current field CIT. It should be an immediate goal of the Japanese CIT examiners and researchers to improve the probative value of the field CIT by introducing statistical judgment methods and then adding new measures to the current CIT.

Despite improvements in measures and statistical assessment, it is important to remember that the CIT is not a test to judge whether an examinee is guilty or innocent. The CIT can show only with relatively high probability whether the examinee recognizes the crime-relevant item. The examinee may have obtained crime-relevant information by any number of means, only one of which is by being the perpetrator of the crime, while others include accidental exposure via media or interrogations, or exposure via a relationship with the perpetrator of the crime; a good examiner of course pays close attention to remove these possibilities. However, the CIT result can be used as one scientific indicator of whether an individual may have been involved in the crime under investigation. Given the fundamentally sound paradigm of the CIT, and the promise of improvements using more sophisticated statistics and additional measures, we hope that the use of the CIT will increase, with Japan's implementation serving as a useful model.

ACKNOWLEDGMENTS

This study was supported in part by KAKENHI 24730650. We thank Tokihiro Ogawa, Michiko Tsuneoka, and the reviewers for their helpful comments.

REFERENCES

- Abootalebi, V., Moradi, M. H., and Khalilzadeh, M. A. (2006). A comparison of methods for ERP assessment in a P300-based GKT. *Int. J. Psychophysiol.* 62, 309–320.
- Abootalebi, V., Moradi, M. H., and Khalilzadeh, M. A. (2009). A new approach for EEG feature extraction in P300-based lie detection. *Comput. Methods Programs Biomed.* 94, 48–57.
- Adachi, K. (1995). Statistical classification procedures for polygraph tests of guilty knowledge. *Behav. Informetrika* 22, 49–66.
- Allen, J. J. B., and Iacono, W. G. (1997). A comparison of methods for the analysis of event-related potentials in deception detection. *Psychophysiology* 34, 234–240.
- Allen, J. J. B., Iacono, W. G., and Danielson, K. D. (1992). The identification of concealed memories using the event-related potential and implicit behavioral measures: a methodology for prediction in the face of individual differences. *Psychophysiology* 29, 504–522.
- Allen, J. J. B., and Mertens, R. (2008). Limitations to the detection of deception: true and false recollections are poorly distinguished using an event-related potential procedure. *Soc. Neurosci.* 4, 473–490.
- Ben-Shakhar, G. (1985). Standardization within individuals: a simple method to neutralize individual differences in skin conductance. *Psychophysiology* 22, 292–299.
- Ben-Shakhar, G. (2002). “A critical review of the control questions test (CQT),” in *Handbook of Polygraph Testing*, ed. M. Kleiner (San Diego: Academic Press), 103–126.
- Ben-Shakhar, G. (2011). “Countermeasures,” in *Memory Detection: Theory and Application of the Concealed Information Test*, eds B. Verschueren, G. Ben-Shakhar, and E. Meijer (Cambridge: Cambridge University Press), 200–214.
- Ben-Shakhar, G., and Dolev, K. (1996). Psychophysiological detection through the guilty knowledge technique: effects of mental countermeasures. *J. Appl. Psychol.* 81, 273–281.
- Ben-Shakhar, G., and Elaad, E. (2002). Effects of questions’ repetition and variation on the efficiency of the guilty knowledge test: a reexamination. *J. Appl. Psychol.* 87, 972–977.
- Ben-Shakhar, G., and Elaad, E. (2003). The validity of psychophysiological detection of information with the guilty knowledge test: a meta-analytic review. *J. Appl. Psychol.* 88, 131–151.
- Ben-Shakhar, G., and Meijer, E. (2012). Skin conductance, respiration, heart rate, and P300 in the concealed information test: a meta analysis. *Int. J. Psychophysiol.* 85, 324.
- Bradley, M. T., Barefoot, C. A., and Arsenault, A. M. (2011). “Leakage of information to innocent suspects,” in *Memory Detection: Theory and Application of the Concealed Information Test*, eds B. Verschueren, G. Ben-Shakhar, and E. Meijer (Cambridge: Cambridge University Press), 90–113.
- Bradley, M. T., and Janisse, M. P. (1981). Accuracy demonstrations, threat, and the detection of deception: cardiovascular, electrodermal, and pupillary measures. *Psychophysiology* 18, 307–315.
- Carmel, D., Dayan, E., Naveh, A., Raveh, O., and Ben-Shakhar, G. (2003). Estimating the validity of the guilty knowledge test from simulated experiments: the external validity of mock crime studies. *J. Exp. Psychol. Appl.* 9, 261–269.
- Coan, J. A., and Allen, J. J. B. (2004). Frontal EEG asymmetry as a moderator and mediator of emotion. *Biol. Psychol.* 67, 7–49.
- Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *Am. Psychol.* 34, 571–582.
- Ekman, P. (2001). *Telling Lie. Clues in the Marketplace, Politics, and Marriage*. New York: Norton.
- Elaad, E. (1990). Detection of guilty knowledge in real-life criminal investigations. *J. Appl. Psychol.* 75, 521–529.
- Elaad, E., and Ben-Shakhar, G. (2006). Finger pulse waveform length in the detection of concealed information. *Int. J. Psychophysiol.* 61, 226–234.
- Elaad, E., Ginton, A., and Jungman, N. (1992). Detection measures in real-life criminal guilty knowledge tests. *J. Appl. Psychol.* 77, 757–767.
- Farwell, L. A., and Donchin, E. (1991). The truth will out: Interrogative polygraph (“lie detection”) with event-related brain potentials. *Psychophysiology* 28, 531–547.
- Fukuda, K. (2001). Eye blinks: new indices for the detection of deception. *Int. J. Psychophysiol.* 40, 239–245.
- Gamer, M. (2011a). “Detecting concealed information using autonomic measures,” in *Memory Detection: Theory and Application of the Concealed Information Test*, eds B. Verschueren, G. Ben-Shakhar, and E. Meijer (Cambridge: Cambridge University Press), 151–170.
- Gamer, M. (2011b). “Detecting of deception and concealed information using neuroimaging techniques,” in *Memory Detection: Theory and Application of the Concealed Information Test*, eds B. Verschueren, G. Ben-Shakhar, and E. Meijer (Cambridge: Cambridge University Press), 27–45.
- Gamer, M. (2012). Task relevance and recognition of concealed information have different influences on electrodermal activity and event-related brain potentials. *Psychophysiology* 47, 355–364.
- Gamer, M., Gödert, H. W., Keth, A., Rill, H.-G., and Vossel, G. (2008a). Electrodermal and phasic heart rate responses in the guilty actions test: comparing guilty examinees to informed and uninformed innocents. *Int. J. Psychophysiol.* 69, 61–68.
- Gamer, M., Verschueren, B., Crombez, G., and Vossel, G. (2008b). Combining physiological measures in the detection of concealed information. *Physiol. Behav.* 95, 333–340.
- Gamer, M., Kosiol, D., and Vossel, G. (2010). Strength of memory encoding affects physiological responses in the guilty actions test. *Biol. Psychol.* 83, 101–107.
- Gamer, M., Rill, H. G., Vossel, G., and Gödert, H. W. (2006). Psychophysiological and vocal measures in the detection of guilty knowledge. *Int. J. Psychophysiol.* 60, 76–87.
- Hira, S., and Furumitsu, I. (2002). Polygraphic examinations in Japan: applications of the guilty knowledge test in forensic investigations. *Int. J. Police Sci. Manag.* 4, 16–27.
- Honts, C. R., Devitt, M. K., Winbush, M., and Kircher, J. C. (1996). Mental and physical countermeasures reduce the accuracy of the concealed knowledge test. *Psychophysiology* 33, 84–92.
- Kobayashi, T., Yoshimoto, K., and Fujihara, S. (2009). Jitsumu polygraph kensa no genjyo [The contemporary situation of field polygraph tests]. *Jpn. J. Physiol. Psychol. Psychophysiol.* 27, 5–15.
- Krapohl, D. J. (2011). “Limitations of the concealed information test in criminal cases,” in *Memory Detection: Theory and Application of the Concealed Information Test*, eds B. Verschueren, G. Ben-Shakhar, and E. Meijer (Cambridge: Cambridge University Press), 293–302.
- Nahari, G., and Ben-Shakhar, G. (2011). Psychophysiological and behavioral measures for detecting concealed information: the role of memory for crime details. *Psychophysiology* 48, 733–744.
- Nose, I., Murai, J., and Taira, M. (2009). Disclosing concealed information on the basis of cortical activations. *Neuroimage* 44, 1380–1386.
- Lubow, R. E., and Fein, O. (1996). Pupillary size in response to a visual guilty knowledge test: new technique for the detection of deception. *J. Exp. Psychol. Appl.* 2, 164–177.
- Lykken, D. T. (1959). The GSR in the detection of guilt. *J. Appl. Psychol.* 43, 385–388.
- Matsuda, I., Hirota, A., Ogawa, T., Takasawa, N., and Shigemasu, K. (2006). A new discrimination method for the concealed information test using pretest data and within-individual comparisons. *Biol. Psychol.* 73, 157–164.
- Matsuda, I., Hirota, A., Ogawa, T., Takasawa, N., and Shigemasu, K. (2009a). Within-individual discrimination on the concealed information test using dynamic mixture modeling. *Psychophysiology* 46, 439–449.
- Matsuda, I., Nittono, H., Hirota, A., Ogawa, T., and Takasawa, N. (2009b). Event-related brain potentials during the standard autonomic-based concealed information test. *Int. J. Psychophysiol.* 74, 58–68.
- Matsuda, I., Nittono, H., and Ogawa, T. (2011). Event-related potentials increase the discrimination performance of the autonomic-based concealed information test. *Psychophysiology* 48, 1701–1710.
- Matsuda, I., Nittono, H., and Ogawa, T. (2012). Decomposing cognitive processes underlying the concealed information test by event-related potentials. *Int. J. Biomed. Soft Comput. Hum. Sci.* 18, 5–11.
- Matsuda, I., and Ogawa, T. (2011). Improved method for calculating the respiratory line length in the concealed information test. *Int. J. Psychophysiol.* 81, 65–71.
- Meijer, E. H., Verschueren, B., and Ben-Shakhar, G. (2011). “Practical guidelines for developing a CIT,” in *Memory Detection: Theory and Application of the Concealed Information Test*, eds B. Verschueren, G. Ben-Shakhar, and E. Meijer (Cambridge: Cambridge University Press), 293–302.

- Osugi, A. (2010). Gap and connection between laboratory research and field application of the CIT in Japan. *Int. J. Psychophysiol.* 77, 238.
- Osugi, A. (2011). "Daily application of the concealed information test: Japan," in *Memory Detection: Theory and Application of the Concealed Information Test*, eds B. Verschueren, G. Ben-Shakhar, and E. Meijer (Cambridge: Cambridge University Press), 253–275.
- Pollina, D., Dollins, A., Senter, S., Brown, T., Pavlidis, I., Levine, J., et al. (2006). Facial skin surface temperature changes during a "concealed information" test. *Ann. Biomed. Eng.* 34, 1182–1189.
- Reid, J. E. (1947). A revised questioning technique in lie-detection tests. *J. Crim. Law Criminol.* 37, 542–547.
- Rosenfeld, J. P. (2011). "P300 in detecting concealed information," in *Memory Detection: Theory and Application of the Concealed Information Test*, eds B. Verschueren, G. Ben-Shakhar, and E. Meijer (Cambridge: Cambridge University Press), 63–89.
- Rosenfeld, J. P., Cantwell, B., Nasman, V. T., Wojdac, V., Ivanov, S., and Mazzetti, L. (1988). A modified, event-related potential-based guilty knowledge test. *Int. J. Neurosci.* 42, 157–161.
- Rosenfeld, J. P., Labkovsky, E., Winograd, M., Lui, M. A., Vandenboom, C., and Chedid, E. (2008). The complex trial protocol (CTP): a new, countermeasure-resistant, accurate, P300-based method for detection of concealed information. *Psychophysiology* 45, 906–919.
- Sartori, G., Agosta, S., Zogmaister, C., Ferrara, S. D., and Castiello, U. (2008). How to accurately detect autobiographical events. *Psychol. Sci.* 19, 772–780.
- Sawada, Y., Tanaka, G., and Yamakoshi, K. (2001). Normalized pulse volume (NPV) derived photoplethysmographically as a more valid measure of the finger vascular tone. *Int. J. Psychophysiol.* 41, 1–10.
- Vandenbosch, K., Verschueren, B., Crombez, G., and De Clercq, A. (2009). The validity of finger pulse line length for the detection of concealed information. *Int. J. Psychophysiol.* 71, 118–123.
- Venables, P. H., and Mitchell, D. A. (1996). The effects of age, sex and time of testing on skin conductance activity. *Biol. Psychol.* 43, 87–101.
- Verschueren, B., Crombez, G., Koster, E. H. W., Van Bockstaele, B., and De Clercq, A. (2007). Startling secrets: startle eye blink modulation by concealed crime information. *Biol. Psychol.* 76, 52–60.
- Verschueren, B., and De Houwer, J. (2011). "Detecting concealed information in less than a second: response latency-based measures," in *Memory Detection: Theory and Application of the Concealed Information Test*, eds B. Verschueren, G. Ben-Shakhar, and E. Meijer (Cambridge: Cambridge University Press), 46–62.
- Verschueren, B., Meijer, E., and De Clercq, A. (2011). Concealed information under stress: a test of the orienting theory in real-life police interrogations. *Legal Criminol. Psychol.* 16, 348–356.
- Winograd, M. R., and Rosenfeld, J. P. (2011). Mock crime application of the complex trial protocol (CTP) P300-based concealed information test. *Psychophysiology* 48, 155–161.
- Zhao, M., Zheng, C., and Zhao, C. (2011). A new approach for concealed information identification based on ERP assessment. *J. Med. Syst.* 36, 2401–2409.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 31 July 2012; accepted: 10 November 2012; published online: 27 November 2012.

*Citation: Matsuda I, Nittono H and Allen JJB (2012) The current and future status of the concealed information test for field use. *Front. Psychology* 3:532. doi: 10.3389/fpsyg.2012.00532*

This article was submitted to Frontiers in Cognitive Science, a specialty of Frontiers in Psychology.

Copyright © 2012 Matsuda, Nittono and Allen. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.



Advancing lie detection by inducing cognitive load on liars: a review of relevant theories and techniques guided by lessons from polygraph-based approaches

Jeffrey J. Walczyk *****, Frank P. Igou, Alexa P. Dixon and Talar Tcholakian

Psychology and Behavioral Sciences, Louisiana Tech University, Ruston, LA, USA

Edited by:

Wolfgang Ambach, Institute for Frontier Areas of Psychology and Mental Health, Germany

Reviewed by:

Eric Postma, Tilburg University, Netherlands

Travis Seymour, University of California Santa Cruz, USA

***Correspondence:**

Jeffrey J. Walczyk, Psychology and Behavioral Sciences, Louisiana Tech University, P.O. Box 10048, Ruston, LA 71272, USA.

e-mail: walczyk@latech.edu

This article critically reviews techniques and theories relevant to the emerging field of “lie detection by inducing cognitive load selectively on liars.” To help these techniques benefit from past mistakes, we start with a summary of the polygraph-based Controlled Question Technique (CQT) and the major criticisms of it made by the National Research Council (2003), including that it is not based on a validated theory and administration procedures have not been standardized. Lessons from the more successful Guilty Knowledge Test are also considered. The critical review that follows starts with the presentation of models and theories offering insights for cognitive lie detection that can undergird theoretically load-inducing approaches. This is followed by evaluation of specific research-based, load-inducing proposals, especially for their susceptibility to rehearsal and other countermeasures. To help organize these proposals and suggest new direction for innovation and refinement, a theoretical taxonomy is presented based on the type of cognitive load induced in examinees (intrinsic or extraneous) and how open-ended the responses to test items are. Finally, four recommendations are proffered that can help researchers and practitioners to avert the corresponding mistakes with the CQT and yield new, valid cognitive lie detection technologies.

Keywords: cognition of deception, cognitive lie detection, rehearsed deception, polygraph, inducing cognitive load

The seemingly disparate fields of “polygraph-based lie detection” and “research and theory on social-cognitive aspects of deception” seldom communicate. Still, lessons from the former may benefit attempts made from the latter perspective to detect lies. A goal of this critical review is to advance a new research area of the social-cognitive perspective, “lie detection by inducing cognitive load selectively on liars,” to develop on valid theoretical grounds and avoid other pitfalls that hampered the Controlled Question Technique (CQT), a questioning paradigm used with the polygraph. To this end, the CQT is summarized and major criticisms of it made by the National Research Council (2003) are shared. Some of them are that it is not based on a valid theory and is highly susceptible to countermeasures. Also summarized is the more successful polygraph-based Guilty Knowledge Test (GKT, a.k.a. the Concealed Knowledge Test), which overcomes many of these concerns (Lykken, 1998). With these lessons in mind and to help load-inducing lie detection efforts to develop on valid theoretical grounds, the critical review begins with discussion of models and theories relevant to the cognition of deception for their insights on cues to deception. Next, we consider the specific proposals appearing in the literature that try to make it cognitively more difficult to lie than to tell the truth, especially for their susceptibility to countermeasures. Then, a taxonomy of load-inducing lie detection is presented to organize these proposals and open up new research avenues. Coming full circle, we conclude with four

recommendations for researchers and practitioners to avoid the corresponding problems with the CQT.

SUMMARY OF POLYGRAPH-BASED LIE DETECTION: ITS USES, PITFALLS, AND SUCCESSES

This section is not part of the review. Rather, it is a summary of certain aspects of polygraph-based lie detection. Critical reviews in this area are available elsewhere (e.g., Lykken, 1998; National Research Council, 2003). The *polygraph* is a device that continuously records psycho-physiological arousal as assessed by pulse rate, blood pressure, respiration rate, and skin conductivity, which has been applied to uncover deception. The most common questioning paradigm used with it for detecting lies is the CQT. In a typical test, an examinee is given a pretest interview for gathering information that can serve as the basis for control questions. Once questions are chosen, the examiner will preview them with the examinee to ensure that the questions are understood and do not surprise the examinee when asked later. During the exam, irrelevant questions are asked such as “What is your name?” along with control questions that most people tend to lie to. For example, “Have you ever stolen anything from the workplace?” Finally, relevant questions probe the issue central to the exam (e.g., “Did you kill . . .?”). The questions usually elicit brief answers. A liar is hypothesized to show more arousal to relevant questions than to control questions, whereas an innocent individual (truth teller)

should show more arousal to control questions than to relevant questions (Lykken, 1998). Law enforcement and federal agencies in the USA use the CQT as a screening device for hiring and retaining employees and as a tool for criminal investigations. The CQT has been used to verify victim's statements, evaluate the veracity of witnesses, and to exonerate suspects. Still, test results are largely inadmissible in US courtrooms (National Research Council, 2003).

The validity of the CQT has been challenged in a 2003 report by a distinguished panel of scientists of the National Research Council, which reviewed all available scientific studies and offered several criticisms. Among the most serious, the administration procedures for the CQT have not been standardized. The paradigm has a high rate of false positives (honest individuals misclassified as liars), is highly susceptible to countermeasures, and the results of examinations are subjectively scored. The criticism most germane here is that *the CQT is not based on a theory of deception that has been validated*. For instance, an assumption central to the CQT, that lying causes more sympathetic nervous system arousal than truth telling, is unsubstantiated. The panel called for research on alternatives. Many of these criticisms are translated later in this article into specific recommendations for advancing cognitive load-inducing lie detection techniques in ways that overcome these criticisms.

Partly in response to the validity concerns with the CQT, the GKT was proposed. It is a questioning paradigm that can be used with the polygraph to uncover the false denials of examinees by exposing whether they possess "guilty knowledge", presumably resulting from their participation in a crime (Lykken, 1998). During a GKT, the examinee is presented with multiple-choice questions, each having one relevant alternative (correct answer) and several neutral alternatives (plausible distractors). The latter should be chosen such that an innocent person could not discriminate them from the relevant alternative (Lykken, 1998). An example of a relevant question is "How was the victim killed?" with the response alternatives of "shot," "stabbed," "struck," "strangled," or "poisoned." This question could be re-asked multiple times, along with other questions probing different aspects of a crime scene. The examinee does not even need to answer. If heightened arousal occurs consistently to relevant responses, then the examinee may be concealing knowledge as the perpetrator. The GKT assumes that innocent examinees could not have acquired guilty knowledge indirectly and that guilty examinees encoded guilty knowledge and have retained it (Elaad, 1990).

Some validity concerns with the CQT were resolved in the GKT, including more standardization of the procedure, more appropriate control alternatives, fewer false positives, and a stronger theoretical basis (Lykken, 1998; Carmel et al., 2003). Also, beyond the psycho-physiological measures of the polygraph, guilty knowledge has been demonstrated with the diverse cues of response time (Seymour et al., 2000; Seymour and Kerlin, 2008; Seymour and Fraynt, 2009), event-related potentials (Rosenfeld et al., 1988, 2006), and pupil dilation (Dionisio et al., 2001), among others. The relative success of the GKT also offers lessons for the development of load-inducing lie detection techniques, especially that they should be based on a valid theory. Still, the GKT is limited in

the deception it can uncover to the false denials of those possessing guilty knowledge.

MODELS/THEORIES ADVANCING UNDERSTANDING OF THE COGNITION OF DECEPTION

Recalling that the CQT is not based on a validated theory of deception, we next review models and theories offering insights on the cognition of deception to help new load-inducing lie detection techniques to advance on solid theoretical ground. As will be discussed later, most of them lack such a foundation. Some accounts were proposed to explain social aspects of deceit, but offer important cognitive insights.

SELECTED ACCOUNTS UNDERGIRDING THE GUILTY KNOWLEDGE TEST

Various theoretical accounts of how the GKT works have been proposed. Two are particularly relevant to the cognition of deception. The first, Orienting Response Theory, focuses on attentional processes. According to it, individuals tend to orient and attend carefully to environmental stimuli that are novel or emotionally significant to them, thereby preparing themselves to respond adaptively as necessary (Sokolov, 1963). Applied to the GKT, an orienting response naturally occurs in guilty examinees on exposure to relevant knowledge, as evidenced, for instance, by a lowering of heart rate, but not to neutral alternatives (Verschuere et al., 2010). It can manifest behaviorally in longer response times to process a stimulus (Seymour et al., 2000) and in other ways. A defensive response to a relevant option is also possible if an examinee feels threatened, characterized by increased heart rate and by other signs of arousal. The orienting response to guilty knowledge is hypothesized to be automatic and hard to suppress (Lykken, 1998).

Seymour (2001) proposed a memory-based alternative to Orienting Response Theory called the Parallel Task Set (PTS) model, which explains the "guilty knowledge effect" via response competition. PTS holds that an examinee's responses to the alternatives of a question of the GKT consist of the following: memory processes, response selection, response preparation, and motor execution. These four components comprise a task set. Two task sets are hypothesized to occur independently and in parallel for each question. The *familiarity task set* occurs quickly and involves automatic priming mechanisms. The *recollection task set*, on the other hand, occurs more slowly, is under conscious control, and draws on cognitive resources. In the case of the relevant alternative, two inconsistent response requests can be received by a particular response processor (e.g., that controlling verbal utterances). In this case, the one received from the familiar task set is for a truthful response while another from the recollection task set is for a deceptive response. One response can also be received while the other response is underway. In both cases, response conflict occurs. Hiding guilty knowledge is postulated to activate conflict resolution, which involves the examinee overriding the familiar response and executing the intended response of denying the guilty knowledge. This model explains the longer response times needed to do so as resulting from the additional processing steps of the recollection task set and the resolution of response conflict. The general insights that the PTS model offers for lie detection are to underscore the centrality of memory processes in deception and truth

telling and the fact that the inhibition of a familiar response is often part of deception. Both accounts of the GKT imply that the possession of guilty knowledge manifests in implicit memory measures, which are subtle and hard to hide (Anderson, 2000).

FOUR-FACTOR THEORY

Zuckerman et al. (1981) proposed the influential Four-Factor Theory of deception. It postulates that deception involves (a) generalized arousal, (b) anxiety, guilt, and other emotions accompanying deception, (c) cognitive components, and (d) liars' attempts to control verbal and non-verbal cues to appear honest. Although these authors speculate that lying imposes greater cognitive load than truth telling, which can result in longer response times, more pupil dilation, and in other signs of load, the theory does not detail the cognitive mechanisms of lying. Still, it highlights the complex, multidimensional nature of deception, and the many types of behavior (e.g., cognitive, physiological, emotional) that are potential cues.

INTERPERSONAL DECEPTION THEORY

Interpersonal Deception Theory (Buller and Burgoon, 1996; Burgoon and Buller, 2008) focuses on the dynamic, interdependent nature of verbal and non-verbal exchanges between the liar and target (the intended recipient of a deception). Specifically, it describes deception as involving (a) an interaction in which each party of a communicative dyad is monitoring the behavior and responding to cues from the other. (b) The use of strategic deception is postulated to impose a cognitive load on liars absent in truth tellers. Deceivers must consciously manipulate information to create a plausible message, appear honest as they share it, monitor targets' reactions, and perform other mental tasks. (c) Too many concurrent tasks produce "cognitive overload," resulting in some behavior going unmonitored. (d) Signs of deception include *uncertainty* and *vagueness* in the detail of a false narrative, *non-immediacy* of responses that involve frequent pausing, and *withdrawal* by sitting away from targets. Liars use *disassociations* to distance themselves from acts of deception, for instance, by describing their actions in a false narrative as going along with the group rather than as resulting from a personal choice.

Four-Factor Theory and Interpersonal Deception Theory posit that a leakage of cues can accompany liars' strategic control over behavior, especially under high cognitive load. In their review, Zuckerman et al. (1981) found the most reliable leaked cues were the use of self-adaptors (fidgeting hand movements), increased blinking and pupil dilation, heightened voice pitch, and speech errors (grammatical mistakes, slips of the tongue), pausing, and other speech hesitations, and discrepancies between verbal and non-verbal channels. Some cognitive load-inducing techniques we review exploit the fact that under high cognitive load, it is hard for examinees to monitor and control certain channels of behavior, which may maximize the leakage of non-verbal cues as a result.

PREOCCUPATION MODEL OF SECRECY

The cognitive load of lies of omission is central to Lane and Wegner's (1995) Preoccupation Model of Secrecy. It postulates that when individuals keep secrets, for instance, one from a spouse about having been unfaithful, (a) the strategy most often used is

thought suppression. ("I will stop thinking about having cheated to avoid accidentally blurting it.") (b) Over time, this ongoing suppression can cause the secret to intrude in the thoughts of the individual. ("I can't stop thinking about what I did.") (c) Intrusive thoughts renew attempts at thought suppression. ("I will try harder to block the memory.") (d) This cycle can escalate to the point that the individual obsesses over the memory long after a secret has been divulged. As does the PTS model, this account notes the difficulty often involved in concealing "guilty knowledge."

In four studies, Lane and Wegner (1995) found support for steps *a* through *d* and evidence that keeping a secret over time, ironically, increases its accessibility above other memories. Although this model focuses on lies of omission, it has relevance to deception generally. Since most lying involves keeping a secret by withholding some truth, it may help explain the fact that an allocation of cognitive resources is often required to inhibit responding truthfully (Pennebaker and Chew, 1985; Johnson et al., 2004; Kozel et al., 2004; Osman et al., 2009), just as it occurs in thought suppression. Expanding this account, for instance, by integrating it with the PTS model, should increase understanding of when lying requires cognitive resources to inhibit the truth and thereby help pinpoint when cognitive load indices make the most reliable cues to deception. Also, if secret truths become more accessible over time, they may be inadvertently blurted under high cognitive load, an implication of this model for lie detection. Finally, like the PTS model, the Preoccupation model emphasizes memory processes in deception, which in this case is for the active suppression of the truth.

SELF-PRESENTATION THEORY

DePaulo (1992) proposed a Self-Presentation Theory of individuals' control over their non-verbal behavior to create specific impressions in the minds of others, including deceptive ones. Three cognitive phases are thought to occur. (a) First, an intention to regulate one's behavior is formed to create a desired impression. (b) Then, the intended self-presentation is translated into non-verbal behaviors. (c) Finally, performance is appraised by the individual, if possible, and lessons are learned for the improvement of future performance. There are obstacles to steps (b) and (c). To note a few, many non-verbal behaviors are hard to monitor, control, or inhibit continuously, such as the expression of basic emotions on the face or the tone of one's voice (Ekman, 2001). Moreover, the non-verbal behaviors emitted are often more accessible to observers than to those producing them, which makes self-appraisal difficult (DePaulo, 1992).

Although not all self-presentations are deceptive, this account can be regarded as a theory of non-verbal deception. It is consistent with other accounts in that deception involves the intent to misrepresent (Ekman, 2001). Also, like Four-Factor Theory and Interpersonal Deception Theory, it posits that leaked non-verbal cues can signal deception. This account offers more insights than the others on the thought processes involved in looking at a potential lie from the target's perspective.

A WORKING MEMORY MODEL OF DECEPTION

Sporer and Schwandt (2006, 2007) recently offered a Working Memory Model of deception, which is based on Baddeley's (1992,

2000) influential working memory theory. It too contends that lying imposes a greater load than truth telling due to its heavier processing requirements. Truth telling involves retrieving and reconstructing a memory. When lying, deceivers must invent new stories or modify those available from past experiences or scripts. A deceptive narrative must be plausible and not contradict itself or what the target knows. When no personal memories or scripts are available for lie construction, the working memories of liars will be heavily burdened, reducing capacity for speech production. Liars must also monitor listeners for signs of suspiciousness.

This model's most unique insights regard the information sources liars use to construct deceptive narratives. It also suggests that cognitive load indices can make reliable cues when examinees are surprised by test items probing details that are likely to be part of the memory of a truthful experience, but not a deceptive narrative.

THE ACTIVATION-DECISION-CONSTRUCTION MODEL OF ANSWERING DECEPTIVELY

The Activation-Decision-Construction Model (ADCM; Walczuk et al., 2003, 2005, 2009) describes answering questions deceptively, which theoretically includes the multiple-choice questions of the GKT. The model analyzes the act into three components. First, a question heard or read activates the truth from long-term memory, usually automatically. Second, based on the activated truth and social context, a decision to lie may be made, usually to advance liars' interests. Truthful answering will then be actively inhibited, especially for well practiced truths that can proactively interfere with lying. Such response competition is elegantly described by the PTS model. Third, a context-appropriate lie is constructed that must be coherent and plausible. When possible, memories of the truth are altered slightly for the sake of lie plausibility and to minimize the cognitive load of lie construction. Finally, a lie is shared.

Walczuk et al. (2009) expanded the ADCM to account for the rehearsal of deceptive answers. "Deciding to lie" becomes "remember to lie," with relevant questions and social contexts serving as the memory cues. "Lie construction" becomes "lie recall," followed by tweaking of the deceptive answers to fit the prevailing social context, both entailing lower loads than spontaneous lying. Responses to questions using the CQT are usually made in less than a second (Lykken, 1998). The expanded ADCM can easily account for this as follows. Either before the exam or during the preview of questions, a deceptive examinee will decide which questions she/he will lie to and will construct deceptive answers. Delivering them during an exam involves cued recall, which typically occurs automatically and quickly (Anderson, 2000).

Several elements of the ADCM have been supported. Walczuk et al. (2003) found, according to self-reports, when participants answered questions deceptively that the truth entered working memory automatically and interfered with lying, consistent with the activation and decision components. Walczuk et al. (2009) demonstrated that individuals lying about well practiced truths had the most difficulty due to a Stroop-like interference. In having participants answer questions about various aspects of their lives either deceptively or truthfully, Walczuk et al. (2005) showed that having to decide to lie adds to cognitive load, and constructing

a lie caused greater load than truth telling. One of the ADCM's implications for lie detection is that when the truth can be pre-activated in examinees and questions are asked that examinees do not anticipate, the processes of deciding to lie and lie construction will manifest as higher cognitive load in liars alone.

GENERAL EVALUATION OF THE MODELS/THEORIES

The range of models and theories above illustrates the multifariousness of deception. No single theory could account for all of its cognitive complexity. Generally, these accounts are most relevant to spontaneous (unrehearsed) lying. In such cases, the cues to deception tend to be the richest, including longer response times and more pupil dilation (DePaulo et al., 2003). To be relevant to load-inducing lie detection, they *must be expanded to account for rehearsed deception*, a likely countermeasure. For instance, Interpersonal Deception Theory holds that liars actively monitor their behavior and that of the targets. This may not apply to highly skilled or practiced liars. As suggested by the expanded ADCM, the memory processes of encoding and retrieval will be central to these expansions and become highly automated with practice (Anderson, 2000).

LIE DETECTION VIA INDUCING COGNITIVE LOAD

Recently an innovative general approach to lie detection has emerged: cognitive load-inducing techniques designed to elicit greater mental effort in liars than in truth tellers (Walczuk et al., 2005; Vrij et al., 2008a). Whereas polygraph-based questioning paradigms rely on elevation in physiological arousal to gauge deception, these use the heightening of indices of cognitive load as the primary cues. Another contrast, although surprising examinees with questions is discouraged in the CQT, given the high rate of false positives that can result (Lykken, 1998), surprising (not shocking) examinees with questions or the task used to access the truth is central to many load-inducing techniques to make it hard to lie. Some techniques below elicit brief responses, as do the CQT and GKT. Others elicit more open-ended responding, such as narratives.

The models and theories above can advance these techniques by showing when and why load indices provide reliable cues (Vrij et al., 2008a). Rather than reviewing all published variations on a common theme, generally only distinctive research-based proposals are discussed, along with their pitfalls and limitations. The results of the experiments testing them show that liars and truth tellers can be classified beyond chance. However, we do NOT discuss the rates of false positives or false negatives for these techniques, because it is far too early in their development to estimate such parameters accurately. This is especially true given that most research is based on college students, not suspects under police interrogation or other authentic samples. Thus, such estimates would be misleading.

TIME RESTRICTED INTEGRITY-CONFIRMATION

Walczuk et al. (2005) proposed a load-inducing technique called *Time Restricted Integrity-Confirmation* (TRI-Con). It is based explicitly on a theoretical account of the differences in mental states between liars and truth tellers, the ADCM. TRI-Con selectively enhances load of liars by surprising examinees with

unanticipated questions and by requiring quick responses. These specific guidelines apply to examinations (Walczuk et al., 2005, 2009). (a) Examinees are prompted about the focus of the question set to follow (e.g., “The next 11 questions concern your activities at the time of the crime”). By priming relevant episodic “truths,” prompts reduce examinees’ need to search memory to answer honestly, making cognitive load indices less ambiguous cues that show when a decision to lie and lie construction have occurred. Prompting also reduces the emotional surprise that might be caused by blindsiding examinees with questions probing sensitive issues or incriminating information. (b) Still, the specific questions are not disclosed until asked during an exam, thus surprising examinees cognitively and reducing the rehearsal of lies. (c) Questions are written when possible to be unclear regarding what truths are targeted until they are fully asked. This reduces further examinees’ chance of preparing lies. (d) To obtain clear assessment of the cognitive load needed to answer completely, questions are written to be answerable with one or a few words. (e) Examinees are instructed to answer as quickly as possible to limit further their opportunity to deceive. The high cognitive load of rapid responding to surprise questions may increase cue leakage in the form of voice pitch elevation, pupil dilation, reduced blinking, and long response times because of the limited opportunity for liars to self-monitor and control (Zuckerman et al., 1981; Buller and Burgoon, 1996; Burgoon and Buller, 2008) and may increase accidental blurring of the truth (Lane and Wegner, 1995). (f) Without adequate preparation, liars’ deceptive accounts should be incomplete. Questions are asked and then re-asked along with logically interrelated questions to increase liars’ cognitive load. Contradictions should occur with liars (Granham and Hartwig, 2008). (g) Behavioral baselines for ground-truth answers are established for all cognitive load indices for comparison with levels of these cues of answers suspected of deception. This practice controls for individual differences in behavioral base rates and improves the accuracy of lie detection (Walters, 1996; Bond and DePaulo, 2006).

Studies have shown the effectiveness of TRI-Con for uncovering deception. Following these guidelines, Walczuk et al. (2005) instructed adults to lie or tell the truth to questions about various aspects of their lives (e.g., employment history, performance on standardized tests). Using response time as the cue, discriminant analyses allowed classification of liars and truth tellers well above chance. Walczuk et al. (2009) tested TRI-Con again by asking participants to lie or tell the truth about their lives and included a rehearsal condition in which participants prepared deceptive answers. The consistency of answers across interrelated questions was added as a cue. Liars and truth tellers were classified up to 89% accurately. The analyses showed that rehearsed deception is detectable. Finally, Walczuk et al. (2012) tested TRI-Con in a forensically relevant context. “Witnesses” observed actual crime videos, then later told the truth or lied rehearsed or unrehearsed about them during interrogation. The cognitive cues were response time, answer consistency, eye movements, and pupil dilation. Discriminant analyses allowed classification of the three conditions 69% accurately, 33% expected by chance.

Despite these promising results, TRI-Con has limitations. For instance, extended narratives given by examinees provide valuable verbal cues to deception (Buller and Burgoon, 1996; Sporer and

Schwandt, 2007) that the short answers of TRI-Con are unlikely to tap. Moreover, pupil dilation, blinking rate, voice pitch elevation, and other reliable cues not only measure cognitive load but emotional responses as well (DePaulo et al., 2003). TRI-Con and the techniques to be described may elicit not only cognitive load but also anxiety in examinees. This fact is not problematic when it can be assumed that both anxiety and cognitive load co-vary with deception (Vrij et al., 2010b). Finally, TRI-Con does not allow participants to qualify their answers during the exam, unlike open-ended responses. However, these limitations can be overcome by combining diverse techniques, a possibility discussed later.

COUNTERMEASURES

After new methods of lie detection are introduced, information about them disseminates, and countermeasures are devised. This occurred with the polygraph (Lykken, 1998; National Research Council, 2003) and is occurring with cutting edge approaches, such as functional magnetic resonance imaging (Simpson, 2008; Ganis et al., 2011). Noting this, Walczuk et al. (2005, 2009, 2012) argued that a likely countermeasure against load-inducing lie detection is the rehearsal of a lie, a load reduction strategy (O’Hair et al., 1981; Greene et al., 1985). All research and theory in this area must consider rehearsal. For TRI-Con, other possible countermeasures include examinees intentionally not complying with instructions to answer quickly (e.g., ask that a question be repeated). Likely countermeasures against other load-inducing proposals are discussed as they are presented.

ASKING UNANTICIPATED QUESTIONS AND SOLICITING SURPRISE DRAWINGS

Asking questions that examinees do not expect may increase cognitive load. Vrij et al. (2009) instructed pairs of participants to lie or tell the truth about having had lunch together. All pairs then prepared for an interview, which included anticipating likely questions. General and unanticipated questions were later asked, the latter probing minor details like these. What color shirt was worn? Who arrived first? Who sat closest to the door? Inconsistencies in answers to such questions enabled observers to classify liars and truth tellers beyond chance, as did discrepancies in the surprise pictures that the pairs were asked to draw of the layout of the restaurant. Although investigators did not measure the cognitive loads elicited by surprising participants with unexpected questions or the drawing task, we regard both to be load-inducing techniques, because respondents likely had to think a lot when answering or drawing to ensure plausibility and consistently since responding to both was unrehearsed (DePaulo et al., 2003). Recently, Vrij et al. (2012b) observed that truth tellers’ drawings of their workplaces contained more plausible details, especially those involving their coworkers, than liars doing the same.

These results are encouraging. Still, asking unanticipated questions has limitations. Recall that once knowledge of this technique disseminates, liars may include spatial and other obscure details into their deceptive narratives in anticipation of such questions. Second, memory for minor details can easily go unnoticed by truth tellers (Loftus, 2007), making the response “I can’t remember.” plausible when given by liars. The same concerns hold for drawing pictures. Liars can practice drawing them in advance or

plausibly deny having noticed spatial details. Still, refinement of these techniques may overcome such concerns.

MAINTAINING EYE CONTACT WITH THE EXAMINER

Having to maintain eye contact with another can elevate cognitive load and anxiety in liars. In support, Vrij et al. (2010b) directed some participants to lie to interview questions; others told the truth. Some were further instructed to maintain continuous eye contact with the interviewer. Observers of videotapes of the interviews were better at discriminating liars from truth tellers when eye contact was maintained, suggesting that it imposed greater cognitive load and anxiety on liars.

One possible countermeasure is practicing lying while maintaining eye contact with another, which may lessen liar-truth teller differences. Also, sustaining eye contact might prove ineffective with Japanese and those of other non-Western cultures for whom this behavior goes against societal norms. It might induce inordinately high levels of anxiety and be distracting, even for truth tellers (McCarthy et al., 2006). Thus, it is unclear how effective this proposal can be as a general load-inducing technique for distinguishing liars and truth tellers.

RECOUNTING EVENTS IN REVERSE CHRONOLOGICAL ORDER

The temporal order in which events are recalled can magnify cues to deception. Vrij et al. (2008b) directed half their participants to lie and the other half to tell the truth about what happened during a staged event. Some participants of each condition were further instructed to report events in reverse chronological order. Others reported in chronological order only. More cues to deception emerged and were noticed by observers in the reverse order recounting. The authors noted that recalling in reverse order runs contrary to the typical forward chronological encoding of events and thus imposes a heavy load, especially for liars. Vrij et al. (2012a) extended this technique by asking individuals to lie or tell the truth about a route they took in chronological and in reverse chronological order. More cues to deception again emerged and were noted by observers in the reverse order retelling.

If liars practice lying in reverse chronological order will the cues to deception be as rich? Another likely countermeasure to cover their involvement in crimes, clever perpetrators might base their false alibis on episodic memories of actual events, altering details as needed (Sporer and Schwandt, 2007; Leins et al., 2012). The reverse chronological retelling of these liars might then be similar in cognitive load to that of truth tellers doing the same.

DUAL-TASKING (DOING TWO THINGS AT ONCE)

Asking examinees to perform a concurrent task during interrogation was a novel approach to load induction tested by Patterson (2010). If lying draws more on attention and working memory than truth telling, then a dual task might interfere more with the former. In this study, truth tellers followed written instructions to go to the university book store, perform specific tasks, and later honestly describe and answer questions about what they did. Liars were shown these instructions but prepared deceptive narratives as if they had been followed, which they later conveyed and answered questions about. During the interview phase, all participants had to perform a concurrent math task. Math response times and

accuracies were the dependent measures. Regarding the results, dual task interference was minimal. No liar-truth teller differences were found for math response times, but there was slightly higher math accuracy for truth tellers. Videos of selected interviews were later shown to observers. When interviewees were engaged in a secondary task, observers were slightly more accurate in assessing the veracity of responses and attributed higher loads to liars. This technique is innovative, and more research is needed. However, no theoretical rationale was given for the choice of concurrent task, which may partially explain the weak findings, a theme expanded on later.

OVERALL EVALUATION OF LOAD-INDUCING PROPOSALS

Our general impression of the load-inducing approaches to date is that they are innovative and promising. However, it is too soon in their development to accurately gauge their applicability to forensic settings and other real world contexts where detecting deception is vital. Once again, more research is needed on their susceptibility to rehearsal and other countermeasures and on whether the use of such countermeasure is detectable. Also, recall that most of the studies above involved college students who were offered extra credit in exchange for their participation. Their motivation to succeed in their lies was low compared to actual perpetrators trying to persuade detectives with their false alibis or innocent suspects attempting to convince detectives of their innocence. The cognitive loads of guilty liars and innocent truth tellers may both be so high that load-inducing interventions do not differential well between the two (Van Koppen, 2012; Vrij and Granagh, 2012a,b). Research testing these techniques on authentic samples is clearly needed.

A THEORETICAL TAXONOMY OF COGNITIVE LOAD-INDUCING LIE DETECTION

Sufficient promising results have been published on load-inducing lie detection (see Vrij et al., 2010a) to justify the proposal of a theoretical taxonomy that can help organize, direct, and advance future validation, refinement, and innovation. It is based on the important distinctions of the type of cognitive load each proposal induces, intrinsic versus extraneous, and how open-ended are the responses each permits in examinees, closed-ended (e.g., short answers, key strokes) versus open-ended (e.g., narratives, drawings).

Two key terms are first defined, both adapted from Cognitive Load Theory (Merriënboer and Sweller, 2005). “Intrinsic cognitive load” refers to the inherent demands on the cognitive resources of attention and working memory needed to lie well. Items 1 through 9 of **Table 1** convey some important factors adding to the intrinsic load of lying, organized by whether they relate to preparing a deceptive message or to delivering the message to a target. “Extraneous cognitive load” means any demands on or loss of cognitive resources due to tasks or factors external to the act of lying that makes it more difficult. For example, extreme anxiety in an examinee can decrease available cognitive resources, effectively imposing cognitive load (see Item 10 of **Table 1**).

The extent to which Items 1 through 10 of **Table 1** apply to an instance of lying depends on the complexity of its social context (DePaulo et al., 2004). Everyday lies are told without imposing

Table 1 | The cognitive load of lying versus truth telling.**SOME FACTORS ADDING TO THE COGNITIVE LOAD OF LYING*****Preparing a deceptive message**

1. Does formation of the lie require that details be kept internally consistent (no contradictory information, Granhag and Hartwig, 2008)?
2. Is the narrative externally consistent (congruent with what the target knows; DePaulo et al., 2003)?
3. Is the narrative detailed enough with multimodal info., a realistic timeline, etc. to convince the target (Vrij et al., 2010a)?
4. Beyond going undetected, are lies based on the deceptive narrative likely to achieve the liar's goal, for instance, obtaining money from a naïve target (Walczuk et al., 2012)?

Appearing sincere while delivering a deceptive message to the target

5. Is the motivation high to lie successfully (Vrij and Mann, 2001)?
6. Not taking credibility for granted, how much monitoring of and control over the self is the liar exercising to appear truthful and to stay in the deceptive role (Zuckerman et al., 1981; Buller and Burgoon, 1996; Vrij et al., 2010a)?
7. How much is the liar monitoring the target's behavior to, see if the lie is believed (Buller and Burgoon, 1996; Vrij et al., 2008a, 2010a)?
8. Is the truth deeply entrenched, does it elicit strong emotions, or is honest responding well practiced so that proactive interference with deceptive responding occurs (Lane and Wegner, 1995; Morgan et al., 2009; Osman et al., 2009; Walczuk et al., 2009)?
9. Is an adequate deceptive narrative unavailable or is the lie unrehearsed (Vrij et al., 2010a)?
10. Is the liar highly anxious (Eysenck, 1992; Beilock and Carr, 2005)?

SOME FACTORS ADDING TO THE COGNITIVE LOAD OF TRUTHTELLING*

11. Does recalling the truth to working memory require retrieving memories that have not been accessed in a long time or details that have decayed (Anderson, 2000; Wixted, 2004)?
12. Is a lie well rehearsed compared to its corresponding truth (O'Hair et al., 1981; Greene et al., 1985)?
13. Does a truthful response require elaboration or qualification to be accurately understood by the target compared to a corresponding lie (Gombos, 2006)?
14. Does a truthful response require the generation of a novel opinion, judgment, evaluation, attitude, or emotional reaction (DePaulo, 1992; DePaulo et al., 2003; Gombos, 2006)?
15. Is the truth teller highly motivated to be believed (Van Koppen, 2012; Vrij and Granhag, 2012a,b)?

*These lists are not exhaustive.

high cognitive loads. Liars typically have little concern about getting caught and rarely monitor their behavior or the targets' (DePaulo et al., 1996). Thus, few items apply. However, serious lies have greater interpersonal consequences and entail heavier loads (DePaulo et al., 2004; Burgoon and Buller, 2008). More items will apply, especially when lying is spontaneous. On the other hand, skilled or well rehearsed liars telling serious lies may not need to monitor their behavior or the targets' (Items 6 and 7), instead relying on their fluent delivery to carry them through (DePaulo, 1992).

For cognitive load-inducing lie detection to succeed, it is important to note when truth telling imposes a greater cognitive load than telling a corresponding lie. For instance, Walczuk et al. (2005) found that college students took longer to recall their actual standardized test scores than to lie about them. The bottom of **Table 1** lists five factors adding to the cognitive load of honesty. Only when they can be discounted during an examination is lying more likely to manifest in heightened load indices. For instance, questions asked in load-inducing lie detection exams need to be written with Items 11 through 15 in mind so that the cognitive load of lying is higher than for truth telling.

Table 2 provides the full Taxonomy of Load-Inducing Lie Detection and shows where the proposals above and others to be discussed fall within it. Despite the severe limitations of some of them, all proposals are included for the sake of comprehensiveness. A question that should guide their refinement is "*Under what testing conditions are cognitive load indices unambiguous cues*

to deception?" To illustrate, such a condition is when "prompting" occurs, which makes cognitive load indices clearer cues by reducing the need for all examinees to search memory for a truth and by reducing the emotional surprise to questions during the exam (Walczuk et al., 2005).

INTRINSIC COGNITIVE LOAD-INDUCING TECHNIQUES

The proposals under this heading seek to make the act of lying harder by surprising examinees cognitively with test items, with the memory task used to access the truth, or by requiring quick responses. TRI-Con, which elicits closed-ended responses, falls into this category. When examinees have not anticipated the questions, they must decide which ones to lie to and generate deceptive answers on the fly, all adding to cognitive load (Walczuk et al., 2005, 2009). Vrij et al.'s (2008b) proposal of having examinees convey narratives in reverse chronological order fits here, as does instructing them unexpectedly to draw pictures (Vrij et al., 2009). Memories related to the truth are being probed in unusual ways that liars may not have anticipated. Because narratives and drawings can be as elaborate as examinees choose, we consider them to be open-ended. Examinees can pace themselves, monitor, and control their behavior, hopefully causing related cues to emerge (DePaulo et al., 2003).

Although not specifically proposed as a load-inducing technique, Seymour et al. (2000) tested a variant of the GKT with response time as the cue to deception, the *Response Time GKT*, which qualifies as one. Participants partook in a mock crime

Table 2 | A taxonomy of load-inducing lie detection: the type of cognitive load induced and the response open-endedness permitted.

Intrinsic load	Extraneous load
Closed-ended responding (encourages the leakage of hard-to-control non-verbal cues and blurring of the truth)	
TRI-Con	Maintain eye contact
Answering unanticipated short-answer questions about minor details	Dual-tasking-articulatory suppression, n-back task
Implicit personality/attitude tests	Dual-tasking-operate driving simulator or do a concurrent math task
Autobiographical implicit association tests	Have examinee give short answers in front of a mirror
Response time GKT	
Open-ended responding (encourages verbal signs of deception, monitoring of self and the target, signs of the attempted control of behavior)	
Have examinee relate surprise narrative	Maintain eye contact
Surprise task of examinees drawing picture of alibi	Have examinees give narrative answers or draw pictures in front of a mirror
Recall events in reverse chronological order	Dual-tasking-operate driving simulator or do a concurrent math task
Recall from a different physical perspective	

involving a computer. They also learned two-word phrases related to the crime as well as other two-word phrases later in the experiment. During a subsequent phrase classification task, with instructions to respond *as quickly as possible*, participants were asked to press a key with their right index finger if an item was on the list that had been learned later. All other items required a key press with their left index finger, some of which were from the mock crime. The latter responses were the equivalent of having to conceal guilty knowledge. Responding to guilty knowledge items took about 300 ms. longer than responding to neutral items. Discriminant analyses correctly classified guilty and innocent trials 95% accurately.

Self-report measures of personality and attitudes (e.g., toward members of minority groups) are highly susceptible to deception as some examinees respond to test items to create a false positive image of themselves to obtain jobs and other rewards. Even so, their actual personalities and attitudes are rehearsed through repeated ways of thinking and acting in their daily lives that form strong associations among ideas and emotions in memory (Banse and Greenwald, 2007). Although not proposed as a load-inducing technique either, as an alternative to self-report measures, *Implicit Personality and Attitude Tests* qualify as well. They put examinees under time pressure in responding (Banse and Greenwald, 2007). This increases intrinsic cognitive load by making it harder to deceive. For dishonest examinees, a proactive interference can occur when their true attitudes and personalities conflict with the impressions they want to make, which usually manifests as slower response times for items lied to. Moreover, responses are typically closed-ended, limited to a forced choice between two options. How quickly individuals respond, for example, when instructed to associate the word “good” with the faces of individuals with dark complexions co-presented on a computer screen among many pairings of stimuli, can reveal racism in those responding slowly. This technique has been used successfully in employee selection (Banse and Greenwald, 2007). A variant of it, the *autobiographical Implicit Association Test* (aiAT), was proposed and examined by Sartori et al. (2008). It requires examinees to respond rapidly to test sentences presented one at a time on a computer screen describing autobiographical events that are either true or false for them. Across six experiments, aiAT had accuracies up

to 91% in revealing concealed knowledge of true autobiographic events. Still, when examinees use the countermeasure of strategically slowing down when responding truthfully, this accuracy drops dramatically (Verschuere et al., 2009).

We now propose another way of accessing truths that could impose higher intrinsic load on liars and may inspire the development of similar proposals by others. It is based on the encoding-specificity hypothesis (see Anderson, 2000) and part of the “Cognitive Interview,” a well validated set of four memory strategies for assisting individuals in recalling accurately and fully prior events, without inducing memory distortions (Fisher and Geiselman, 1992; Geiselman and Fisher, 1997). First, the interviewer tries to reinstate the physical and mental state of the witnessed event, for instance, by asking the interviewee to form a mental picture of the context of the event and recall how she/he felt. Second, the interviewee is encouraged to recall every detail of an event she/he can, even seemingly insignificant memory fragments. By the third principle, the interviewee is encouraged to recall in a variety of temporal orders. Recall that a variant of this principle was applied to lie detection by Vrij et al. (2008b), specifically recounting in reverse chronological order. Fourth, the interviewee is encouraged to recall from a variety of physical locations, for instance, from how things would have appeared “if you had been looking down on the room where the crime occurred from directly above” or “if you had looked at the room from the perpetrator’s perspective.” The latter principle, too, might be applied to lie detection. If asked to recall from different perspectives, truth tellers should have narratives richer in realistic details that are delivered with fewer hesitations than liars (Sporer and Schwandt, 2007).

EXTRANEOUS COGNITIVE LOAD-INDUCING TECHNIQUES

Techniques under this heading seek to induce cognitive load selectively on liars, not by making it harder to lie, but by altering other aspects of the examination procedure or context. “Dual-tasking” was one such proposal considered earlier that is now discussed more deeply. Cognitive scientists have long used this research paradigm to determine when different tasks use a common system or pool of resources (Pashler, 1994; Baddeley, 1996). As a technique for lie detection, it can be used with test items soliciting closed-ended or open-ended responses (Patterson, 2010). Vrij et al.

(2008a) suggested that examinees could “recall their stories whilst conducting a computer driving simulation task at the same time” (p. 41). If deception imposes greater load, then the simulation may interfere more with liars, enhancing cognitive cues. To our knowledge, this study has not been done, but is worth testing.

Meyer and Kieras (1997) evaluated various theoretical accounts of multi-task interference. Of them, Unitary Resource Theory is the one that Patterson (2010) and Vrij et al. (2008a) implicitly subscribe to with their proposals. Its basic assumptions are that (a) attentional capacity is a limited general resource that can be assigned to multiple tasks. (b) The amount of attention allocated depends on the demands of the current activities. (c) Under low levels of task load, attention can easily be divided between tasks, not so when either or both of the tasks are difficult. (d) Finally, attention is controllable and can be allocated dynamically. Meyer and Kieras (1997) also review the major criticisms of this account. The one that is most problematic for the proposals above concerns “difficulty insensitivity.” Varying the difficulty of a primary task often does not interfere with a concurrent task, which should occur if both are dependent on a central, limited resource. For instance, difficulty insensitivity was apparently the case with Patterson (2010), who found that lying was minimally disruptive of a concurrent math task.

A powerful framework for understanding multi-task interference effects, which we embrace, is *Adaptive Executive Control* (AEC; Meyer and Kieras, 1997; Meyer et al., 2002). It overcomes the criticisms of Unitary Resource Theory, is instructive regarding what concurrent tasks theoretically should interfere with lying more than truth telling, and is well supported. Five components underlie the framework. (a) It is based on a comprehensive information processing architecture that incorporates all of the known characteristics of human cognition. (b) It is also based on a production system formalism that expresses actions as If-Then rules, which succinctly capture procedural knowledge. The “If” part specifies the conditions under which actions are executed. The “Then” portion specifies the actions in their proper order. (c) Importantly, no assumption of a limited general cognitive resource or capacity is made. (d) Rather, AEC attributes dual task interference to the flexible strategies individuals adopt to fulfill their task priorities as handled by supervisory executive processes. In effect, one task is put on hold while another task higher in priority takes precedence and executes. (e) Finally, AEC explicitly takes account of the constraints in processing imposed by perceptual and motor systems during multi-task performance. For instance, concurrent tasks both requiring verbal responses will naturally interfere. A higher priority utterance will precede the lower priority utterance. Those interested in AEC are referred to Meyer and Kieras (1997) as well as Meyer et al. (2002). To summarize, the major sources of interference are competition between concurrent tasks for the same perceptual or motor response systems or the executive process performing one task before another due to its higher priority given the performer’s goals. If the AEC framework is valid, then the dual-tasking, load-inducing proposals above are unlikely to be effective, because competition for a limited central resource, as they intend, is not the basis of interference.

We now propose a potentially interfering task suggested by the Working Memory Model of deception (Sporer and Schwandt,

2006, 2007). It does not assume a competition for limited attention. Rather, it prevents liars from using a specialized working memory store needed for lying. In research on working memory’s phonological loop, *articulatory suppression* prevents the rehearsal of memory items (Baddeley, 1996), for instance, by instructing participants to continuously repeat a simple word such as “one.” However, repeating a single, familiar syllable might quickly become automated and be minimally disruptive of lying. Having to repeat a sequence of unfamiliar syllables, such as “Bah-Bay-Boo-Bee,” would lessen this problem (Gordon and Meyer, 1987). Will continuously repeating such a sequence interfere more with lying? To test whether, a study can be conducted in which recorded questions are asked through headphones and answers, yes or no, are given non-verbally as keystrokes so as not to impair articulatory suppression. Thus, responding is closed-ended. According to the Working Memory Model, lying requires more access than truth telling to an unencumbered phonological loop for language production. If so, then when articulatory suppression is added, lying should entail longer response times, more pupil dilation, and less blinking than truth telling due to interference caused by competition for this specialized working memory store. Another theoretically based dual task was tested by Ambach et al. (2011) with the GKT: the n-back procedure (deciding whether a stimulus was presented n trials previously). Both tasks were hypothesized to compete for working memory’s central executive. The n-back task enhanced the detection of concealed knowledge as measured by electrodermal activity. Researchers are encouraged to follow these two examples and develop other theoretically based dual task proposals.

Requiring examinees to hold eye contact with the examiner can impose extraneous cognitive load in liars, perhaps by provoking anxiety and an allocation of cognitive resources to monitor the self and the target (Vrij et al., 2010b). As noted previously, this load-inducing technique may not work well with members of some cultures and possibly other segments of the general population for whom maintaining eye contact goes against a social norm (McCarthy et al., 2006). When it is appropriate, it can be used with closed- and open-ended responses. Another way to induce extraneous cognitive load on liars might be to have examinees answer questions while sitting in front of a mirror, which could increase their self-monitoring and the emergence of related cues (Buller and Burgoon, 1996). An examiner would still be needed in the room. This proposal has not been tested.

COMBINING INTRINSIC AND EXTRANEous TECHNIQUES; OPEN- AND CLOSED-ENDeD RESPONDING

Intrinsic and extraneous load-inducing techniques can be combined in an exam, thereby gaining the advantages of each. For example, examinees can be tested under the intrinsic load-inducing conditions of TRI-Con while following instructions to maintain eye contact with the interrogator (Vrij et al., 2010b). Unanticipated questions about spatial information and other details can be asked. Examinees can also be asked to draw pictures of the physical layout (Vrij et al., 2009).

Closed- and open-ended items can also complement each other (Toris and DePaulo, 1984). Recall that TRI-Con is intended to assess the truthfulness of the short answers given to closed-ended questions, which allows unambiguous assessment of the cognitive

load needed to answer using response time, pupil dilation, and other load indices. Moreover, the time pressure on responding may increase leakage of non-verbal cues, which manifest primarily under high cognitive load (Buller and Burgoon, 1996; Burgoon and Buller, 2008) and may increase the chance that a truth is inadvertently blurred (Lane and Wegner, 1995). On the other hand, open-ended questions eliciting narratives can be rich in verbal cues such as vagueness and dissociations (Buller and Burgoon, 1996; Burgoon and Buller, 2008) and in signs of the attempted control of behavior by liars (DePaulo et al., 2003; Sporer and Schwandt, 2006; Vrij et al., 2010a). Interrogators might, for example, have a suspect provide a narrative of an alibis at the time of the crime, followed by a TRI-Con exam with questions probing details of an alibi. If the verbal and control cues from the open-ended portion and the cognitive load indices of the closed-ended questioning all point to deception, strong converging evidence will exist. Also, the more reliable cues to deception used, the more accurate its detection tends to be (DePaulo et al., 2003). It may be worthwhile to combine psycho-physiological and cognitive load cues to enhance lie detection.

ADVANCING COGNITIVE LIE DETECTION BY AVOIDING THE PITFALLS OF THE CQT

Four recommendations for researchers and practitioners and their justifications appear below to help the emerging field of cognitive lie detection to avert four of the major weaknesses of the CQT (National Research Council, 2003) and profit from the strengths of the GKT. To recall the criticisms, the CQT is not based on a validated theory and is easily susceptible to countermeasures. Its administration has not been standardized, and the scoring of results is largely subjective.

1. Cognitive load-inducing lie detection techniques should be based on explicitly stated, well-specified, and validated cognitive models of deception (see McCornack, 1997). To many readers this recommendation may be obvious. However, *to date, few load-inducing techniques are based on models or theories that were made explicit in the research reports*. Perhaps the models were implicit in some cases, but this is not helpful to readers wishing to understand the reasons why load-inducing manipulations work or when experimental findings apply to authentic settings. In fact, historically many researchers have sought out reliable cues to deception with minimal regard for their basis in theory (DePaulo et al., 2003). This risks repeating this mistake of the CQT with load-inducing lie detection. Recall that refined cognitive models, supported by data, can illuminate the conditions when load-inducing interventions are likely to succeed, necessary for generalizing the results of experiments to the field (Vrij et al., 2008a). Since deception is multifarious (e.g., verbal, non-verbal, lies of omission) and is driven by many motives (e.g., protect another, conceal wrongdoing, exploit others; DePaulo et al., 2004), no single cognitive account can explain all of its forms (Ekman, 2001; DePaulo et al., 2003).

The accounts we reviewed can serve as building blocks for narrowly focused models directly applicable to specific authentic contexts like the interrogation room of a police department. To recap those we regard as most applicable, the PTS model

specifies the response competition that occurs when examinees falsely deny possessing guilty knowledge. Response competition also likely underlies much of lying, especially when deceiving about well practiced truths. Interpersonal Deception Theory (Buller and Burgoon, 1996; Burgoon and Buller, 2008) highlights the cognitive load of having to monitor the behavior of the self and the target and postulates the leakage of cues under high cognitive load. DePaulo's (1992) Self-Presentation Theory posits the leakage of cues and delineates between non-verbal behaviors that are easy to control versus those that are not, the latter providing the best cues. Sporer and Schwandt (2006, 2007)'s Working Memory Model elaborates on the sources of information (e.g., scripts, personal memories) used in lie construction. The ADCM is informative about the encoding and retrieval processes related to truth telling, the decision to lie, and lie construction and offers insights regarding the rehearsal of deception (Walczuk et al., 2009, 2012). The Preoccupation Model of Secrecy advances understanding of why cognitive resources are often needed to inhibit truthful responding. Still, these cognitive accounts must be expanded to address the motivation to lie, rehearsal, and other important moderators of cues to deception to be maximally relevant to lie detection (DePaulo et al., 2003).

2. Countermeasures will be devised by deceptive examinees for beating any new method of lie detection as knowledge of it disseminates, nor can this be entirely prevented (Lykken, 1998; National Research Council, 2003; Rosenfeld et al., 2004; Simpson, 2008; Verschueren et al., 2009; Ganis et al., 2011). Researchers and practitioners concerned with cognitive load-inducing lie detection should note that the rehearsal of deception is a serious countermeasure, continue to find ways to minimize it, as well as ways to expose rehearsal when it occurs. Other countermeasures need to be uncovered as well. If this recommendation too seems obvious, it is noteworthy that *few studies testing load-inducing techniques have seriously considered rehearsal or included a rehearsal condition in the research*. We discussed several load-inducing proposals to minimize it such as surprising examinees with test items, with memory tasks, or having examinees respond quickly. Still, rehearsal cannot be prevented completely. Constructing deceptive narratives before an exam as well as anticipating questions and preparing deceptive answers are likely in intelligent, motivated liars (Vrij and Mann, 2001; Vrij et al., 2010a).

The countermeasure of the rehearsal of deception can be overcome if research can identify "behavioral signatures" that it occurred. More research is needed on the effects of this countermeasure on pupil dilation, voice pitch, response time, blinking rate, and other correlates of cognitive load. Some encouraging findings are that rehearsed liars can have response times falling below those of unrehearsed liars and truth tellers (O'Hair et al., 1981; Greene et al., 1985), as well as reduced eye movements so they can focus on memory retrieval undistracted by the visual environment (Walczuk et al., 2012). Rehearsed liars also have brain activation patterns distinguishable from those of unrehearsed liars and truth tellers (Ganis et al., 2003). The more distinguishing cues that can be developed, the more accurately rehearsed deception can be exposed (DePaulo et al.,

- 2003). Research on the effects of other countermeasures, such as intentionally not complying with instructions to maintain eye contact or answer quickly (Verschueren et al., 2009), may reveal distinctive cognitive-behavioral signatures too.
- Research shows that even law enforcement officers, among other human observers, generally make poor lie detectors (Ekman and O'Sullivan, 1991; Garrido et al., 2004; Bond and DePaulo, 2006, 2008). This is partly because people tend to focus on unreliable cues like gaze aversion or nervousness and miss genuine cues that are often subtle (DePaulo et al., 2003). From police detectives interviewing witnesses to federal agents interrogating suspected terrorists, human lie detectors will be relevant for the foreseeable future. Most of the studies on inducing cognitive load we reviewed have wisely sought to improve the accuracy of human observers, but still report detection rates that are rather low (Vrij et al., 2010a).
 - Cognitive load-inducing lie detection involving closed-ended responding offers an alternative to both the CQT and the use of human lie detectors. Recall that administration of the CQT hasn't been standardized. To elaborate, most polygraph testing procedures allow for uncontrolled variation in test administration (e.g., creation of the emotional climate, selecting questions) that can be expected to result in variations in accuracy and limit the level of accuracy that can be consistently achieved (National Research Council, 2003, p. 213).
 - The guidelines of TRI-Con and those that might result from refinement of other closed-ended, load-inducing proposals, such as the aiAT and the Response Time GKT, can help standardize lie detection. This does not mean imposing a "predictability" that liars can rely on to foil exams. For instance, unanticipated questions can still be asked. Rather, standardization means following procedures that disambiguate cognitive load indices as cues to deception. TRI-Con, for instance, can be implemented on a laptop computer. Reading and computer skills are not required. Examinees wear a microphone-headset connected to a computer. Questions can be digitally recorded in advance. The assessment of answer response times, pupil dilation, voice pitch elevation, and other cues can be automated with technology now available (Walczuk et al., 2012). Accordingly, we recommend that lie detection exams be developed or further refined to follow standardized cognitive load-inducing procedures and that they be as automated as possible to sidestep the severe limitations of human lie detectors. Human examiners should still be present to oversee administration. Their presence can also induce load selectively on liars as deceptive examinees may feel compelled to monitor them for signs of suspiciousness (Burgoon and Buller, 2008).
 - Appropriate automated analytical procedures should be used to determine the "deceptiveness" or "honesty" of answers. A benefit of standardizing and automating lie detection and assessing the changes in cognitive load between known truthful answers and those suspected of deception is that statistical or other analytic procedures can be used to "decide" if examinees are lying or truth telling. This avoids the subjective scoring of the polygraph (Iacono and Lykken, 1997; Lykken, 1998; National Research Council, 2003). Of course, data must be collected on authentic

samples of liars and truth tellers who are highly motivated to convince authorities.

CONCLUSION

The polygraph-based CQT lacks a strong scientific basis (Iacono and Lykken, 1997; Lykken, 1998; National Research Council, 2003), unlike the more successful GKT. We believe that cognitive load-inducing techniques are promising alternatives to the CQT, especially if lessons can be learned from the latter. To help them advance, we reviewed many models and theories relevant to the cognition of deception, their implications for lie detection, and evaluated specific proposals for selectively inducing cognitive load on liars, particularly their susceptibility to rehearsal and other countermeasures. The taxonomy proposed classifies these proposals according to the type of cognitive load induced and the breadth of the responses permitted, which may help organize and advance the field by opening up new research areas. Along these lines, new proposals were also suggested. Finally, four recommendations were shared to assist this promising general approach in averting the corresponding pitfalls of the CQT. To date, researchers in this area often have not heeded warning implicit in the report of the National Research Council (2003).

Another contribution of this article is its modest attempt to merge the seemingly disparate fields of "polygraph-based lie detection" on the one hand and "social-cognitive perspectives on deception" on the other. As we have argued, the former has a long history (Lykken, 1998) with many valuable lessons for understanding deception and advancing lie detection. Scholars from each perspective are encouraged to consider research and theory from the other perspective for useful insights and opportunities for cross-fertilization.

Finally, there are several obstacles to societal acceptance of new forensic tools like these cognitive load-inducing lie detection proposals. Will judges, lawyers, solicitors, police officers, victims, suspects, or witnesses accept them? Recalling events in reverse chronological order, drawing pictures, the guidelines of TRI-Con, maintaining eye contact, or answering questions while performing a concurrent task might be rejected for lacking face validity. The most formidable obstacles come with legal tests like the US Supreme Court's 1993 decision of "Daubert v. Merrill Dow Pharmaceuticals" on the admissibility of scientific evidence in the courtroom. A new method that gives rise to such evidence must (a) have been empirically tested within the applicable field as described in publications in peer-reviewed outlets, (b) have a known potential error rate, (c) have established standards and safeguards for its use, and (d) be widely accepted within the scientific community (Solomon and Hackett, 1996). Much refinement and validation will be necessary to meet these standards. If load-inducing techniques can do so, their acceptance by stakeholders will likely follow.

ACKNOWLEDGMENTS

This research was funded by grant #648375 from the National Science Foundation. Any opinions, findings, conclusions, or recommendations expressed are those of the authors, not NSF. The authors would like to express their gratitude to the reviewers for the comments that greatly improved this manuscript.

REFERENCES

- Ambach, W., Stark, R., and Dieter, V. (2011). An interfering n-back task facilitates the detection of concealed information with EDA but impedes it with cardiopulmonary physiology. *Int. J. Psychophysiol.* 80, 217–226.
- Anderson, J. R. (2000). *Learning and Memory*. New York: Wiley and Sons.
- Baddeley, A. (1996). Exploring the central executive. *Q. J. Exp. Psychol.* 49, 5–28.
- Baddeley, A. D. (1992). Working memory. *Science* 255, 556–559.
- Baddeley, A. D. (2000). The episodic buffer: a new phase of working memory? *Trends Cogn. Sci. (Regul. Ed.)* 4, 417–423.
- Banse, R., and Greenwald, A. G. (2007). Personality and implicit social cognition research: past, present, and future. *Eur. J. Pers.* 21, 371–382.
- Beilock, S. L., and Carr, T. H. (2005). When high-powered people fail: working memory and “choking under pressure” in math. *Psychol. Sci.* 16, 101–105.
- Bond, C. F. Jr., and DePaulo, B. M. (2006). Accuracy of deception judgments. *Pers. Soc. Psychol. Rev.* 10, 214–234.
- Bond, C. F. Jr., and DePaulo, B. M. (2008). Individual differences in judging deception: accuracy and bias. *Psychol. Bull.* 134, 477–492.
- Buller, D. B., and Burgoon, J. K. (1996). Interpersonal deception theory. *Commun. Theory* 6, 203–242.
- Burgoon, J. K., and Buller, D. B. (2008). “Interpersonal deception theory,” in *Engaging Theories in Interpersonal Communication: Multiple Perspectives*, eds L. A. Baxter and D. O. Braithwaite (Thousand Oaks, CA: Sage Publications), 227–239.
- Carmel, D., Dayan, E., Naveh, A., Raveh, O., and Ben-Shakhar, G. (2003). Estimating the validity of the guilty knowledge test from simulated experiments: the external validity of mock crime studies. *J. Exp. Psychol. Appl.* 9, 261–269.
- DePaulo, B. M. (1992). Nonverbal behavior and self-presentation. *Psychol. Bull.* 111, 203–243.
- DePaulo, B. M., Ansfield, M. E., Kirkendol, S. E., and Boden, J. M. (2004). Serious lies. *Basic Appl. Soc. Psych.* 26, 147–167.
- DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., and Epstein, J. A. (1996). Lying in everyday life. *J. Pers. Soc. Psychol.* 70, 979–995.
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., and Harris, C. (2003). Cues to deception. *Psychol. Bull.* 129, 74–112.
- Dionisio, D. P., Granholm, E., Hillix, W. A., and Perrine, W. F. (2001). Differentiation of deception using pupillary response as an index of cognitive processing. *Psychophysiology* 38, 205–211.
- Ekman, P. (2001). *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage*. New York: Norton.
- Ekman, P., and O’Sullivan, M. (1991). Who can catch a liar? *Am. Psychol.* 46, 913–920.
- Elaad, E. (1990). Detection of guilty knowledge in real-life criminal investigations. *J. Appl. Psychol.* 75, 521–529.
- Eysenck, M. W. (1992). *Anxiety: The Cognitive Perspective*. Hove: Erlbaum.
- Fisher, R. P., and Geiselman, R. E. (1992). *Memory-Enhancing Techniques for Investigative Interviewing: The Cognitive Interview*. Springfield, IL: Thomas.
- Ganis, G., Kosslyn, S. M., Stose, S., Thompson, W. L., and Yurgelun-Todd, D. A. (2003). Neural correlates of different types of deception: an fMRI investigation. *Cereb. Cortex* 13, 830–836.
- Ganis, G., Rosenfeld, J. P., Meixner, J., Kievit, R. A., and Schendan, H. E. (2011). Lying in the scanner: covert countermeasures disrupt deception detection by functional magnetic resonance imaging. *Neuroimage* 55, 312–319.
- Garrido, E., Masip, J., and Herrero, C. (2004). Police officers’ credibility judgments: accuracy and estimated ability. *Int. J. Psychol.* 39, 254–275.
- Geiselman, R. E., and Fisher, R. P. (1997). “Ten years of cognitive interviewing,” in *Intersections in Basic and Applied Memory Research*, eds J. W. Payne and J. S. Carroll (Mahwah, NJ: Lawrence Erlbaum), 91–126.
- Gombos, V. A. (2006). The cognition of deception: the role of executive processing in producing lies. *Genet. Soc. Gen. Psychol. Monogr.* 132, 197–214.
- Gordon, P. C., and Meyer, D. E. (1987). Control of serial order in rapidly spoken syllable sequences. *J. Mem. Lang.* 26, 300–321.
- Granhang, P. A., and Hartwig, M. (2008). A new theoretical perspective on deception detection: on the psychology of instrumental mind-reading. *Psychol. Crime Law* 14, 189–200.
- Greene, J. O., O’Hair, H. D., Cody, J. J., and Yen, C. (1985). Planning and control of behaviors during deception. *Hum. Commun. Res.* 11, 335–364.
- Iacono, W. G., and Lykken, D. T. (1997). The validity of the lie detector: two surveys of scientific opinion. *J. Appl. Psychol.* 82, 426–433.
- Johnson, R., Barnhardt, J., and Zhu, J. (2004). The contribution of executive processes to deceptive responding. *Neuropsychologia* 42, 878–901.
- Kozel, F. A., Padgett, T. M., and George, M. S. (2004). A replication study of neural correlates of deception. *Behav. Neurosci.* 118, 252–256.
- Lane, J. D., and Wegner, D. M. (1995). The cognitive consequences of secrecy. *J. Pers. Soc. Psychol.* 69, 237–253.
- Leins, D. A., Fisher, R. P., and Ross, S. J. (2012). Exploring liars’ strategies for creating deceptive reports. *Legal Criminol. Psych.* doi:10.1111/j.2044-8333.2011.02041.x
- Loftus, E. F. (2007). “Memory distortions: problems solved and unsolved,” in *Do Justice and Let the Sky Fall: Elizabeth Loftus and her Contributions to Science, Law, and Academic Freedom*, eds M. Garry and H. Hayne (Mahwah, NJ: Lawrence Erlbaum Associates Publishers), 1–14.
- Lykken, D. T. (1998). *A Tremor in the Blood: Uses and Abuses of the Lie Detector*. New York: McGraw-Hill.
- McCarthy, A., Lee, K., Itakura, S., and Muir, D. W. (2006). Cultural display rules drive eye gaze during thinking. *J. Cross Cult. Psychol.* 37, 717–722.
- McCornack, S. A. (1997). “The generation of deceptive messages: laying the groundwork for a viable theory of interpersonal deception,” in *Message Production: Advances in Communication Theory*, ed. J. O. Greene (Mahwah, NJ: Lawrence Erlbaum), 91–126.
- Merriënboer, J., and Sweller, J. (2005). Cognitive load theory and complex learning: recent developments and future directions. *Educ. Psychol. Rev.* 17, 147–177.
- Meyer, D. E., and Kieras, D. E. (1997). A computational theory of executive processes and multiple task performance: I. Basic mechanisms. *Psychol. Rev.* 104, 3–65.
- Meyer, D. E., Kieras, D. E., Lauber, E., Schumacher, E. H., Glass, J., Zurnbrüggen, E., et al. (2002). “Adaptive executive control: flexible multiple-task performance without pervasive immutable response-selection bottlenecks,” in *Cognitive Modeling*, eds T. A. Polk, A. Thad, and C. M. Seifert (Cambridge, MA: MIT Press), 101–128.
- Morgan, C. J., SeSage, J. B., and Kosslyn, S. M. (2009). Types of deception revealed by individual differences in cognitive abilities. *Soc. Neurosci.* 4, 554–569.
- National Research Council. (2003). *The Polygraph and Lie Detection. Committee to Review the Scientific Evidence on the Polygraph*. Washington, DC: The National Academies Press.
- O’Hair, H. D., Cody, M. J., and McLaughlin, M. L. (1981). Prepared lies, spontaneous lies, Machiavellianism, and nonverbal communication. *Hum. Commun. Res.* 7, 325–339.
- Osman, M., Channon, S., and Fitzpatrick, S. (2009). Does the truth interfere with our ability to deceive? *Psychon. Bull. Rev.* 16, 901–906.
- Pashler, H. (1994). Dual-task interference in simple tasks: data and theory. *Psychol. Bull.* 116, 220–244.
- Patterson, T. (2010). The effect of cognitive load on deception. *Dissertation Abstracts International: Section B: The Sciences and Engineering* 71, 1364.
- Pennebaker, J. W., and Chew, C. H. (1985). Behavioral inhibition and electodermal activity during deception. *J. Pers. Soc. Psychol.* 49, 1427–1433.
- Rosenfeld, J. P., Biroschak, J. R., and Furedy, J. J. (2006). P300-based detection of concealed autobiographical versus incidentally acquired information in target and non-target paradigms. *Int. J. Psychophysiol.* 60, 251–259.
- Rosenfeld, J. P., Cantwell, B., Nasman, V. T., Wojdac, V., Ivanova, S., and Mazzari, L. (1988). A modified, event-related potential-based guilty knowledge test. *Int. J. Neurosci.* 24, 157–161.
- Rosenfeld, J. P., Soskins, M., Bosh, G., and Ryan, A. (2004). Simple, effective countermeasures to P300-based tests of detection of concealed knowledge. *Psychophysiology* 41, 205–219.
- Sartori, G., Agosta, S., Zogmaister, C., Ferrara, S. D., and Castiello, U. (2008). How to accurately detect autobiographical events. *Psychol. Sci.* 19, 772–780.
- Seymour, T. L. (2001). A EPIC model of the ‘guilty knowledge effect’: strategic and automatic processes in recognition. *Dissertation Abstracts International: Section B: The Sciences & Engineering* 61, 5591.
- Seymour, T. L., and Fraynt, B. R. (2009). Time and encoding effects in the concealed knowledge test. *Appl. Psychophysiol. Biofeedback* 34, 177–187.
- Seymour, T. L., and Kerlin, J. R. (2008). Successful detection of verbal and visual concealed knowledge using

- an RT-based paradigm. *Appl. Cogn. Psychol.* 22, 475–490.
- Seymour, T. L., Seifert, C. M., Shafto, M. G., and Mosmann, A. L. (2000). Using response time to measures to assess “guilty knowledge.” *J. Appl. Psychol.* 85, 30–37.
- Simpson, J. R. (2008). Functional MRI lie detection: too good to be true? *J. Am. Acad. Psychiatry Law* 36, 491–498.
- Sokolov, E. N. (1963). *Perception and the Conditioned Reflex*. New York: Macmillan.
- Solomon, S. M., and Hackett, E. J. (1996). Setting boundaries between science and law: lessons from *Daubert v. Merrell Dow Pharmaceuticals, Inc. Sci. Technol. Hum. Values* 21, 131–156.
- Sporer, S. L., and Schwandt, B. (2006). Paraverbal indicators of deception: a meta-analytic synthesis. *Appl. Cogn. Psychol.* 20, 421–446.
- Sporer, S. L., and Schwandt, B. (2007). Moderators of nonverbal indicators of deception: a meta-analytic synthesis. *Psychol. Public Policy Law* 13, 1–34.
- Toris, C., and DePaulo, B. M. (1984). Effects of actual deception and suspiciousness of deception on interpersonal personal perceptions. *J. Pers. Soc. Psychol.* 47, 1063–1073.
- Van Koppen, P. J. (2012). Deception detection in police interrogations: closing in on the context of criminal investigations. *J. Appl. Res. Mem. Cogn.* 1, 124–125.
- Verschueren, B., Meijer, E., and De Clercq, A. (2010). Concealed information under stress: a test of the orienting theory in real-life police interrogations. *Legal Criminol. Psych.* 16, 348–356.
- Verschueren, B., Prati, V., and De Houwer, J. (2009). Cheating the lie detector: faking in the autobiographical implicit association test. *Psychol. Sci.* 20, 410–413.
- Vrij, A., Fisher, R., Mann, S., and Leal, S. (2008a). A cognitive load approach to lie detection. *J. Investig. Psychol. Offender Profil.* 5, 39–43.
- Vrij, A., Fisher, R., Mann, S., and Leal, S., Milne, B., and Bull, R. (2008b). Increasing cognitive load to facilitate lie detection: the benefit of recalling an event in reverse order. *Law Hum. Behav.* 32, 253–265.
- Vrij, A., and Granhag, P. A. (2012a). Eliciting cues to deception and truth: what matters are the questions asked. *J. Appl. Res. Mem. Cogn.* 1, 119–117.
- Vrij, A., and Granhag, P. A. (2012b). The sound of critics: new tunes, old tunes, and resistance to play. *J. Appl. Res. Mem. Cogn.* 1, 139–143.
- Vrij, A., Granhag, P. A., and Porter, S. (2010a). Pitfalls and opportunities in nonverbal and verbal lie detection. *Psychol. Sci. Public Interest* 11, 89–121.
- Vrij, A., Mann, S., Leal, S., and Fisher, R. (2010b). “Look into my eyes”: can an instruction to maintain eye contact facilitate lie detection? *Psychol. Crime Law* 16, 327–348.
- Vrij, A., Leal, S., Granhag, P. A., Mann, S., Fisher, R. P., Hillman, J., et al. (2009). Outsmarting the liars: the benefits of asking unanticipated questions. *Law Hum. Behav.* 33, 159–166.
- Vrij, A., Leal, S., Mann, S., and Fisher, R. (2012a). Imposing cognitive load to elicit cues to deceit: inducing the reverse order technique naturally. *Psychol. Crime Law* 18, 579–594.
- Vrij, A., Mann, S., Leal, S., and Fisher, R. (2012b). Is anyone there? Drawings as a tool to detect deceit in occupation interviews. *Psychol. Crime Law* 18, 377–388.
- Vrij, A., and Mann, S. (2001). Telling and detecting lies in a high-stakes situation: the case of a convicted murderer. *Appl. Cogn. Psychol.* 15, 187–203.
- Walczuk, J. J., Mahoney, K. T., Doverpike, D., and Griffith-Ross, D. A. (2009). Cognitive lie detection: response time and consistency of answers as cues to deception. *J. Bus. Psychol.* 24, 33–49.
- Walczuk, J. J., Roper, K., Seemann, E., and Humphrey, A. M. (2003). Cognitive mechanisms underlying lying to questions: response time as a cue to deception. *Appl. Cogn. Psychol.* 17, 755–774.
- Walczuk, J. J., Schwartz, J. P., Clifton, R., Adams, B., Wei, M., and Zha, P. (2005). Lying person to person about life events: a cognitive framework for lie detection. *Pers. Psychol.* 58, 141–170.
- Walczuk, J. J., Griffith, D. A., Yates, R., Visconte, S. R., Simoneaux, B., and Harris, L. L. (2012). Lie detection by inducing cognitive load: eye movements and other cues to the false answers of false answers of “witnesses” to crimes. *Crim. Justice Behav.* 39, 887–909.
- Walters, S. B. (1996). *Principles of Kinesic Interview and Interrogation*. New York: CRC Press.
- Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annu. Rev. Psychol.* 55, 235–269.
- Zuckerman, M., DePaulo, B. M., and Rosenthal, R. (1981). “Verbal and nonverbal communication of deception,” in *Advances in Experimental Social Psychology*, Vol. 14, ed. L. Berkowitz (New York: Academic Press), 1–59.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 28 July 2012; accepted: 08 January 2013; published online: 01 February 2013.

Citation: Walczuk JJ, Igou FP, Dixon AP and Tcholakian T (2013) Advancing lie detection by inducing cognitive load on liars: a review of relevant theories and techniques guided by lessons from polygraph-based approaches. *Front. Psychology* 4:14. doi: 10.3389/fpsyg.2013.00014

This article was submitted to Frontiers in Cognitive Science, a specialty of Frontiers in Psychology.

Copyright © 2013 Walczuk, Igou, Dixon and Tcholakian. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.



When interference helps: increasing executive load to facilitate deception detection in the concealed information test

George Visu-Petra¹, Mihai Varga¹, Mircea Miclea^{1,3} and Laura Visu-Petra^{2*}

¹ Applied Cognitive Psychology Center, Department of Psychology, Babes-Bolyai University, Cluj-Napoca, Romania

² Developmental Psychology Lab, Department of Psychology, Babes-Bolyai University, Cluj-Napoca, Romania

³ COGNITROM Ltd, Cluj-Napoca, Romania

Edited by:

Matthias Gamer, University Medical Center Hamburg-Eppendorf, Germany

Reviewed by:

Bruno Verschueren, Ghent University, Belgium

Xiaoqing Hu, Northwestern University, USA

***Correspondence:**

Laura Visu-Petra, Department of Psychology, Republicii Street No 37, Cluj-Napoca 400015, Romania.

e-mail: laurapetra@psychology.ro

The possibility to enhance the detection efficiency of the Concealed Information Test (CIT) by increasing executive load was investigated, using an interference design. After learning and executing a mock crime scenario, subjects underwent three deception detection tests: an RT-based CIT, an RT-based CIT plus a concurrent memory task (CITMem), and an RT-based CIT plus a concurrent set-shifting task (CITSift). The concealed information effect, consisting in increased RT and lower response accuracy for probe items compared to irrelevant items, was evidenced across all three conditions. The group analyses indicated a larger difference between RTs to probe and irrelevant items in the dual-task conditions, but this difference was not translated in a significantly increased detection efficiency at an individual level. Signal detection parameters based on the comparison with a simulated innocent group showed accurate discrimination for all conditions. Overall response accuracy on the CITMem was highest and the difference between response accuracy to probes and irrelevants was smallest in this condition. Accuracy on the concurrent tasks (Mem and Shift) was high, and responses on these tasks were significantly influenced by CIT stimulus type (probes vs. irrelevants). The findings are interpreted in relation to the cognitive load/dual-task interference literature, generating important insights for research on the involvement of executive functions in deceptive behavior.

Keywords: deception detection, concealed information test, interference design, executive functions, cognitive load

INTRODUCTION

There is a growing body of behavioral, psychophysiological, and neuroimaging evidence revealing that lying is a complex, cognitively demanding behavior. Most of this evidence reflects an overall increase in executive control demands imposed by lying, as compared to truth-telling. Truth-telling is considered a baseline, almost automatic cognitive state (Spence, 2004). To support this claim, lying has been proven to take longer than truth-telling (Spence et al., 2001), necessitating greater cognitive effort (see Vrij et al., 2011, for a recent review). Furthermore, it activates a wider network of prefrontal neural areas linked to executive functioning (see Christ et al., 2009; Gamer, 2011, for reviews). However, recent research questions the “cognitive complexity” view of deception (Gombos, 2006), revealing that in certain contexts lying might not be that cognitively demanding, especially as a result of extensive practice (e.g., Hu et al., 2012a,b; Van Bockstaele et al., 2012). This raises the need for developing deception detection tools less vulnerable to the effects of practice. One interesting possibility is to increase the cognitive workload experienced during deceptive behavior (Vrij et al., 2006). Inducing an overall increase in cognitive/executive load, such as by asking participants to narrate their deceptive stories backwards has been shown to interfere with lying, facilitating the process of lie detection by enhancing verbal and non-verbal cues to deception (Vrij et al., 2008). However, the backwards recall technique has been questioned with regard

to the accuracy and completeness of the retrieved information (Dando et al., 2011), suggesting that a global interference with deceptive and memory processes might induce some unwanted collateral effects. An ingenious recent study (Debey et al., 2012) actively manipulated executive control, using an ego depletion procedure *prior* to detecting deception and inducing goal neglect *during* the deception task (by using longer response-stimulus intervals). Across two experiments, goal neglect, but not the ego depletion procedure facilitated deception detection efficiency, generating longer deceptive response speed (but not consistently lower accuracy).

Vrij et al. (2006) suggested that requiring interviewees to perform a *concurrent* secondary task while being interviewed might provide a useful tool to enhance lie detection. There have been some preliminary experimental attempts to add a parallel task aimed at disrupting the executive functions involved in the deceptive act, yielding mixed evidence in terms of effects on deception. In a Concealed Information Test (CIT, see the description below), Ambach et al. (2008) added a parallel inhibition (Go/No-Go) task. This manipulation was supposed to interfere with the very subprocesses of response inhibition that are required for deceptive responses. However, the physiological and behavioral measures of deception (RTs, error rates) were not significantly affected by introducing this additional measure (see Ambach et al., 2008 for a discussion of these negative findings). In a recent investigation,

Ambach et al. (2011) pursued this line of reasoning, but they introduced a working memory (WM) task in parallel with the deception test. This manipulation affected RTs to critical items to a larger extent when compared to irrelevants. Considering the limitations induced by the very long RTs specific to the psychophysiological measurement design, the authors suggested that a faster pace of the task (asking the subjects to respond within a second) would enhance this preliminary documented effect. This idea was recently tested by introducing an interfering inhibition (dot-probe) task within each trial of the Reaction Time-based (RT-based) CIT, which led to an increase in its detection efficiency (Hu et al., 2013). The present study aimed at further testing this prediction, using the RT-based CIT at a faster pace, and interfering with two different executive functions shown to be involved in the deceptive act (WM updating and shifting). Moreover, rather than introducing a parallel task, peripheral to the deception detection task, in the present study the concurrent task targeted the same items used in the deception detection test, presumably creating a larger interference with deceptive behavior.

The abovementioned “high cognitive workload” studies have used a variety of deception detection paradigms, ranging from naturalistic interviewing settings to elaborated experimental contexts. The use of a unique and well-supported research paradigm which has also been used in ecological settings would substantially benefit the integration of various investigations targeting cognitive control in the deceptive act. This research context could be provided by the CIT, which is one of the most widely adopted techniques by nowadays deception research (Verschuere et al., 2011; Ben-Shakhar, 2012). Originally known as the Guilty Knowledge Test (Lykken, 1959, 1974), this procedure is an interrogation technique designed to test individuals for knowledge that only a guilty person could possess. The subject is presented with several multi-choice questions. For each question, there are several equally plausible alternatives, only one being correct. Hence, the test is based on the rationale that the critical alternative is recognized only by the guilty suspects. A different version of this test based on measuring reaction times was proposed by Seymour et al. (2000), now known as the RT-based CIT (Seymour and Kerlin, 2008; Verschuere et al., 2010). In this procedure, the subject is required to give speeded responses to three types of items: probes, targets, and irrelevants. Probe items are selected from the crime itself and are supposed to represent relevant details of the crime; the irrelevant items share a variable degree of categorical similarity with the relevant items, and are usually several times more numerous. The deceptive subject denies recognition of both irrelevant and probe items. Target items (explicitly learned and recognized as such) are used in order to prevent the subject from entering an automatic mode of responding; they also share categorical similarity with the other two types of items. A number of studies have suggested that this procedure can successfully differentiate between truthful and deceptive responses, or between guilty and innocent participants on the basis of RTs, supporting the validity of the RT-based CIT (see Verschuere and De Houwer, 2011 for a recent review).

The main aim of the present study was to systematically investigate whether introducing a concurrent executive load targeting the very CIT items, rather than a parallel interfering task, would better differentiate between truthful and deceptive responses in

the RT-based CIT. The current investigation used an interference design, introducing tasks involving two executive functions evidenced to be relevant for the deceptive act: memory updating and flexible set-shifting (Morgan et al., 2009; Visu-Petra et al., 2012). In order to efficiently plan and execute a deceptive act, a person needs to continuously monitor and update memory contents in order to distinguish truthful from deceptive responses, and to flexibly alternate between these mental sets in producing the deceptive response (Walczek et al., 2003). A third executive functioning dimension (according to the model proposed by Miyake et al., 2000), namely inhibition, has been documented to be involved in deception (Verschuere et al., 2007; Hu et al., 2013), but it was not directly targeted by the current study.

Consistent with previous findings by Ambach et al. (2011), we hypothesized an increase in CIT detection accuracy due to the introduction of the concurrent memory load condition. We anticipated that the introduction of the requirement to hold on to a memory load while performing recognition judgments would interfere with WM updating processes, and disrupt their efficiency by slowing them down (Logan, 1979). The manipulation was supposed to affect deceptive responses to a greater degree than truthful responses, because they required a larger amount of executive resources compared to simple visual recognition skills necessary for responses to irrelevants and targets, and the executive resources are depleted by the concurrent task. This would be evidenced by an increase in difference scores (RTs) between probes and irrelevants in the CIT plus memory condition. A second research interest was to investigate whether this effect could be replicated when introducing a concurrent task which required flexible set-shifting. We explored whether performance slowing would be further increased in this context, because flexibly shifting between responses in a trial-to-trial manner could place greater executive demands than a simple memory load. In addition to the main measure derived from the CIT (the RT), we wanted to explore whether response accuracy would discriminate between truthful and deceptive responses in the three experimental conditions. Finally, we wanted to see whether performance on the concurrent tasks itself would be more impaired on the trials containing probes than on the trials with irrelevants, thus reflecting the reciprocal interference generated by deception-related increased executive demands.

MATERIALS AND METHODS

PARTICIPANTS

Participants ($N = 75$, 62 females) were recruited from general psychology classes by using an online recruitment system and received credit for their participation. All participants underwent the mock crime procedure described below, followed by the three CIT conditions. Data from the CIT plus memory test of one participant were lost due to a technical failure, and data from one participant were discarded altogether from the analysis because he remembered less than four of the five probes used in this experiment. A remaining total of 73 participants (62 females) were included in the data analyses. The age of participants ranged from 19 to 43 years, and the mean age was 22.76 years ($SD = 4.79$). Participants had normal or corrected-to-normal vision and wore glasses or contact lenses if necessary.

MATERIALS

Concealed information test items were two-word phrases: five probes, five targets, and 20 irrelevants (four corresponding to each probe), which were generated for this study and very similar to items used in previous studies (e.g., Farwell and Donchin, 1991; Seymour et al., 2000; see Appendix). They were displayed using the E-Prime software on a 17" monitor. Each word pair subtended 0.85° of vertical visual angle, and ranged from 2" to 4.2" of horizontal visual angle depending on word pair length, from a viewing distance of approximately 60 cm.

TASKS

All participants completed a series of tasks as follows: they read the instructions for the mock crime, they executed the mock crime, then completed a filler task; afterward, they studied and learned the target items and finally resolved the three CIT conditions (the order of presentation was counterbalanced across subjects).

Mock crime

The participants were initially required to read and sign the informed consent form. Afterward, the mock crime scenario was presented. Written instructions were used at this time, according to which they had to pretend to be a student of Psychology who was about to take a previously failed exam at an important course in the following day. Because of some personal issues, he/she had been unable to study. However, in the previous day, the student had presumably visited the professor's office for a meeting. There he/she noticed a paper on the desk and saw the login Id (Psiho MCC, where the MCC abbreviation stands for – in Romanian – Cognitive Behavioral Modifications, the actual name of the course) and password (*patru verde/four green*) for the discipline's e-mail account which is hosted on the faculty's official web site. With this information, he/she was instructed to access the course e-mail account from a *café* (*Café Amber*) placed in certain *street* (*Bicaz Street*; all locations were chosen from another city in order to avoid previous exposure). After accessing the account (which was created to be identical to a real course application on the actual faculty website), the participant had to search the Inbox for the e-mail with the exam subjects that the professor had sent to the course *tutor* (*Amalia Ciucă*; the name of the actual tutor was used, with her and the professor's consent) for multiplying exam papers. The participant had to forward this message with the attachment to their personal e-mail account.

Subjects read these written instructions twice and memorized (emphasized) the five critical items (i.e., the probes). Afterward, they were asked to go into a distant room of the same building (designated as *Café Amber*) and perform the actions from the scenario (access the e-mail account with the username and password, forward the e-mail). The interface was a mock program designed for this study and was deactivated after the completion of the study.

Following the mock crime, a non-verbal reasoning test taken from a standardized battery was used as a filler task, lasting for about 12–15 min. This data was not analyzed further.

In the target learning phase, the participants learned a sequence of five items similar to the probes. They were instructed to memorize the items in order to reproduce and recognize them. In order to obtain a good memory for the target items the participant was

asked to complete two pencil-and-paper cued recall tests after the memorizing phase: in the first run, they were presented with the first word of the two-word phrase, and in the second run they were presented with the second word. In each run the participant completed the missing item. This was followed by a free recall test. If wrong answers were given at any time, they were again presented with the items and asked to memorize them. A final verbal recall was performed to ensure a good retention of each item.

RT-based CIT

After the mock crime and the target learning phase, the participants undertook the three CIT procedures designed for this study: a classical RT-based CIT, a CIT with a concurrent memory task (CITMem), and a CIT with a concurrent set-shifting task (CITShift).

The items utilized in this study were two-word phrases belonging to three categories of items: *probes* (the five critical items from the mock crime), *targets* (five to-be-recognized items, also from the same category as the probes), and *irrelevants* (items from the same category as the probes, not previously encountered). For each probe, four similar irrelevants were selected. The items were matched on number of syllables across the three categories (see Appendix). In each of the three conditions, each item was repeated four times, generating a total of 120 trials/condition. The participants were instructed to press Yes when presented with the targets, indicating recognition, and No to any other item encountered. The two response keys were counterbalanced across subjects. Item presentation was randomly established by the E-Prime software for the CIT and CITShift conditions. For the CITMem, a randomized list was generated and kept constant across subjects, to allow for verbal recall accuracy to be checked by the experimenter with a response key.

In the CITShift, the primary task remained the same, but the stimuli themselves appeared written in bold or in italics. Subjects had to press the answers to the CIT *once* if the item was written with *bold* and *twice* if the item was written with *italics*. Stimuli were presented equally often in bold or italics. The assignment of number of presses to the respective fonts was also counterbalanced across subjects.

In the CITMem condition, the task was spaced in sequences consisting in groups of three items, with items randomly divided over sequences. The subject again had to press Yes or No to each item according to CIT instructions, but additionally he/she had to memorize the last word of each two-word item. After each three items sequence, a blank screen appeared. The subject had to verbally reproduce the three words he/she had memorized. After this, the participant pressed the space bar in order to initiate the next three items sequence. The experimenter verified the accuracy of verbal answers with an answer-key. A total of 40 memory checks were performed.

Each condition began with a training phase identical in length (16 trials). For each condition, written instructions were presented and verbally clarified by the experimenter. The instructions for the CIT were identical for all the three tasks. For the CITMem and CITShift, general CIT instructions were followed by specific instructions referring to the additional task. A shortened version of the instructions also appeared on the computer screen before

the practice trials. The items used in the training phase were similar to the subsequent CIT items (three probes, three targets, 10 irrelevants).

The inter-stimulus interval randomly varied between 500, 800, and 1100 ms in order to discourage automatic responses or preparation effects (cf. Seymour et al., 2000). If a response was not made within 1200 ms, a “Too slow” message appeared. The 1200 ms interval was established after a pilot study in which shorter stimulus presentation RTs were associated with floor levels of performance on the CITMem and CITShift. No feedback was given (except for the practice trials, where the participant received feedback after every response). Each item remained on the screen until a response was made.

Scoring

For each condition, accuracy and RT (for accurate responses) on the CIT according to stimulus type represented the main collected measures. On the CITMem, an additional index of memory for each stimulus type across trials, and also for mixed groups of three was added. For each group of three items, we checked whether they recalled the last word for irrelevants, probes, or target items, and whether the group of three items was also correctly recalled. For the CITShift, accuracy in pressing once/twice the answer according to stimulus font was calculated; however, an inaccurate shift was not considered to be an error on the CIT (e.g., if the subject pressed once the answer No when presented with a probe it was scored as a shifting error, if the task was to press twice, but it was not scored as a CIT error). However, in the analysis of RTs, only time until first press was recorded and analyzed (for correct CIT responses).

RESULTS

RESPONSE TIME

Group effects

In order to analyze the RT data, an elimination of outliers was first conducted. Since there was an established upper limit for RTs of 1200 ms, we only eliminated responses faster than 200 ms as

outliers. Descriptive data for accuracy and response time according to stimulus type are presented in **Figure 1**. In the subsequent analyses only the comparison between probes and irrelevants is considered, similar to other studies using the RT-based CIT (Seymour et al., 2000; Seymour and Kerlin, 2008; Verschueren et al., 2010; Visu-Petra et al., 2012).

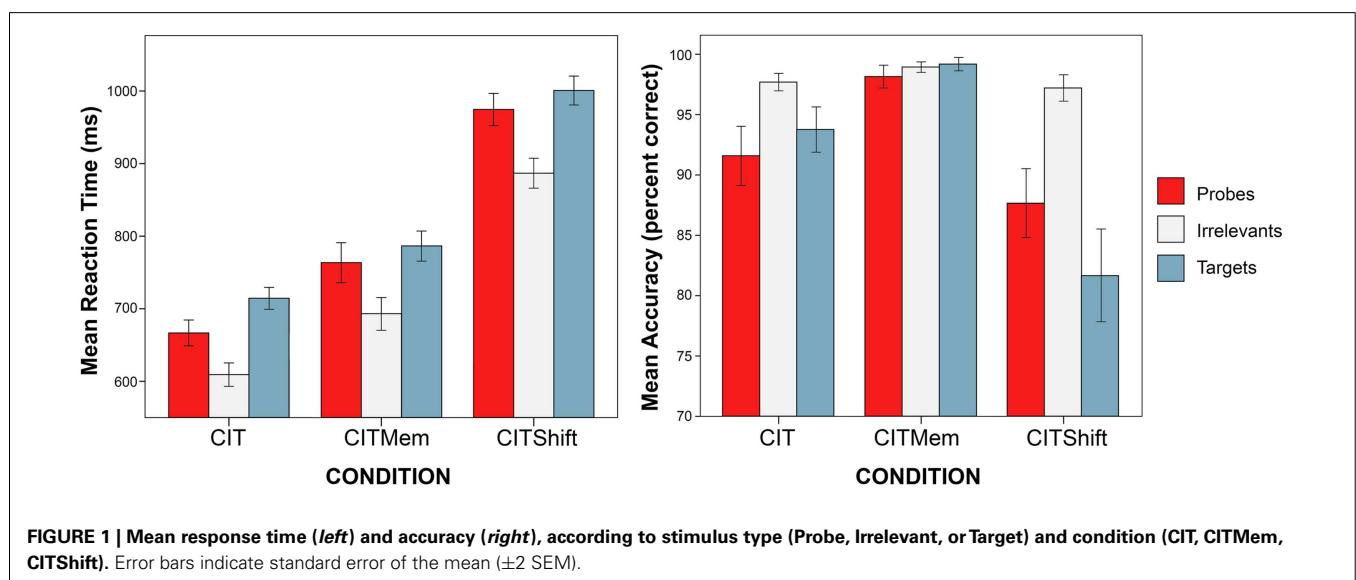
A two-way repeated-measures ANOVA with Condition (CIT vs. CITMem, and CITShift) and Stimulus type (probe vs. irrelevant) as within-subject factors was conducted for the mean RT data. The results showed that there was a significant effect of Condition, $F(2, 144) = 341.91, p < 0.001, \text{MSE} = 9600.04, \text{partial } \eta^2 = 0.83$. Post hoc pairwise comparisons (with a Bonferroni correction) indicated that subjects were significantly faster on the traditional CIT than on both the CITMem, and the CITShift, $p < 0.001$. They were also significantly faster on the CITMem than on the CITShift.

There was a significant main effect of Stimulus type, $F(1, 72) = 288.63, p < 0.001, \text{MSE} = 1958.06, \text{partial } \eta^2 = 0.80$. Across conditions, subjects were faster in responding to irrelevants than to probes, $p < 0.001$ (see **Figure 1**).

Finally, there was a significant Condition \times Stimulus type interaction, $F(2, 144) = 12.5, p < 0.001, \text{MSE} = 678.88, \text{partial } \eta^2 = 0.15$. There was a significant increase across tasks in RTs to both irrelevants, and probes, respectively, with the fastest responses on the CIT, followed by responses on the CITMem, and by longest responses on the CITShift, $p < 0.001$ in each case. To investigate the magnitude of the difference between RTs for irrelevants and probes across conditions, difference scores (difference between mean RTs for probes minus mean RTs for irrelevants) were calculated for each condition. Post hoc paired *t*-tests revealed that RT differences were smaller in the CIT than in the CITMem, $t(72) = 2.12, p = 0.04$, and in the CITShift, $t(72) = 5.23, p < 0.001$. Additionally, difference scores were significantly larger in the CITShift compared to the CITMem, $t(72) = 2.80, p < 0.007$.

Detection efficiency

Measurement of response latency differences across experimental conditions can lead “to an increased likelihood of finding



spurious overadditive interactions" (Faust et al., 1999, p. 777), which could determine an artificial inflation of effect size. The authors recommended *z*-score transformations to augment traditional analyses of raw response latencies. Also, Bush et al. (1993) recommended the use of the *z*-score to remove the influence of individual differences in overall mean response latency within a single group. To eliminate individual differences in responsivity, within-question standardized scores were computed by subtracting the mean of all five responses (one probe and four irrelevants) from the response to the probe and dividing that by the standard deviation of all five values (Ben Shakhar, 1985; Meijer et al., 2007). These standardized scores were then averaged over questions in order to produce a single detection score for the CIT, CITMem, and CITShift (Meijer et al., 2007).

According to signal detection theory, the efficiency of detection may be assessed by considering the degree of separation between the distributions of the detection measure for the innocent and the guilty conditions. Although we included only guilty participants in our study, the distribution of the detection score for innocent individuals can be estimated (Carmel et al., 2003; see also Meijer et al., 2007). Our signal detection parameters were based on a comparison with a simulated innocent group consisting of 73 participants. Following the procedure proposed by Carmel et al. (2003), we generated an innocent group by drawing five values randomly from a standard normal distribution. One value (as the "probe") was standardized relative to the mean and standard deviation of all five values. The computation was repeated five times and the new values were averaged to obtain a score for one innocent participant (Meijer et al., 2007).

We also analyzed the possibility of increasing detection efficiency by combining measures of concealed information. Using the method described by Nahari and Ben-Shakhar (2011) and by Hu and Rosenfeld (2012), we averaged the *z scores* from CIT and the *z scores* from CIT Shift into a new combined measure.

After we computed the distance (in standard deviation units) between the centers of the two distributions (d'), we derived the *area under the receiver operating characteristic – ROC* (Ben Shakhar and Elaad, 2003). The *area under the ROC curve (AUC)* represents the degree of separation between the distributions of the response time from guilty and innocent participants. It varies between 0 and 1 (perfect detection level), with a chance level of 0.5 (Hu and Rosenfeld, 2012). The d' and the *AUC* for each condition are displayed in Table 1.

Intraindividual bootstrap analysis

To allow for a more in-depth testing of probe versus irrelevant differences within an individual, data from each condition were bootstrapped (Wasserman and Bockenholt, 1989) and hit rates were subsequently calculated. After excluding incorrect behavioral responses and artifacts, a computer program draws, with replacement, a set of individual probe reaction times equal to the number of accepted probe trials in each block and also draws (with replacement) an equal number of irrelevant reaction times, selected randomly from the irrelevant trials. Next, a difference score is obtained by subtracting the mean irrelevant reaction times from the mean probe reaction times. This process is repeated 500 times (Verschueren et al., 2009), resulting in a distribution of 500

Table 1 | Means, Standard deviations, standardized differences (d') and area under the curve (AUC) for CIT, CITMem, CITShift, and the combination of CIT and CITShift for the Guilty and Innocent Conditions.

Measure	Mean <i>z</i> guilty	Standard deviation guilty	Mean <i>z</i> innocent	Standard deviation innocent	d'	AUC
CIT	0.48	0.37	-0.04	0.31	1.54	0.86
CITMem	0.53	0.43	-0.03	0.38	1.39	0.83
CITShift	0.59	0.26	-0.02	0.43	1.71	0.88
CIT and CITShift	0.53	0.32	-0.08	0.37	1.75	0.89

Table 1 reveals that d' values for the CIT, CITMem, and CITShift were 1.54, 1.39, and 1.71, respectively. The d' value for the combination of CIT and CITShift was 1.75. The areas under the ROC curve (AUC) were 0.86 for the CIT, 0.83 for the CITMem, 0.88 for the CITShift, and 0.89 for the combination between CIT and CITShift (all other combinations had equal or lower AUCs compared to individual measures).

differences scores. If the mean difference score minus 1.29 times the standard deviation is greater than zero, it can be concluded with 90% confidence that the probe reaction times are slower than the irrelevant ones.

Bootstrapping of the CIT reaction times resulted in a hit rate of 67%, i.e., for 49 out of 73 participants concealed information was detectable through their slower responses on probe stimuli. For the CITMem, a hit rate of 64% was computed, while for CIT-Shift, 68% of the participants displayed a reaction time for probes that sufficiently deviated from that for irrelevant stimuli to be of diagnostic value.

RESPONSE ACCURACY

CIT accuracy

Additional analyses regarding performance accuracy according to stimulus type were conducted, in order to ensure the comparability of the current procedure with previous data reported by studies using similar methodology (e.g., Seymour et al., 2000). First, mean percent correct for responses to irrelevants and for deceptive responses to probes were calculated (see Figure 1). In order to directly compare percentages for the two stimulus types, an arcsine transformation was then applied to this percent correct data (Cohen, 1988, cf. Gamer et al., 2007).

First, a two-way repeated-measures ANOVA with Condition (CIT vs. CITMem vs. CITShift) and Stimulus type (probe vs. irrelevant) as within-subject factors was conducted. The results showed that there was a significant effect of Condition, $F(2, 144) = 31.30$, $p < 0.001$, $MSE = 0.03$, partial $\eta^2 = 0.30$. Post hoc pairwise comparisons (with a Bonferroni correction) indicated that subjects were significantly less accurate on both the CIT and the CITShift than on the CITMem (although accuracy on the CIT and on the CITShift did not differ).

There was also a significant main effect of Stimulus type, $F(1, 72) = 80.86$, $p < 0.001$, $MSE = 0.01$, partial $\eta^2 = 0.53$. Across conditions, accuracy in responses to irrelevants was higher than accuracy in responses to probes, $p < 0.001$ (see Figure 1).

Finally, there was a significant Condition \times Stimulus type interaction, $F(2, 144) = 23.59, p < 0.001$, $MSE = 0.01$, partial $\eta^2 = 0.25$. Accuracy in response to irrelevants differed across tasks, $F(2, 144) = 10.65, p < 0.001$, $MSE = 0.01$, partial $\eta^2 = 0.13$, with responses on the CITMem being more accurate than on both CIT and CITShift, $p < 0.05$. Accuracy in response to probes also significantly differed across tasks, $F(2, 144) = 33.67, p < 0.001$, $MSE = 0.03$, partial $\eta^2 = 0.32$. Again, *post hoc* contrasts revealed that accuracy to probes on the CIT and CITShift was significantly lower than accuracy to probes on the CITMem, $p < 0.05$. To investigate the magnitude of the difference between accuracy for irrelevants and probes across conditions, difference scores (accuracy for irrelevant minus accuracy for probes) were calculated for each condition. *Post hoc* paired *t*-tests revealed that the difference between irrelevants and probes was larger on the CITShift, compared to both CIT, $t(72) = 2.37, p < 0.02$, and to CITMem, $t(72) = 6.58, p < 0.001$, respectively. This difference was also larger in the CIT, compared to the CITMem, $t(72) = 4.93, p < 0.001$.

Accuracy on the concurrent tasks

A final step was to check for accuracy on the secondary tasks (Mem and Shift). Results showed that accuracy for recalling groups of three on the CITMem was high, mean percent correct = 93.37, $SD = 5.62$. Comparing memory for probes versus irrelevants (after the arcsine transformation of percent correct data), we found that subjects were significantly more accurate in recalling the last word of the probes, than of the irrelevants, $t(72) = 7.85, p < 0.001$.

Overall accuracy in shifting between responses to stimuli written in bold or italics was also high, mean percent correct = 87.24, $SD = 11.06$. This time, accuracy in shifting responses to probes was lower than accuracy in shifting responses to irrelevants (after the arcsine transformation of percent correct data), $t(72) = 6.88, p < 0.001$.

DISCUSSION

The present study analyzed how introducing an additional executive load impacts the accuracy and efficiency of deceptive responses in the RT-based CIT. We hypothesized that the introduction of a concurrent memory load or of flexible shifting demands along with the primary recognition task would selectively interfere with the executive processes required by deception. Therefore, we expected increased detection accuracy of the RT-based CIT in the two conditions with concurrent executive demands, compared to the traditional CIT. We anticipated that the introduction of more complex shifting demands would affect performance to a larger degree than the memory demands. Finally, we also checked whether performance on the concurrent task was itself affected by CIT stimulus type (probe vs. irrelevant).

The results partially confirmed these predictions, but revealed interesting distinctions between group and individual detection efficiency, and between performance accuracy and response time. First, it should be pointed out that the concealed knowledge effect was confirmed across tasks, with subjects presenting longer RTs and lower accuracy on the probes, compared to the irrelevants. This supports the potential of the two RT-based CIT versions (with additional memory load or set-shifting demands) to distinguish between truthful and deceptive responses.

Looking at group differences between conditions in terms of RT, we found that subjects were faster on the CIT than on the versions containing additional memory updating or set-shifting demands. This difference was probably a consequence of the extra time required to deal with the increased cognitive load, which affected preparatory, processing, or execution stages of responses in the dual-task conditions (Pashler, 1994). The result confirms previous findings that have used an interfering WM task in the CIT, which increased RTs to both irrelevants and probes (Ambach et al., 2011). Similar to Ambach's study, the increase in RTs to probes was larger than the increase in RTs to irrelevants, with the outcome of an increased RT-based detection efficiency in the two conditions that contained interfering tasks, compared to the traditional CIT condition – at least at this group level. Since the design did not allow us to directly contrast the influence of the additional cognitive load on guilty versus innocent participants' behavior, a next step was to simulate a hypothetical group of innocent subjects.

The comparison between distributions of the guilty group and the simulated innocent group showed that the CIT d' value was slightly below the average effect size ($d' = 1.55$) computed in the meta-analysis made by Ben Shakhar and Elaad (2003) for the psychophysiological CIT. The ability of all the measures to differentiate between guilty and innocent participants was evident from the d' values. The values of 1.39, 1.54, 1.71, 1.75 for the concealed information measures in this study represent a large effect size (Cohen, 1988). Also, the computed AUC showed accurate discrimination for all conditions, with the highest rate for the combined measure (CIT + CITShift). Among the two interfering tasks, the demand to flexibly shift responses on a trial-to-trial basis created the largest discrepancy between responses to probes and to irrelevants, and was also associated with the highest hit rate (68%) among the three conditions, although the differences among them were not significant.

In terms of performance *accuracy*, subjects had fewer errors in a CITMem, compared to the traditional CIT and to the CIT-Shift conditions. Importantly, this effect was visible for all stimulus types, and did not differentiate between them. Increasing demands for attentional control induced by concurrent tasks have not been found to affect simple recognition accuracy (Baddeley et al., 1984; Craik et al., 1996), unless there is a deep encoding of the to-be-recognized items (Hicks and Marsh, 2000). It is plausible that the additional conceptual processing required by the memory task might have led to a deeper encoding and to a better subsequent recognition of the stimuli in the CIT task. Since we could not prioritize one task over another, it could be conjectured that the subjects strategically used a sequential strategy. In this context, not only would the two tasks not disrupt each other, but performance on one task could even be enhanced by a deeper earlier processing of the stimuli within the other. When input stimuli are similar, it is possible for dual-task performance to be enhanced because the "same set of processing machinery could be turned on and used for both" and because overt responses were not incompatible between tasks (Pashler, 1994, p. 221). For instance, when focusing on encoding the last word of the probes, subjects could become more aware of the type of stimulus (probe or irrelevant) for the CIT response. Or conversely, if they first focused on responding to the CIT, recognition of the item as a probe could result in an

enhanced memory for the previously encountered stimulus. This was indeed demonstrated by better overall accuracy in probe recall. The prolonged/more intensive processing of the stimuli in this condition was translated into an increase in response time compared to the CIT, leading to a potential speed-accuracy tradeoff that is often found in dual-task contexts (Schumacher et al., 2001).

Why was a similar effect not visible in the CITShift condition? In this case, overall accuracy wasn't significantly different from the traditional CIT, and further, the crucial difference between truthful and deceptive responses was even enhanced – just as we initially expected for both dual-task conditions. Two types of explanations might account for the different findings resulting in enhanced detection efficiency with this task. Firstly, both lower accuracy on the Shift task and the increased processing time suggest that the concurrent task was more difficult and executive-demanding than the memory load task. This confirms the superior executive demands induced by switching between task sets, when compared to simple memory storage (Oberauer et al., 2003). The conclusion is also supported by the greater reciprocal interference between competing tasks when probes were presented. The result of the interference led to a decrease in performance in both the CIT response (longer RT, lower accuracy compared to irrelevants) and the shifting task (lower accuracy for probes). Secondly, the shifting task targeted the *perceptual* (font type), and not the conceptual dimension of the CIT stimuli. This could have generated a greater incompatibility between the two tasks, affecting the more executive-demanding deception trials more. The literature (Pashler and Christian, 1994) suggests yet another possibility: the two simultaneous overt manual responses elicited by the CITShift interfered to a greater degree than the manual plus vocal – non-simultaneous – response present in the CITMem. Finally, an interesting possibility is that the superior accuracy found in the CITMem could simply be a result of the task providing the participants with regular (self-paced) breaks in order to recall the items. This could help them maintain better focus and diminish the goal neglect induced by the (fast-paced) superimposing of two tasks (such as in the CITShift).

Accuracy on the *concurrent tasks* was high in both conditions, but much higher in the Mem task (93%) compared to both the Shift task (87%) and to the previous investigation of Ambach et al. (2011) using an n-back task (83%). It has been shown that if the memory load is significantly below the subjects' memory span (in this case, three elements), people's ability to retain the memory load is usually unaffected by the concurrent task, except for a relative slowing of overall task performance (Pashler, 1994) also noticeable in the present study. Interestingly, probes were better recalled than irrelevant items. This could be a result of previous exposure to the probe stimuli during the mock crime. The preferential recall of probes could also indicate memory enhancement for stimuli with emotional/motivational significance, compared to neutral stimuli (Kensinger and Corkin, 2003). The Shift task revealed an opposite trend, namely poorer shifting accuracy in response to trials containing probes. We have already proposed some explanations suggesting the higher interference between the CIT and Shift tasks in the more executive-demanding tasks containing probes. The results obtained with the concurrent task underline the importance of equating for interfering task

difficulty; this could be achieved by a control condition containing only the concurrent tasks, but not targeting items from the CIT.

To summarize, by contrasting general detection efficiency between the three conditions, we found the following. According to the group analyses, both dual-task conditions were superior in discriminating between truthful and deceptive responses. Signal detection parameters based on a comparison with the simulated innocent group showed accurate discrimination for all conditions, but did not reveal the same advantage of the dual-task conditions over the traditional RT-based CIT. This apparent inconsistency is not simply a byproduct of the overall slower responses found in the dual tasks, as revealed by our analyses on standardized data. The most plausible explanation is that some participants in the CITShift condition might have presented extremely large probe-irrelevant differences, which were responsible for the group effect. In the light of the current research, the comparison of differences in response latency at a group level need to be interpreted with caution. Combining analyses performed on raw and transformed data can provide important information regarding the most appropriate interpretation of differences in response latency. The computed hit rates for all conditions were slightly higher than those previously found in other RT-based CIT studies (e.g., hit rate of 56%, Verschuere et al., 2009). However, the hit rates computed in our study were still modest (as compared to 95% discrimination accuracy found by Seymour et al., 2000, although different estimating methods were used in that study). Looking at combinations between CIT versions, a combined measure including both CIT and CITShift showed the highest discrimination efficiency. In terms of accuracy, the demand to flexibly shift between types of responses generated the largest discrepancy between probes and irrelevants, while the additional memory load led to ceiling levels of performance accuracy on the CIT (98% for both probes and irrelevants). Performance accuracy on the concurrent tasks was affected by the type of trial (truthful or deceptive), revealing that these tasks could themselves provide valuable clues for deception detection.

The study extends the existing literature dealing with the impact of interfering tasks on the CIT (Ambach et al., 2008, 2011; Hu et al., 2013) in several ways. In the Ambach and collaborators' studies the exposure time for each CIT (pictorial) stimulus was very large (10 s) in order to collect physiological measures. The authors themselves state that the longer mean RT to CIT stimuli than in other studies could be responsible for the surprising results of shorter RTs for probes than for irrelevants, possibly suggesting strategic alterations of responses in order to appear innocent. In our study the use of the RT-based CIT, the faster pace of the task (1.2 s per verbal stimulus) led to a greater temporal overlap between the primary and the concurrent task, probably generating a stronger interference. However, the self-paced nature of the task induces a potential confound: individual response speed influences overall task speed because by responding earlier the subject receives the following item earlier. A basic measure of psychomotor speed could be introduced to investigate the impact of this individual difference. A further potential confound might be introduced by using three ISIs. However, an analysis of RTs to each stimulus type separated by the preceding ISI interval (500, 800, or 1100 ms) revealed no significant differences.

Second, in both Ambach et al.'s (2008, 2011), and the Hu et al. (2013) studies the inhibition task was peripheral and did not involve the CIT stimuli themselves. In the present study, the concurrent task involved processing the CIT stimuli themselves (remembering the second word of each item in the CITMem and shifting between CIT stimuli written in bold or italics in the CIT-Shift). Again, this might have increased the interference between the two tasks, and led to differences in accuracy/RT between conditions. Finally, Ambach et al. (2011) consider the assignment of conditions (with or without parallel tasks) to large blocks as a potential limitation of their initial study. They favored rapid switches between conditions in the second study. However, we believe that in our design, this manipulation would have induced additional trial-by-trial switching costs, which would obscure the specific effect of memory load/shifting interference. Thus, a large blocks counterbalanced design was chosen.

In the present study, the detection efficiency (compared to a simulated innocent group) in the conditions in which interfering memory ($AUC = 0.83$) or shifting ($AUC = 0.88$) demands were introduced was not significantly larger than in the pure RT-based CIT ($AUC = 0.86$). This final value is strikingly similar to the one obtained by Hu et al. (2013) for their pure RT-based CIT in the comparison with an authentic innocent group ($AUC = 0.86$). However, in their case, the introduction of an interfering inhibition task led to a significant improvement in detection efficiency ($AUC = 0.94$). One possibility is that inhibition plays a more crucial role in the deception processes required by the CIT than memory updating or switching, so that interfering with these inhibitory processes leads to greater disruption of deceptive responses. However, design differences between ours and the Hu et al. (2013) study (e.g., a substantially larger number of trials in our case, the use of a peripheral rather than central interference task in their case, and the use of a simulated versus a real innocent group) makes it problematic to directly contrast the findings of these two studies. Further research is needed to disentangle the differential contributions of inhibition, switching, and memory updating demands to the production and execution of deceptive responses in the CIT, preferably in a unitary interference design such as the one proposed in the present study.

An important limitation of the present study is the fact that the exposure time for each stimulus exceeded 1 s, allowing for potential strategic alterations of response speed (Seymour et al., 2000). This could account for the relatively smaller difference between RTs for irrelevants (50–100 ms) and for probes than those obtained in other studies, in which the difference was approximately 200 ms (Seymour et al., 2000; Seymour and Kerlin, 2008, but see Verschuere et al., 2010, for similar difference values to ours). Other limits include the uneven distribution of the sample by gender, which can affect the generalization of the data from the present investigation.

In accordance with the conclusions made by Meijer et al. (2007), our results also indicate that it is worthwhile to combine several different types of lie detection measures. Future studies should make a direct comparison between the incremental validity of RT-based CIT and an RT-based CIT plus a CITShift. The inclusion of an authentic, rather than a simulated innocent group is also recommended.

Both the RT-based CIT, and the "cognitive load" paradigms are recent developments in deception detection research. They are supported by a growing body of evidence (so far, mostly laboratory) that can inform research into the cognitive mechanisms involved in the deceptive act. The use of an interference design can deepen this understanding, creating a selective disruption of a particular executive skill involved in deception. Theoretically, by experimentally introducing different concurrent tasks, one can speculate with regards to the extent to which a particular executive skill is essential to the deceptive act when disrupted, and thus inform research into the neurocognitive mechanisms involved in deception. An implicit assumption which guides the interpretation of our results is that there is a general mechanism subserving both executive functioning and deceptive responses (Johnson et al., 2004), so that disrupting the efficiency of executive functions would directly impact the way a person constructs and executes the deceptive response. However, there is an open question regarding the possibility to dissociate the executive processes underlying deceptive behavior based on such interference designs. Considering the differences between the two dual-task experimental conditions (different fonts used only in the CITShift, regular breaks provided only by the CITMem), their differential impact on deception detection cannot be directly contrasted. Alternative interference designs which would equate for all these experimental variables and would also separately test for individual proficiency in distinct executive functions could offer valuable insights into the executive mechanisms underlying deceptive behavior.

Could the RT-based CIT plus a concurrent task be potentially implemented in field settings? Our results caution us not to transfer this procedure without further documenting its impact upon both RT and accuracy of responses. In the case of CITMem, while the RT for correct responses (the main output) supports the potential of such interference designs to enhance deception detection, an analysis of response accuracy reveals that there are also more correct responses, which makes their comparison with the CIT questionable in terms of RT. As suggested previously, it is possible that this effect might be a result of the CITMem targeting the very contents of the CIT, leading to an increased/prolonged processing of these contents, and to a better performance in deceptively denying their recognition. Further research should confirm whether the introduction of a CITMem peripheral to the CIT in the rapid-paced version of the CIT might provide an optimal candidate for detecting deception.

The demand to flexibly shift between two types of motor responses in accordance to a perceptual characteristic of the CIT stimuli was found to discriminate best between truthful and deceptive responses (at least at a group level). This result has potential implications for interviewing techniques, especially for those involving visual stimuli. Interviewers can alternate between relevant and irrelevant questions regarding critical stimuli from an investigation. It has been shown that rapid alternations between question types (e.g., relevant and irrelevant/unanticipated questions), differentially affects liars' and truth-tellers' responses (Vrij et al., 2009). However, a cautionary note relates to the possibility of using the CITShift as a countermeasure. More specific, deceptive subjects might deliberately focus

on the perceptual characteristics of the stimulus and ignore their contents, undermining the deception detection process. An important detail is to use a strict response timing deadline that would not permit the participants to strategically alter their responses (as stressed by Seymour et al., 2000). In addition, the fact that the participant would focus only on the secondary task as a counter-measure and ignore the CIT task (leading to higher error rates) would be reflected in an increased accuracy on this concurrent task and facilitate the detection of deliberate faking.

Finally, our results suggest that in any potential application of the RT-based CIT, participants' responses should be videotaped

REFERENCES

- Ambach, W., Stark, R., Peper, M., and Vaitl, D. (2008). An interfering Go/No-go task does not affect accuracy in a concealed information test. *Int. J. Psychophysiol.* 68, 6–16.
- Ambach, W., Stark, R., and Vaitl, D. (2011). An interfering n-back task facilitates the detection of concealed information with EDA but impedes it with cardiopulmonary physiology. *Int. J. Psychophysiol.* 80, 217–226.
- Baddeley, A. D., Lewis, V., Eldridge, M., and Thomson, N. (1984). Attention and retrieval from long term memory. *J. Exp. Psychol. Gen.* 113, 518–540.
- Ben Shakhar, G. (1985). Standardization within individuals: a simple method to neutralize individual differences in skin conductance. *Psychophysiology* 22, 292–299.
- Ben Shakhar, G., and Elaad, E. (2003). The validity of psychophysiological detection of information with the guilty knowledge test: a metaanalytic review. *J. Appl. Psychol.* 88, 131–151.
- Ben-Shakhar, G. (2012). Current research and potential applications of the concealed information test: an overview. *Front. Psychol.* 3:342. doi:10.3389/fpsyg.2012.00342
- Bush, L. K., Hess, U., and Wolford, G. (1993). Transformations for within-subject designs: a Monte Carlo investigation. *Psychol. Bull.* 113, 566–579.
- Carmel, D., Dayan, E., Naveh, A., Raveh, O., and Ben-Shakhar, G. (2003). Estimating the validity of the guilty knowledge test from simulated experiments: the external validity of mock crime studies. *J. Exp. Psychol. Appl.* 9, 261–269.
- Christ, S. E., Van Essen, D. C., Watson, J. M., Brubaker, L. E., and McDermott, K. B. (2009). The contributions of prefrontal cortex and executive control to deception: evidence from activation likelihood estimate meta-analyses. *Cereb. Cortex* 19, 1557–1566.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd Edn. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Craik, F. I. M., Govoni, R., Naveh-Benjamin, M., and Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *J. Exp. Psychol. Gen.* 125, 159–180.
- Dando, C. J., Ormerod, T. C., Wilcock, R., and Milne, R. (2011). Change temporal order retrieval: help or hindrance. *Cognition* 121, 416–421.
- Debey, E., Verschuere, B., and Crombez, G. (2012). Lying and executive control: an experimental investigation using ego depletion and goal neglect. *Acta Psychol.* 140, 133–141.
- Farwell, L. A., and Donchin, E. (1991). The truth will out: interrogative polygraphy ("lie detection") with event-related potentials. *Psychophysiology* 28, 531–547.
- Faust, M. E., Balota, D. A., Spieler, D. H., and Ferraro, F. R. (1999). Individual differences in information processing rate and amount: implications for group differences in response latency. *Psychol. Bull.* 125, 777–799.
- Gamer, M. (2011). "Detecting of deception and concealed information using neuroimaging techniques," in *Memory Detection: Theory and Application of the Concealed Information Test*, eds B. Verschuere, G. Ben-Shakhar, and E. H. Meijer (Cambridge: Cambridge University Press), 90–113.
- Gamer, M., Bauermann, T., Stoeter, P., and Vossel, G. (2007). Covariations among fMRI, skin conductance and behavioral data during processing of concealed information. *Hum. Brain Mapp.* 28, 1287–1301.
- Gombos, V. A. (2006). The cognition of deception: the role of executive processes in producing lies. *Genet. Soc. Gen. Psychol. Monogr.* 132, 197–214.
- Hicks, J. L., and Marsh, R. L. (2000). Toward specifying the attentional demands of recognition memory. *J. Exp. Psychol. Learn.* 26, 1483–1498.
- Hu, X., Chen, H., and Fu, G. (2012a). A repeated lie becomes a truth? *Front. Psychol.* 3:488. doi:10.3389/fpsyg.2012.00488
- Hu, X., Rosenfeld, J. P., and Bodenhausen, G. V. (2012b). Combating automatic autobiographical associations: the effect of instruction and training in strategically concealing information in the autobiographical implicit association test. *Psychol. Sci.* 23, 1079–1085.
- Hu, X., Evans, A., Wu, H., Lee, K., and Fu, G. (2013). An interfering dot-probe task facilitates the detection of mock crime memory in a reaction time (RT)-based concealed information test. *Acta Psychol. (Amst.)* 142, 278–285.
- Hu, X., and Rosenfeld, J. P. (2012). Combining the P300-complex trial-based concealed information test and the reaction time-based autobiographical implicit association test in concealed memory detection. *Psychophysiology* 49, 1090–1100.
- Johnson, R. Jr., Barnhardt, J., and Zhu, J. (2004). The contribution of executive processes to deceptive responding. *Neuropsychologia* 42, 878–901.
- Kensinger, E. A., and Corkin, S. (2003). Memory enhancement for emotional words: are emotional words more vividly remembered than neutral words? *Mem. Cogn.* 31, 1169–1180.
- Logan, G. D. (1979). On the use of a concurrent memory load to measure attention and automaticity. *J. Exp. Psychol. Hum.* 5, 189–207.
- Lykken, D. T. (1959). The GSR in the detection of guilt. *J. Appl. Psychol.* 43, 385–388.
- Lykken, D. T. (1974). Psychology and the lie detection industry. *Am. Psychol.* 29, 725–739.
- Meijer, E. H., Smulders, F. T. Y., Johnston, J. E., and Merckelbach, H. L. G. J. (2007). Combining skin conductance and forced choice in the detection of deception. *Psychophysiology* 44, 814–822.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., and Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. *Cogn. Psychol.* 41, 49–100.
- Morgan, C. J., LeSage, J. B., and Kosslyn, S. M. (2009). Types of deception revealed by individual differences in cognitive abilities. *Soc. Neurosci.* 4, 554–569.
- Nahari, G., and Ben-Shakhar, G. (2011). Psychophysiological and behavioral measures for detecting concealed information: the role of memory for crime details. *Psychophysiology* 48, 733–744.
- Oberauer, K., Süß, H.-M., Wilhelm, O., and Wittmann, W. W. (2003). The multiple faces of working memory: storage, processing, supervision, and coordination. *Intelligence* 31, 167–193.
- Pashler, H. (1994). Dual-task interference in simple tasks: data and theory. *Psychol. Bull.* 116, 220–244.
- Pashler, H., and Christian, C. (1994). *Bottlenecks in Planning and Producing Vocal, Manual and Foot Responses*. Center for Human Information Processing Technical Report.
- Schumacher, E. H., Seymour, T. L., Glass, J. M., Fencsik, D. E., Lauber, E. J., Kieras, D. E., et al. (2001). Virtually perfect time sharing in dual-task performance: uncorking the central cognitive bottleneck. *Psychol. Sci.* 12, 101–108.
- Seymour, T. L., and Kerlin, J. R. (2008). Successful detection of verbal and visual concealed knowledge using an RT-based paradigm. *Appl. Cogn. Psychol.* 22, 475–490.
- Seymour, T. L., Seifert, C. M., Shafto, M. G., and Mosmann, A. L. (2000). Using response time measures to assess "guilty knowledge." *J. Appl. Psychol.* 85, 30–37.
- Spence, S. A. (2004). The deceptive brain. *J. R. Soc. Med.* 97, 6–9.
- Spence, S. A., Farrow, T. F. D., Herford, A. E., Wilkinson, I. D., Zheng, Y., and

and analyzed in terms of response accuracy, consistency, and speed, because the outputs from multiple deception indexes do not necessarily converge.

ACKNOWLEDGMENTS

The present study was supported by the National University Research Council and by COGNITROM Ltd. We thank Oana Ciornoi, Elena Cristiana Stefan, and Mihai Bucovschi for their assistance with the data collection process. We are also grateful to Dr. Bruno Verschuere for his helpful comments on an earlier version of the manuscript.

- Woodruff, P. W. R. (2001). Behavioral and functional anatomical correlates of deception in humans. *Neuroreport* 12, 2849–2853.
- Van Bockstaele, B., Verschuere, B., Moens, T., Suchotzki, K., Debey, E., and Spruyt, A. (2012). Learning to lie: effects of practice on the cognitive cost of lying. *Front. Psychol.* 3:526. doi:10.3389/fpsyg.2012.00526
- Verschuere, B., Ben-Shakhar, G., and Meijer, E. (2011). *Memory Detection: Theory and Application of the Concealed Information Test*. Cambridge: Cambridge University Press.
- Verschuere, B., Crombez, G., Degrootte, T., and Rosseel, Y. (2010). Detecting concealed information with reaction times: validity and comparison with the polygraph. *Appl. Cogn. Psychol.* 24, 991–1002.
- Verschuere, B., Crombez, G., Koster, E., Van Bockstaele, B., and De Clercq, A. (2007). Startling secrets: startle eye blink modification by concealed crime information. *Biol. Psychol.* 76, 52–60.
- Verschuere, B., and De Houwer, J. (2011). “Detecting concealed information in less than a second: response-latency based measures,” in *Memory Detection. Theory and Application of the Concealed Information Test*, eds B. Verschuere, G. Ben-Shakhar, and E. Meijer (London: Cambridge University Press), 46–62.
- Verschuere, B., Rosenfeld, J. P., Wino-grad, M., Labkovsky, E., and Wiersema, R. (2009). The role of deception in P300 memory detection. *Legal Crim. Psychol.* 14, 253–262.
- Visu-Petra, G., Miclea, M., and Visu-Petra, L. (2012). RT-based detection of concealed information in relation to individual differences in executive functioning. *Appl. Cogn. Psychol.* 26, 342–351.
- Vrij, A., Fisher, R., Mann, S., and Leal, S. (2006). Detecting deception by manipulating cognitive load. *Trends Cogn. Sci. (Regul. Ed.)* 10, 141–142.
- Vrij, A., Granhag, P. A., Mann, S., and Leal, S. (2011). Outsmarting the liars: toward a cognitive lie detection approach. *Curr. Dir. Psychol. Sci.* 20, 28–32.
- Vrij, A., Leal, S., Granhag, P. A., Mann, S., Fisher, R. P., Hillman, J., et al. (2009). Outsmarting the liars: the benefit of asking unanticipated questions. *Law Hum. Behav.* 33, 159–166.
- Vrij, A., Mann, S., Fisher, R., Leal, S., Milne, B., and Bull, R. (2008). Increasing cognitive load to facilitate lie detection: the benefit of recalling an event in reverse order. *Law Hum. Behav.* 32, 253–265.
- Walczuk, J. J., Roper, K. S., Seemann, E., and Humphrey, A. M. (2003). Cognitive mechanisms underlying lying to questions: response time as a cue to deception. *Appl. Cogn. Psychol.* 17, 755–774.
- Wasserman, S., and Bockenholt, U. (1989). Bootstrapping: applications to psychophysiology. *Psychophysiology* 26, 208–221.
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Received: 29 August 2012; accepted: 07 March 2013; published online: 28 March 2013.*
- Citation: Visu-Petra G, Varga M, Miclea M and Visu-Petra L (2013) When interference helps: increasing executive load to facilitate deception detection in the concealed information test. *Front. Psychol.* 4:146. doi: 10.3389/fpsyg.2013.00146*
- This article was submitted to Frontiers in Cognitive Science, a specialty of Frontiers in Psychology.*
- Copyright © 2013 Visu-Petra, Varga, Miclea and Visu-Petra. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.*

APPENDIX

Table A1 | Items (English translation) derived from the mock crime and used in the RT-based CIT.

Probes	Targets	Irrelevants		
<i>Psiho MCC*</i>	<i>Psiho GEN</i>	Psiho DEZ	Psiho COG	Psiho JUD
<i>Four green</i>	<i>Nine brown</i>	Six yellow	Five orange	One red
<i>Amalia Ciucă</i>	<i>Bianca Coman</i>	Emilia Anton	Marcela Petre	Lavinia Dinu
<i>Café Amber</i>	<i>Café Dante</i>	Café Tonka	Café Zebra	Café Antic
<i>Bicaz Street</i>	<i>Tomis Street</i>	Tusnad Street	Arges Street	Bacau Street
				Galati Street

*The abbreviation *Psiho XXX* stands for the name of the course in Romanian, for example: *Psiho GEN*, general psychology; *MCC*, cognitive behavioral modifications; *DEZ*, developmental psychology; *COG*, cognitive psychology; *JUD*, forensic psychology; *SOC*, social psychology; the students are used with these abbreviations for their courses names.



Detecting concealed information from groups using a dynamic questioning approach: simultaneous skin conductance measurement and immediate feedback

Ewout H. Meijer^{1,2*}, Gary Bente³, Gershon Ben-Shakhar² and Andreas Schumacher¹

¹ Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, Netherlands

² Department of Psychology, The Hebrew University of Jerusalem, Jerusalem, Israel

³ Department of Social Psychology and Media Psychology, University of Cologne, Cologne, Germany

Edited by:

Matthias Gamer, University Medical Center Hamburg-Eppendorf, Germany

Reviewed by:

Judith Peth, University Medical Center Hamburg-Eppendorf, Germany
Michael T. Bradley, University of New Brunswick, Canada

***Correspondence:**

Ewout H. Meijer, Forensic Psychology Section, Department of Clinical Psychological Science, Faculty of Psychology and Neuroscience, Maastricht University, PO BOX 616, 6200 MD Maastricht, Netherlands.
e-mail: eh.meijer@maastrichtuniversity.nl

Lie detection procedures typically aim at determining the guilt or innocence of a single suspect. The Concealed Information Test (CIT), for example, has been shown to be highly successful in detecting the presence or absence of crime-related information in a suspect's memory. Many of today's security threats, however, do not come from individuals, but from organized groups such as criminal organizations or terrorist networks. In this study, we tested whether a plan of an upcoming mock terrorist attack could be extracted from a group of suspects using a dynamic questioning approach. One-hundred participants were tested in 20 groups of 5. Each group was asked to plan a mock terrorist attack based on a list of potential countries, cities, and streets. Next, three questions referring to the country, city, and street were presented, each with five options. Skin conductance in all five members of the group was measured simultaneously during this presentation. The dynamic questioning approach entailed direct analysis of the data, and if the average skin conductance of the group to a certain option exceeded a threshold, this option was followed up, e.g., if the reaction to the option "Italy" exceeded the threshold, this was followed up by presenting five cities in Italy. Results showed that in 19 of the 20 groups the country was correctly detected using this procedure. In 13 of these remaining 19 groups the city was correctly detected. In 7 of these 13, the street was also correctly detected. The question about the country resulted in no false positives (out of 20), the question about the city resulted in two false positives (out of 19), while the question about the streets resulted in two false positives (out of 13). Furthermore, the two false positives at the city level also yielded a false positive at the street level. Even though effect sizes were only moderate, these results indicate that our dynamic questioning approach can help to unveil plans about a mock terrorist attack.

Keywords: lie detection, concealed information test, guilty knowledge test, searching concealed information test

INTRODUCTION

The Concealed Information Test (CIT; Lykken, 1959; Verschueren et al., 2011) uses physiological responding to determine the presence or absence of crime-related information in a suspect's memory. In a typical CIT, questions concern crime details known only to the perpetrator and the investigative authorities, but not to an innocent suspect. With each question, several answer options are presented serially, while peripheral autonomic nervous system activity is recorded. Answer options include the correct, but also several plausible but incorrect ones (e.g., "Was the victim dumped . . . (a) on a construction site, (b) in a pond, (c) on a beach, (d) in a dumpster, (e) in the trunk of a car"). For an innocent suspect, all options are equally plausible and will therefore elicit similar physiological responses. For a guilty suspect, the correct option is salient and significant, and will therefore elicit an enhanced orienting response (Verschueren et al., 2004). Such an orienting response is reflected by several psychophysiological responses, such as an increased skin conductance response (SCR; Lynn, 1966). Thus, a

consistent pattern of stronger responding to the correct options indicates knowledge of intimate crime details, from which guilt can be inferred.

Historically, the CIT has been used to infer guilt or innocence using information known to the investigative authorities. However, the CIT can also be employed when the correct option is not known, and the purpose of the investigation is to detect which of several options is the correct one. In this case, a series of options is presented to the suspect, and the option that evokes the largest physiological response warrants further investigation. This approach is often referred to as the Searching-CIT (S-CIT; Osugi, 2011) and can be used to discover, for example, the location of the body of a murder victim when the perpetrator is known (Nakayama, 2002). Applying the S-CIT to a terrorism scenario, Meixner and Rosenfeld (2011) asked 12 participants to choose a type of bomb, a location, and a date for a mock terrorist attack from a list, resulting in 36 to be detected details. Using a CIT based on the P300 component of the event related potential, they were

able to correctly identify 21 out of these 36 details, with no false positives.

Meijer et al. (2010) applied a variant of the S-CIT to a group of mock terrorism suspects. The idea behind this study was that the CIT and S-CIT are typically used to render a decision at the individual level. Yet many of today's security threats come from terrorist networks and organized crime. In these cases there may often be a group of people suspected of either planning or committing a crime. In Meijer et al. (2010), 12 participants were instructed to pretend they were members of a terrorist organization. They received information about the target, location, and date of an upcoming terrorist attack, and were then subjected to the CIT. An analysis at the group level showed that the correct option elicited a significantly larger average SCR, and as such information about an upcoming mock terrorist attack could be extracted from the group. Using a similar group approach but with a standard CIT, Bradley and Barefoot (2010), tested whether they could correctly identify exposure to one of three mock village scenarios. Groups of participants viewed tea making, bomb-making, or no activity, while building a card house. The CIT results showed that on the basis of group average SCRs, 80% of the bomb-making groups, and 75% of the tea making groups were correctly identified.

While Meijer et al. (2010) showed that the CIT can be used to elicit sensitive information from groups, their approach may be of limited applicability because the CIT format requires a limited number of plausible answer options. In some cases the number of available options may be naturally limited; while in others, the available options could be reduced by police work. Yet, the potential for real life application of the group variant of the S-CIT would be increased considerably if the content of test questions administered to suspects could be made contingent on their physiological responding to previous questions. For example, if the location of an upcoming attack is of interest, the first question could entail different countries, the next question could entail regions, then cities, etc. However, using series of questions requires an immediate feedback about which option evoked the largest mean physiological response. In the current experiment, we tested whether such an approach could be used to identify details of a mock terrorist attack. To enable immediate feedback, we performed an experiment in which we simultaneously measured skin conductance of groups of participants. These group data were analyzed immediately after each question, and the next CIT question presented was selected based on the responses to the previous question.

MATERIALS AND METHODS

PARTICIPANTS

Participants were 105 students of the University of Cologne, who received 10€ for their participation. Participants were tested in groups of five. Data of one group was discarded due to technical failure. Thus, the remaining sample consisted of 100 participants (28 men) with a mean age of 23.7 years ($SD = 3.66$).

All participants received written information about the procedure of the experiment before coming to the lab and read and signed a letter of informed consent before participating. The experiment was approved by the ethical committee of the Faculty of Psychology and Neurosciences of Maastricht University.

PROCEDURE

Once all five participants of a group arrived in a room located next to the laboratory, they were instructed by the experimenter to treat the experiment as a role playing game and imagine being members of a terrorist network whose job is to select a location for an attack. No reference to the type of attack was made. The group was informed that once they had selected their location, they would be subjected to a lie detection test, and their task was to try to conceal the information from the experimenter. The experimenter stressed that it was crucial to the study that everybody remembered their choice, and that they would be given a memory check after the test. Next, the experimenter instructed the group on how to select a location, and left the room. The group was given 10–15 min to make their selection.

The location of the attack consisted of a country, a city within this country, and a street within this city. First, the group had to open a sealed envelope labeled "Countries." This envelope contained a list with five European countries. Together, they had to decide on a country for their attack. Next, they opened a second envelope which contained five separate envelopes, one for each country. They opened only the envelope for the country they had chosen, and this envelope contained a list of five cities within that country. They chose one of these cities, and proceeded with opening the last envelope labeled with the city of their choice. This last envelope contained a list five streets in the chosen city, from which they selected one. This procedure was used to ensure participants were not exposed to the cities and streets that were not part of their chosen location¹. Once the group had selected their location, they listed it on a form signed by all members. This form served as the ground truth criterion. One member of the group held on to this form, and gave it to the experimenter at the end of the experiment. Thus, the experimenter was unaware of the details the group had chosen.

Once the group had completed the steps described above, they came to the testing room where the experimenter was waiting. Participants were seated in five cinema chairs facing a wall, and separated by room dividers so they could not see each other. Sensors measuring skin conductance were attached, and the S-CIT was performed. During the S-CIT, the experimenter was seated behind the cinema chairs. Upon completion of the CIT, the participants filled out a free recall memory check, and were thanked and paid for their participation.

SEARCHING-CONCEALED INFORMATION TEST

The S-CIT consisted of one example question and three test questions. The example question dealt with the day of the week ("Today is ... Monday ... Tuesday ... Wednesday ... Thursday ... Friday ...") and served to familiarize the participants with the procedure. Test questions referred to the country ("With this question we will determine in which country the attack will take place. Is it ...?"), the city ("With this question we will determine in which city the attack will take place. Is it ...?"), and the street ("With this question we will determine at which street the attack will take place. Is

¹This also means participants were exposed to all the correct and incorrect options of their plan. Research by Verschuere and Crombez (2008), however, showed that such previewing did not affect detection efficiency.

it . . .?"). Each question was presented for 10 s and followed by six options, each presented for 7 s. A random Inter Stimulus Interval ranging between 16 and 24 s was used. The first option presented within each question served as a buffer, and was excluded from all analyses. The following five options were presented in a random order. These five options were identical to the five options the group could choose from during the planning phase. Examples of options are France, The Netherlands, Belgium, Italy, and England for the countries, Reims, Bordeaux, Lille, Marseille, and Toulouse for the cities and Rue de Vesles, Rue Buirette, Rue de L'etape, Rue Carnot, and Rue des Murs for the streets. Obvious options such as capital cities and well known streets were avoided. Each question was repeated a number of times, depending on the outcome (see below). All stimuli were presented in a bimodal fashion; auditory via headphones; and visual text projected on the wall using a beamer. Each participant received a slider box, and was instructed to push this slider down with their right hand representing a "no" answer. This was done to encourage participants to focus their attention to the test. No data were, however, recorded from these slider boxes.

SKIN CONDUCTANCE MEASUREMENT, RESPONSE SCORING, AND ANALYSIS

Skin conductance was measured using dry electrodes with 1V DC system (Wild devine IOM), and sampled at 31 Hz. Sensors were placed on the tip of the index finger and the ring finger of the left hand of each participant. SCR's were defined as the maximum positive deflection in the 1–7 s window after stimulus onset. To eliminate individual differences in responsivity, the raw SCRs were transformed to a within-participants standard scores (Ben-Shakhar, 1985). Specifically, the SCR to each option was standardized relative to the mean and standard deviation of the SCRs across all five options within each question. Next, the z-scores for each option were averaged across the five participants, yielding a single z-score for each option. These z-scores were then averaged across repetitions.

The analysis described above was performed after each question, and the outcome was used by the experimenter to determine the next question to be presented. The following *a priori* rule was used to determine the choice of the follow-up questions. Each question was repeated twice. If after these two repetitions the average z-score of one option exceeded 0.4, this option determined the following question. If more than one option exceeded the 0.4 threshold, the option yielding the largest z-score was followed up. If no option exceeded the threshold, the question was repeated for a third time, and the option exceeding an average of 0.4 was followed up. If still no option exceeded the threshold, the test was stopped, and the verdict deemed "no decision."

RESULTS

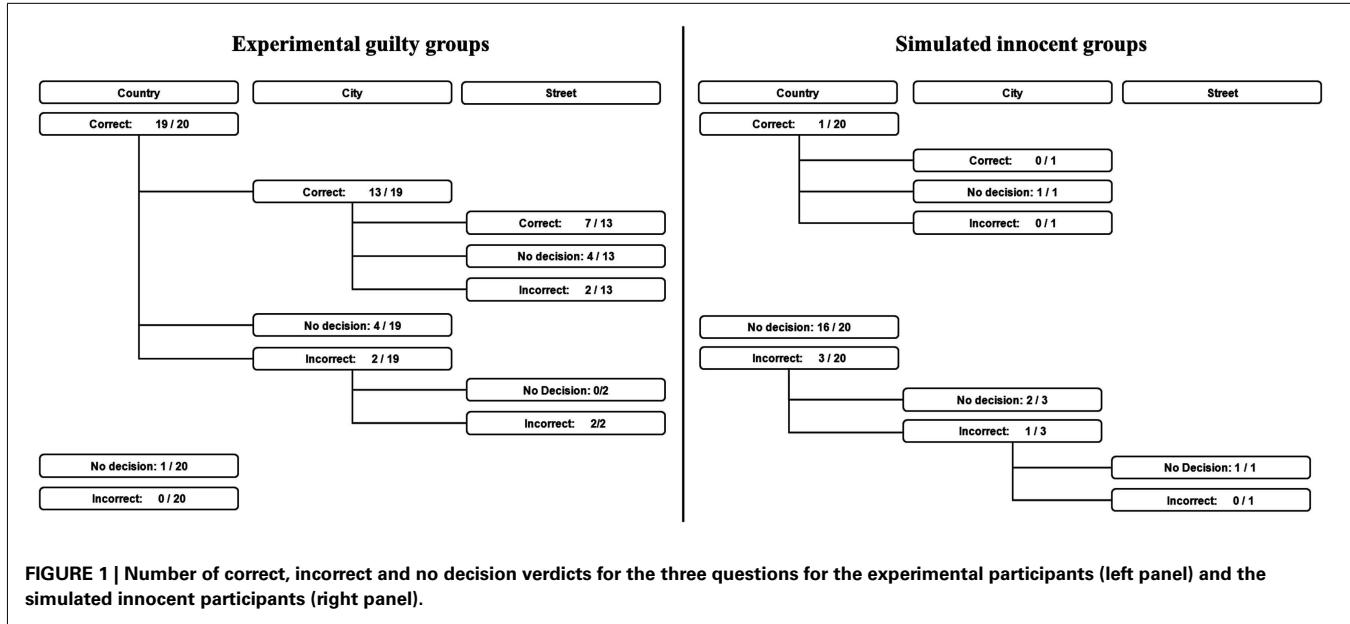
Correct recall on the memory check after the test was 100%. Average number of repetitions for question 1 was 2.05, for question 2 2.32, and for question 3 2.47. The results of the experimental groups are displayed on the left panel of **Figure 1**. The country was correctly identified in 19 of the 20 groups and in the remaining group no decision was made. The results of the second stage revealed that among the 19 groups for which the country was

correctly detected, the city was correctly identified in 13. In four groups, no decision was made, while in two an incorrect option exceeded the threshold. Among these 13 groups, the street was correctly identified in seven, while in four groups no decision was made, and in two an incorrect street name exceeded the threshold. In the two groups where an incorrect city was identified and consequently followed up, an incorrect street exceeded the threshold. When averaging over repetitions, the question containing the correct alternative was presented to a group in 52 cases (20 for the country, 19 for the city, and 13 for the street). In 39 of these cases (75%) the correct option was identified. In 9 cases (17.3%) a "no decision" verdict was rendered, and in 4 cases (7.7%) an incorrect option was identified. In the 2 cases where a question without the correct option was presented an incorrect option was identified. For 7 out of the 20 groups (35%), the correct location (country, city, and street) was successfully identified, for 9 groups (45%) a "no decision" verdict was rendered, while in 4 (20%) an incorrect location was identified.

To compare these results with outcomes that would be expected under a condition of chance level performance, we applied the simulation procedure outlined by Meijer et al. (2007). Adopting this procedure to the present data, we randomly drew five values from a standard normal distribution, representing one participant's responses to the five options of one question. These values were analyzed using the same steps used for the analysis of the experimental participants' data, i.e., each of these five values was standardized relative to the mean and standard deviation of all five values. This was repeated five times representing a group of five participants. Next, the standardized values were averaged across these five "participants," yielding a single z-value for each "option." This entire procedure was repeated representing repetitions, and the z-values were averaged across two "repetitions" if one of the values was greater than 0.4, and averaged over three "repetitions" if no value exceeded 0.4.

Repeating this simulation for 10,000 groups of five participants yielded a "no decision" verdict in 78.8% of all simulations, while in the remaining 21.2%, one option was identified. Among these 21.2% of the simulations, for which an option was identified, it was the correct one in 4.2%, and the incorrect one in 17%. These percentages are displayed on the right panel of **Figure 1** (as applied to 20 groups, rounded off to the nearest integer) such that they can be compared with the results obtained for the experimental groups. For example, while in 35% of the experimental groups the precise location (country, city, and street) was correctly identified such perfect identification was not obtained in any of the simulated groups.

To compare these results with those reported in other studies we computed the effect size based on individual data, using the ground truth criterion. This was done by simulating data to represent an innocent group that was matched to the experimental data, using the procedure outlined by Meijer et al. (2007). For example, the experimental data for question 1 (country) consisted of 95 participants (individual data for one group was not recoverable) of whom 90 were presented with two repetitions and 5 were presented with three repetitions. An innocent group consisting of the same number of participants and the same number of repetitions per participant was simulated. For each question



only the groups for which the correct option was actually presented were included. Cohen's d was calculated by subtracting the mean z -value of a randomly chosen option for the simulated data from the mean z -score to the correct option for the experimental participants and dividing this difference by the pooled standard deviation. Question 1 (Country; 95 participants) yielded an effect size of 1.12. Question 2 (City; 95 participants) yielded an effect size of 0.53. Question 3 (Street; 65 participants) also yielded an effect size of 0.53.

Finally, effect sizes (Cohen's d) comparing the group averaged z -values were computed, including only those groups for whom the question with the correct option was presented. Effect sizes were 2.70 for question 1 (Country; 20 groups), 1.67 for question 2 (City; 19 groups), and 1.66 for questions 3 (Street, 13 groups). To check for the effect of habituation, we also compared the group averaged z -values of the first and the second repetition within each question using paired t -tests. Only for question 2 was there a significant decrease in differential responding between the two repetitions [$t(18) = 2.27, p = 0.04$]. The decrease in differential responding between the repetitions in question 1 and 3 were not significant [$t(19) = 1.50, p = 0.15$ and $t(12) = 1.77, p = 0.10$, respectively]. Effect sizes of the group averaged z -values based on only the first repetition decreased to 2.15 for question 1, 1.36 for question 2, and 1.29 for question 3.

DISCUSSION

The goal of this experiment was to examine the possibility of applying a variant of the S-CIT to detect concealed information from groups of suspects using a sequence of questions, such that the content of a question is contingent on the physiological responding to the previous question. To enable immediate feedback, we collected skin conductance data simultaneously from multiple participants and analyzed the responses immediately following each question. Results showed that the precise location of a mock terror attack planned by the participants was correctly detected

in 35% of the groups, while in 20% an incorrect location was identified. The remaining groups (45%) rendered a "no decision" verdict.

Although the procedure performed above chance level, it led to a relatively high number of incorrect identifications. Two considerations are important here. First, it is important to realize that in the four groups where an incorrect location was identified, the information was still partially correct. In two groups the Country was correctly identified, while in the other two both the Country and the City were correctly identified. As such, the test did yield some information gain. Second, in contrast to criminal investigations where the CIT outcome typically addresses the guilt or innocence of a suspect, in the current application of the S-CIT, the costs of missing information about a planned terror attack outweigh the costs of incorrectly identifying a location. Even though due to the design of the current study, an incorrect identification also means missing the correct option, this may not be the case in a real life application, as the correct option may simply not be included. This justifies the use of a non-conservative cut-off point, yielding a relatively large false positive rate as done in this study. Yet, it is important to realize that other applications may warrant a different cut-off point than the one used in this study.

The relatively high number of false identifications and no decisions verdicts is not surprising given that the effect sizes at the individual level were only moderate. A number of explanations may account for this. First, the sensors used were dry electrodes, which may be less sensitive to changes in skin conductance. Secondly, the stimuli may have possessed relatively little signal value. Contrary to mock crime studies, where participants actually perform an act, in the current study participants were required just to pick options from a list. Needless to say, they were aware of the fact that they would not actually act out the scenario. In this sense the paradigm used here resembles the card test or the code words paradigms, which have been shown by Ben-Shakhar and

Elaad (2003) to yield effect sizes (1.35 and 1.16, respectively) of similar magnitudes to those obtained in this study for the first question (1.12).

The first question yielded a higher effect size than the second and third question. Several explanations may be offered for this finding. First, due to habituation differential responding to correct and incorrect options may have decreased over time. Yet, the analysis of the repetitions within each question did not yield strong support for this. Although there was a significant reduction in differential responding for question 2, the effect sizes of all three questions decreased when using only the first repetition, due to the increased standard deviation in the simulated innocent group. An additional explanation for the difference between the questions may be found in the work of Bradley and Janisse (1979). These authors used a standard CIT, with pupillary response as the dependent measure. By giving the participants fake feedback about the test's performance during previous trials, participants were led to believe that the test was either perfectly effective (100% accurate), somewhat effective (33 or 67% accurate), or perfectly ineffective (0% accuracy). Results showed that participants who were led to believe that the test was somewhat effective were easier to detect than those who were led to believe that the test was perfectly effective or perfectly ineffective. In the current experiment, in 19 out of the 20 groups, the cities presented with the second question were in the correct country. This may have served as feedback that the test was accurate to 95% of the participants, which, in line with the findings of Bradley and Janisse (1979) would explain the lower accuracy of the second and third question. Finally, the advantage of the first question over the subsequent questions may be explained in terms of differences in stimulus significance. Names of countries, may have been simply more salient and significant for the participants than names of cities and the streets.

Several limitations of this study deserve some attention. First, because the results were analyzed at the group level, all options need to be identical for all participants. As a consequence we did not check whether some of the items were personally relevant to some of the participants. But this will also characterize realistic situations. Secondly, as the experimenter was blind to what happened during the planning phase, we did not collect any data on social group interaction such as communication and compliance to the final decision. Future studies may incorporate such information, and, for example, test its influence on test efficiency. Finally, in the current experiment we used only guilty

REFERENCES

- Ben-Shakhar, G. (1985). Standardization within individuals: a simple method to neutralize individual differences in skin conductance. *Psychophysiology* 22, 292–299.
- Ben-Shakhar, G., and Elaad, E. (2003). The validity of psychophysiological detection of deception with the guilty knowledge test: a meta-analytic review. *J. Appl. Psychol.* 88, 131–151.
- Bradley, M. T., and Barefoot, C. A. (2010). Eliciting information from groups: Social information and the Concealed Information Test. *Can. J. Behav. Sci.* 42, 109–115.
- Bradley, M. T., and Janisse, M. P. (1979). Pupil size and lie detection: the effect of certainty on detection. *Psychol. A Q. J. Hum. Behav.* 16, 33–39.
- Breska, A., Ben-Shakhar, G., and Gronau, N. (2012). Algorithms for detecting concealed knowledge among groups when the critical information is unavailable. *J. Exp. Psychol: Appl.* 18, 292–300
- Lykken, D. T. (1959). The GSR in the detection of guilt. *J. Appl. Psychol.* 43, 385–388.
- Lynn, R. (1966). *Attention, Arousal and the Orienting Reaction*. New York: Pergamon.
- Meijer, E. H., Smulders, F., and Merckelbach, H. (2010). Extracting concealed information from groups. *J. Forensic Sci.* 55, 1607–1609.
- Meijer, E. H., Smulders, F. T. Y., Johnston, J. E., and Merckelbach, H. L. G. J. (2007). Combining skin conductance and forced choice in the detection of concealed information. *Psychophysiology* 44, 814–822.
- Meixner, J. B., and Rosenfeld, J. P. (2011). A mock terrorism application of the P300-based concealed information test. *Psychophysiology* 48, 149–154.

participants. One may argue that this does not represent realistic situations, where typically some suspects may be innocent and thus not possess any critical information. Thus, in reality the group tested may consist of both informed and uninformed suspects. Recently, Breska et al. (2012) tested the efficiency of two classes of algorithms for analyzing S-CIT data designed to detect critical information and differentiate between guilty and innocent examinees. The first class relied on a simple averaging procedure, while the second class relied on a PCA approach. They applied these algorithms on three data-sets of previous studies that used the standard CIT and demonstrated that in most cases the detection efficiency of both classes of algorithms was similar to that of the standard CIT. Moreover, the algorithms were relatively robust to the introduction of unknowledgeable participants in the sample. Such an analysis could also be applied with our dynamic questioning approach if only some participants possess the relevant knowledge.

The aim of our dynamic questioning approach was to increase the potential for real life application of the group variant of the S-CIT. Yet, due to the nature of the CIT format, even with this dynamic questioning approach the number of potential options needs still be limited somehow. Practically, this can be done by using intelligence gathered by investigative authorities. So it is important to note that even the dynamic questioning approach cannot be applied without at least some prior intelligence.

In sum, this study was a first attempt to use a dynamic questioning approach and despite the modest effect sizes obtained, and the finding that in only 35% of the groups tested the entire plan was correctly identified, we did demonstrate that this usage of the S-CIT can perform above chance level and yield important information gain. Moreover, even with the modest effect size of 1.12, the question referring to the Country of the attack yielded an impressive detection rate of 19 out 20 correct identifications. Although we can only speculate about the magnitude of the effect size to be expected in a field application, the bulk of available research indicates it will most likely be higher than the 0.53 obtained for the questions referring to the City and the Street (Ben-Shakhar and Elaad, 2003). We therefore believe that this approach deserves further research, for example with the use of multiple physiological and behavioral measures which can enhance detection efficiency.

ACKNOWLEDGMENTS

This study was funded by an NWO VENI grant (451-11-038), and a Golda Meir Fellowship awarded to the first author.

- Nakayama, M. (2002). "Practical use of the concealed information test for criminal investigation in Japan," in *Handbook of Polygraph Testing*, ed. M. Kleiner (London: Academic Press), 49–86].
- Osugi, A. (2011). "Daily application of the concealed information test: Japan," in *Memory Detection: Theory and Application of the Concealed Information Test*, eds B. Verschueren, G. Ben-Shakhar, and E. Meijer (Cambridge: Cambridge University Press), 253–275.
- Verschueren, B., Ben-Shakhar, G., and Meijer, E. (eds). (2011). *Memory Detection: Theory and Application of the Concealed Information Test*. Cambridge: Cambridge University Press.
- Verschueren, B., and Crombez, G. (2008). Deja vu! The effect of previewing test items on the validity of the concealed information polygraph test. *Psychol. Crime Law* 14, 287–297.
- Verschueren, B., Crombez, G., De Clercq, A., and Koster, E. H. (2004). Autonomic and behavioral responding to concealed information: differentiating orienting and defensive responses. *Psychophysiology* 41, 461–466.
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Received: 26 July 2012; accepted: 31 January 2013; published online: 15 February 2013.*
- Citation: Meijer EH, Bente G, Ben-Shakhar G and Schumacher A (2013) Detecting concealed information from groups using a dynamic questioning approach: simultaneous skin conductance measurement and immediate feedback. *Front. Psychology* 4:68. doi: 10.3389/fpsyg.2013.00068*
- This article was submitted to Frontiers in Cognitive Science, a specialty of Frontiers in Psychology.*
- Copyright © 2013 Meijer, Bente, Ben-Shakhar and Schumacher. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.*



The autobiographical IAT: a review

Sara Agosta¹ and Giuseppe Sartori^{2*}

¹ Center for Neuroscience and Cognitive Systems, Italian Institute of Technology, Rovereto, Italy

² Department of Psychology, University of Padua, Padua, Italy

Edited by:

Matthias Gamer, University Medical Center Hamburg-Eppendorf, Germany

Reviewed by:

Fren Smulders, Maastricht University, Netherlands
Kristina Suchotzki, Ghent University, Belgium

***Correspondence:**

Giuseppe Sartori, Department of Psychology, University of Padua, Via Venezia, 8, 35100 Padua, Italy
e-mail: giuseppe.sartori@unipd.it

The autobiographical Implicit Association Test (aIAT; Sartori et al., 2008) is a variant of the Implicit Association Test (IAT; Greenwald et al., 1998) that is used to establish whether an autobiographical memory is encoded in the respondent's mind/brain. More specifically, with the aIAT, it is possible to evaluate which one of two autobiographical events is true. The method consists of a computerized categorization task. The aIAT includes stimuli belonging to four categories, two of them are logical categories and are represented by sentences that are always true (e.g., I am in front of a computer) or always false (e.g., I am climbing a mountain) for the respondent; two other categories are represented by alternative versions of an autobiographical event (e.g., I went to Paris for Christmas, or I went to New York for Christmas), only one of which is true. The true autobiographical event is identified because, in a combined block, it gives rise to faster reaction times when it shares the same motor response with true sentences. Here, we reviewed all the validation experiments and found more than 90% accuracy in detecting the true memory. We show that agreement in identifying the true autobiographical memory of the same aIAT repeated twice is, on average, more than 90%, and we report a technique for estimating accuracy associated with a single classification based on the D-IAT value, which may be used in single subject's investigations. We show that the aIAT might be used to identify also true intentions and reasons and conclude with a series of guidelines for building an effective aIAT.

Keywords: implicit, associations, autobiographical memory, intentions, memory detection

Autobiographical memory is the ability to remember events that constitute part of one's life, such as directly experienced events. It is part of the episodic memory, which is, in turn, part of the long-term memory (Tulving, 1983). Available assessment methodologies of autobiographical memories focus on the subject's overall ability to recall past memorized events. For example, the Autobiographical Memory Interview (AMI; Kopelman et al., 1989) consists of a series of questions asking subjects to retrieve personal events related to a target concept. Most techniques for investigating this field are limited to the estimation of the individual/patient's capacity of recalling past autobiographical information rather than measuring the presence/absence of a specific autobiographical memory.

Methods for evaluating single autobiographical memories are limited to a few techniques such as the Guilty Knowledge Test (GKT; Lykken, 1959; Ben-Shakhar and Elaad, 2003) also known as Concealed Information Test (CIT). The GKT largely relies on the orienting response. In a typical GKT examination, participants, while undergoing the polygraph testing (physiological measurements), are shown a series of stimuli, including a salient one, related to a crime. When the stimulus related to the crime is shown, the subject can easily recognize it, thus producing an orienting reflex (e.g., skin conductance increase and heart rate deceleration). For a recent book on this technique, see Verschueren et al. (2011).

A new method that can be used to identify a true autobiographical memory, intentions and reasons that motivate an

act is the autobiographical Implicit Association Test (aIAT), a variant of the Implicit Association Test (IAT; Greenwald et al., 1998). Here, we will review all the published experiments on the aIAT so far. The traditional IAT (Greenwald et al., 1998) is a method for assessing the strengths of automatic associations. The method consists of a computerized task. Participants have to classify stimuli as quickly as possible in four different categories: two target concept categories (e.g., European American vs. African American names) and two attribute categories (pleasant vs. unpleasant) using two keys, one on the right and one on the left side of the keyboard. In one combined block, two categories (one from the target concept and one from the attribute dimension) are mapped on the same response key (e.g., European American names and pleasant words with the same key vs. African American names and unpleasant words with the other key). In a reversed combined block, participants have to classify the same four categories reversely paired (e.g., African American names and pleasant words with a key vs. European American names and unpleasant words with the other key), so that both target concept categories are paired with both attribute categories. The IAT effect is expressed as the difference between the combined and reversed combined blocks. In the block where two associated concepts require the same motor response, reaction times (RTs) will be faster than in the block where the same two concepts require different motor responses. Thus, the typical finding in this experiment is that, for European American participants, the stronger associated concept-attribute pair is the one coupling European

American names and pleasant words: This block should be easier to categorize than the one associating African American names and pleasant words. The reversed pattern is found for African American participants. The IAT has been extensively studied in social psychology to assess implicit beliefs, attitudes, and prejudices to measure self-esteem and self-concept (Nosek et al., 2007).

Clinical applications indicate that the IAT may be an effective technique to identify suicide-prone subjects, Pedophilia sexual orientation, doping, and personality assessment (Gray et al., 2005; Schmukle et al., 2008; Nock et al., 2010; Petròczi et al., 2010). Nock et al. (2010), for example, reported that the IAT might be useful in detecting suicidal ideations in people who attempted suicide. The authors documented that a high implicit association between self and death in suicide attempters is linked to a 6-fold risk increase in committing a suicide attempt in the next 6 months.

The aIAT (Sartori et al., 2008) is a variant of the IAT (Greenwald et al., 1998) that could be used to establish whether an autobiographical memory trace is encoded in the respondent's mind/brain. More specifically, with the aIAT, it is possible to evaluate which one of two autobiographical events is true.

The aIAT differs, for example, from the above European American/African American IAT as the evaluative dimension (pleasant/unpleasant) is substituted by a logical dimension (True/False), which is represented by sentences describing events that are certainly true (e.g., I am sitting in front of a computer) and certainly false (e.g., I am climbing a mountain). Furthermore, the target concept categories (e.g., European American/African American) are represented by sentences describing alternative versions of an autobiographical event (e.g., I went to Paris for Christmas vs. I went to New York for Christmas), only one of which is true. The true autobiographical event is identified because, in a combined block, it gives rise to faster RTs when it shares the same motor response with true sentences. If the participant spent his/her vacation in Paris, the block associating true sentences and sentences related to Paris will be faster than the block associating true sentences and sentences related to New York.

The aIAT is structured in five blocks, three simple blocks (1, 2, 4), and two combined categorization blocks (3 and 5). In simple blocks, each response button is used to classify sentences related to only one category. In double blocks each response button is used to classify sentences related to two different categories.

In Block 1, participants have to classify true and false sentences (e.g., I am in front of a computer vs. I am in front of a television) using two response keys, one on the left and one on the right of the keyboard. In Block 2, participants have to classify autobiographical sentences (e.g., I went to Paris for Christmas vs. I went to New York for Christmas) with the same two response keys. In Block 3 (double categorization block), true sentences and sentences related to the first autobiographical event (e.g., Christmas in Paris) are paired on the same response key and false sentences and sentences related to the second autobiographical event (e.g., Christmas in New York) are classified with the other response key. In Block 4, only autobiographical events are reversely classified with the two response keys. Finally, in

Block 5, participants have to classify both true sentences and sentences related to the second autobiographical event (Christmas in New York) with the same response key, and false sentences and the first autobiographical event (Christmas in Paris) with the other key.

The aIAT/IAT effect is expressed in terms of average RT difference between the two double categorization blocks: the congruent block (pairing the two associated categories) and the incongruent block (pairing the non-associated categories).

Used as a memory detection technique, the aIAT has a number of advantages related to the use of reaction times (Seymour et al., 2000), when compared to traditional psychophysiological techniques of lie detection (e.g., Ben-Shakhar and Elaad, 2003) or fMRI-based lie detection strategies (e.g., Langleben et al., 2005). For instance, it can be administered quickly (10–15 min), it is based on an unmanned analysis (no training for the user is necessary), it requires low-tech equipment (a standard PC is sufficient), and it can be administered remotely to many participants (e.g., via the internet).

DETECTION OF AUTOBIOGRAPHICAL MEMORIES: A REVIEW OF VALIDATION STUDIES

The aIAT accuracy in identifying the true memory has been investigated in a series of validation experiments summarized in **Table 1**. In this table, we separated first from second administration of an aIAT. Here, in order to evaluate the accuracy of the method, we included only experiments that did not include negative statements as subsequent investigations (Agosta et al., 2011c), conducted after the original publication (Sartori et al., 2008), indicated that the use of negative sentences or reminder labels generates unreliable and inaccurate results. For this reason, the following experiments were excluded:

1. Mock crime experiment (experiment 2) in Sartori et al. (2008).
2. Cocaine/heroin experiment (experiment 3) in Sartori et al. (2008).
3. Driving license experiment (experiment 5) in Sartori et al. (2008).
4. Control conditions (innocent or guilty) of experiments 1 to 3 reported by Verschuere et al. (2009).
5. Experiments 1 to 4 in Agosta et al. (2011c) used to verify the accuracy of the aIAT using negative statements.

Moreover, data used to calculate the accuracy refer to administrations of the aIAT prior to or without manipulations (faking, training, EEG-required-modifications of stimulus presentation) and for this reason we decided to exclude:

1. Naive faking and experienced faking groups in experiments 1–4 described in Agosta et al. (2011b).
2. Faking conditions of experiments 1–3 in Verschuere et al. (2009).
3. Intention aIAT combined with EEG (experiment 3 in Agosta et al., 2011a).
4. Second administration of the practice, instruction and training groups in Hu et al. (2012).

Table 1 | In this table, the results from all the validation experiments are summarized.

Experiment	Source	Number of subjects	Participants classified correctly using the D-IAT	Average IAT effect	Average D-IAT	Confidence interval 95%
FIRST ADMINISTRATION						
Card alAT	Sartori et al., 2008	37	35/37 AUC = 0.99	316 ms	0.56	0.45–0.67
Holiday alAT	Sartori et al., 2008	20	18/20	219 ms	0.44	0.25–0.63
Christmas holiday alAT (non-faking group)	Agosta et al., 2011b	14	14/14	955 ms	1.06	0.84–1.28
Mock crime alAT (affirmative sentences; first alAT administered)	Agosta et al., 2011c	40	35/40	297 ms	0.56	0.41–0.71
Intention alAT (Experiment 1)	Agosta et al., 2011a	22	22/22	712 ms	1.16	1.00–1.33
True memory alAT (first alAT administered)	Marini et al., 2012	18	18/18	879 ms	1.02	0.89–1.15
Flashbulb alAT Experiment 1 = outlier	Lanciano et al., 2012	42 14	Experiment 2 42/42 Experiment 1 14/14	Experiment 2 876 ms Experiment 1 1082 ms	Experiment 2 1.48 Experiment 1 3.87	Experiment 2 1.86–1.10 Experiment 1 2.8–4.93
White lies alAT (first alAT administered)	Agosta et al., 2013	20	20/20	444 s	0.55	0.42–0.68
Reasons alAT* (first alAT administered)	Agosta et al., 2013	20	20/20	309 ms	0.46	0.37–0.55
Mock crime alAT	Hu and Rosenfeld, 2012	12 + 12 + 12 = 36	Immediate guilty: 10/12 Innocents: 9/12 Delayed guilty: 7/12	Immediate guilty: 92 ms Innocents: 69 ms Delayed guilty: 82 ms	Immediate guilty: 0.23 Innocents: 0.32 Delayed guilty: 0.32	Immediate guilty: -0.08–0.54 Innocents: 0.13–0.51 Delayed guilty: 0.14–0.50
Mock crime Pretest repetition group	Hu et al., 2012	16	16/16 AUC = 0.98	121 ms	0.52	0.39–0.65
Mock crime Pretest practice group	Hu et al., 2012	16	13/16 AUC = 0.91	114 ms	0.46	0.29–0.63
Mock crime Pretest instruction group	Hu et al., 2012	16	13/16 AUC = 0.95	103 ms	0.47	0.28–0.66
Mock crime Pretest training group	Hu et al., 2012	16	15/16 AUC = 0.98	94 ms	0.51	0.35–0.67
Action alAT (Imagined + not imagined)	Takarangi et al., 2013	79	77/79	Not reported	0.585	0.53–0.64

(Continued)

Table 1 | Continued

Experiment	Source	Number of subjects	Participants classified correctly using the D-IAT	Average IAT effect	Average D-IAT	Confidence interval 95%
Average	17 aIATs	412	92% accuracy		0.58	0.41–0.73
SECOND ADMINISTRATION						
Christmas holiday aIAT (non-faking group with previous aIAT experience not reported in the paper as was only considered a practice)	Agosta et al., 2011b	20	19/20	445 ms	0.64	0.48–0.80
Two cards aIAT (non faking group with previous aIAT experience not reported in the paper as was only considered a practice)	Agosta et al., 2011b	12	11/12	236 ms	0.45	0.26–0.64
Ten cards aIAT (non-faking group with previous aIAT experience not reported in the paper as was only considered a practice)	Agosta et al., 2011b	20	20/20	684 ms	1.13	1.03–1.22
Mock crime aIAT (affirmative sentences; second aIAT administered)	Agosta et al., 2011c	40	35/40	220 ms	0.45	0.33–0.57
True memory aIAT	Marini et al., 2012	18	18/18	606 ms	0.87	0.80–1.08
White lies aIAT (second aIAT administered)	Agosta et al., 2013	20	20/20	280 ms	0.45	0.34–0.56
Reasons aIAT (second aIAT administered)	Agosta et al., 2013	20	19/20	266 ms	0.50	0.39–0.61
Mock crime (Repetition group; second aIAT administered)	Hu et al., 2012	16	15/16	92 ms	0.41	0.28–0.54
Average	8 aIATs	166	94% accuracy		0.67	0.48–0.87

For each experiment, the number of participants together with average D-IAT values are reported. First administrations have been separated from second administrations of an aIAT.

*White lies and Reasons aIATs have been administered to the same participants, but have been included in this analysis not fulfilling the criteria for a systematic review. When excluding the Reason aIAT (second IAT administered to the same subjects), weighted average D-IAT is 0.59 for the first administration and 0.70 for the second administration. As shown, when eliminating the same subjects from analysis there are no substantial changes in the effect size. For each experiment, the number of participants together with average D-IAT values are reported. First administrations have been separated from second administrations of an aIAT.

Repetitions of aIAT administrations to participants were only included in the analysis if there were no manipulations in between. Thus, in **Table 1**, we only report data from participants who either completed only one aIAT or two aIATs without manipulations in between.

In all the experiments, the validity of the aIAT was tested against a known false event. For example, in the card experiment, a card, which was actually chosen by the participant, was compared to the non-selected card. In the autobiographical memory experiment, a real autobiographical event, as assessed through a preliminary questionnaire, was compared to a false event. For this reason, we excluded:

1. Experiment 2 in Agosta et al. (2011a) because evaluating the difference between intentions and hopes and not presenting two contrasting events or intentions (i.e., hopes are true as well as intentions).
2. False memory aIAT in Marini et al. (2012) because comparing two actual false events (one believed to be true).

Two measures can be used for evaluating the diagnostic accuracy: the magnitude of the IAT effect (RTs of the incongruent block minus the RTs of the congruent block) and the D-IAT value (D600; Greenwald et al., 2003). Here, we focused in particular on the D-IAT value. This index combines speed of response and

classification accuracy. It includes a penalty for errors and variability. It expresses the difference in the mean latencies of the double categorization blocks scaled by the standard deviation of response latencies. It is calculated by subtracting corrected mean RTs of the congruent block from corrected mean RTs of the incongruent block and dividing this difference by the inclusive standard deviation for the two blocks.

Effect size was the average D-IAT value. To calculate an average effect size across all the studies, the D-IAT values were weighted by the inverse variance in order to deal with the different and small sample sizes of each study (Lipsey and Wilson, 2001). The only outlier (Flashbulb aIAT, Experiment 1; Lanciano et al., 2012) was identified using the interquartile range and was not included in the calculation of the mean effect size.

For the first-administration studies (17), homogeneity among study results was evaluated using Cochran's Q combined with the I^2 statistic. Cochran's Q value had to be compared to a chi-square distribution with $k-1$ (number of studies -1) degrees of freedom. In our case, it resulted in a value of 9.26, below the critical value for 16 degrees of freedom in a chi-square distribution (26.3). This value indicated low heterogeneity. The interpretation of the I^2 statistic was made following Higgins and colleagues' directions (Higgins et al., 2003) with values of 25% representing low heterogeneity, 50% moderate heterogeneity, and 75% high heterogeneity. Our I^2 is equal to 0%. D-IAT values were combined to obtain a mean effect-size using a fixed-effect approach because of the low heterogeneity. D-IAT average value resulted in 0.57 (95% C.I. 0.41–0.73).

For a total of 8 s administration studies, Cochran's Q was 8.51 (<14.1–7 degrees of freedom) and the I^2 was 0%. Again, we used a fixed-effect model for calculating the mean effect size of 0.67 (95% C.I. 0.48–0.87).

Weighted average D-IAT for the first administration was 0.57, while for the second-administration studies it was 0.67. More studies are needed in order to investigate the effect of repetition of an aIAT. Indeed, in the studies reported here, the same aIAT has never been repeated twice.

To determine the accuracy of the test, we used the direction of the D-IAT values, calculated by subtracting the congruent block from the incongruent one, with negative values indicating an incorrect classification (i.e., the identification of the false memory as true) and positive D-IAT values indicating the correct identification of the true memory.

Accuracy was also calculated across a total of 412 first administrations of the aIAT to participants ($Q = 4.7 < 26.3$; $I^2 = 0\%$). The weighted average classification accuracy was 92% (95% C.I. 83–100%). Across a total of 166 s administrations ($Q = 0.41 < 14.1$; $I^2 = 0\%$), the weighted average accuracy was 94% (95% C.I. 80–100%). Clearly, repetition of the aIAT does not decrease the overall accuracy.

In this small review, we mainly included experiments from the same laboratory. Importantly, in **Table 1**, we have also included data of five mock-crime experiments from two other laboratories (Hu and Rosenfeld, 2012; Hu et al., 2012; Takarangi et al., 2013). These data include preliminary aIATs administered to four groups of participants that were subsequently tested with a variety of manipulations between test and retest, and data on

performed and non-performed actions. Finally, we added data from an associated laboratory (Lanciano et al., 2012).

Recently, a modified version of the IAT/aIAT has been used in order to distinguish between seen and unseen events (eyewitness—Implicit Association Test—eIAT; Freng and Kehn, 2012). The authors tested a total of 18 participants and showed that the eIAT “successfully distinguished between witnessed and non-witnessed details” of a video. In particular, they reported that central and peripheral details of a scene were efficiently identified (central details; $D = 0.5$, peripheral details $D = 0.42$). These data have not been included in **Table 1** because of a lack of details in the text (i.e., average reaction times of congruent and incongruent blocks, accuracy of the D-IAT values in identifying the eye-witnessed event). Results of this experiment show that the aIAT cannot only be used to identify episodic memory of an own action, but also an observed event.

OVERALL ACCURACY AND ACCURACY AS A FUNCTION OF D-IAT

The D-IAT value measures the strength of the IAT effect combining both RTs and errors. The D-IAT value used as classification criterion yields correct classifications in more than 90% of the cases, with a weighted average value of 0.58 for first-administration studies and 0.67 for second-administration studies.

When analysing the relation between classification accuracy and D-IAT values, we found that it varies depending on D-IAT values. For D-IAT values just above zero, classification accuracy is just above 50%, while for D-IAT values larger than 0.6, the classification is almost 100% (please refer to **Figure 1**).

Figure 1 was drawn as follows:

1. Data were used from eight previous validation experiments [first-administration aIAT only and limited to experiments conducted in our research group: Card aIAT in Sartori et al. (2008); Holiday aIAT in Sartori et al. (2008); Christmas holiday aIAT in Agosta et al. (2011b); Mock crime aIAT in Agosta

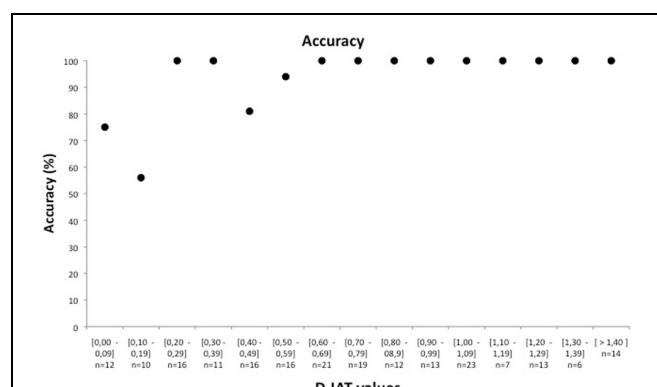


FIGURE 1 | Classification accuracy as a function of the D-IAT value.

Data from eight validation experiments, for a total of 209 subjects, were used to calculate accuracy in identifying the true autobiographical memory on the basis of the D-IAT value. D-IAT values have been grouped in bins of 0.1. In the Y axis, the number of participants for each bin is reported.

et al. (2011c); Intention aIAT in Agosta et al. (2011a); True and false memory aIAT in Marini et al. (2012); White lies aIAT in Agosta et al. (2013); Reasons aIAT in Agosta et al. (2013)] for a total of 209 subjects.

2. For each participant, the D-IAT value was used with information about whether the classification of the target true autobiographical memory was correct (1) or wrong (0).
3. Subjects were ordered for increasing values of absolute D-IAT value.
4. D-IAT values were grouped in bins of 0.1.
5. For each D-IAT value group, the corresponding accuracy was calculated (different number of participants for each D-IAT value group).
6. Each black dot on the figure represents the corresponding accuracy of a specific D-IAT value group.

This D/accuracy figure highlights the close relationship between accuracy and D-IAT value. First, it is important to note that only for a few D-IAT values is the accuracy lower than 0.8 (80% correct classifications), and most of the values with the lower accuracy are included in the window between 0 and 0.2. For this reason, we would advise considering any D-IAT value from 0 to 0.2 as inconclusive. Across the total of 209 subjects, 10% showed an inconclusive result. Moreover, the figure highlights the fact that D-IAT values greater than 0.6 are always classified correctly and, more importantly that the majority of the D-IAT values have 100% accuracy. This D/accuracy function could help in estimating the probability of correct classification depending on the individual test result, thus increasing the confidence of the technique when making inferences on a single test.

An important issue in clinical and forensic single-case investigations is the estimation of the validity of the test results. In short, if a subject's D is equal to 0.43 from **Figure 1**, we would expect that his/her result is in the 0.4–0.49 range and has an average accuracy of 81%.

RELIABILITY OF aIAT

SPLIT-HALF RELIABILITY

Ideally, a good memory detection technique should identify the same true memory from different subsets of items. This feature is assessed with the split-half technique.

aIAT split-half reliability has been computed after separating odd and even stimuli and then deriving, for each test, two D-IAT values. Data were calculated over a subgroup of the previous validation experiments: the first-administration studies. The main result indicates an average 88% of agreement in the identification of the true autobiographical memory (correct or incorrect classification of the subject on the basis of the D-IAT value), of even and odd stimuli (please refer to **Table 2**).

Correlations of the D-IAT values, calculated separately for odd and for even trials, resulted in an average split-half value of $r = 0.52$, with a low correlation between even and odd stimuli in the “Intention aIAT.” There are no apparent reasons for this low correlation, but the agreement in identifying the true autobiographical memory is 90%. Thus, even if the correlation of the D-IAT values is low, both values, derived from the even

Table 2 | Split-half correlation, percentage agreement between classifications derived on even numbers and classification derived from odd numbers in five experiments.

Experiment	Agreement %	Split-half correlation
Card aIAT (Sartori et al., 2008)	73	$r = 0.47 p < 0.003$
Mock Crime (Agosta et al., 2011c)	85	$r = 0.48 p < 0.002$
Christmas holiday aIAT (non-faking group, Agosta et al., 2011b)	100	$r = 0.67 p < 0.006$
Intentions (Experiment 1, Agosta et al., 2011a)	90	$r = 0.17 p = 0.465$
White lie aIAT (Agosta et al., 2013)	90	$r = 0.83 p < 0.001$

and odd stimuli, result in a comparable identification of the true autobiographical memory.

ORDER OF PRESENTATION OF THE CONGRUENT BLOCK

In order to establish if there is an agreement between the results obtained with different orders of the congruent and incongruent blocks (3rd and 5th positions), we also analysed the correlation of D-IAT values of the same aIAT with the congruent block either in the 3rd (direct order) or the 5th position (reversed order), and consequently, the incongruent block in the 5th or the 3rd position. Two experiments (**Table 3**) in which participants were administered both orders (direct and reversed), taken from the previous validation table, were used for this analysis: the “Mock crime” aIAT reported in Agosta et al. (2011c) and the “White lie” aIAT (Agosta et al., 2013). All the participants in the two experiments were administered two aIATs: one in the direct and one in the reversed order. In the “Mock crime” aIATs, the order of presentation of the two aIATs was counterbalanced across subjects, while in the “White lies” aIAT (Agosta et al., 2013) the first aIAT always had the congruent block in the third position.

Results indicated that the agreement in the identification of the true autobiographical memory for the direct and reversed orders (on the basis of the direction of the D values) was high: 95% and 85% for the “White lie” and “Mock crime” experiments, respectively.

Moreover, the correlation of D-IAT values (for the direct and reversed orders) was 0.15 for the “White lie” and 0.63 for the “Mock crime” experiment. For the White lie experiment, as for the Intention experiment presented in the previous section, we do not have an explanation for this low correlation, but the level of agreement is high and there is no reduction in the identification of the true memory.

FACTORS REDUCING ACCURACY AND MODULATING THE aIAT

Further research was conducted in order to highlight the limitations in the use of aIAT. Specifically, the effect of faking, of using

negative sentences and negative labels, has been investigated. The results are summarized below.

EFFECTS OF FAKING ON MEMORY DETECTION

Verschueren et al. (2009) have shown that properly trained participants may alter the test outcome strategically. Participants may be trained to alter the test outcome by speeding up the incongruent blocks and slowing down the congruent block. Verschueren et al. (2009) instructed the guilty participants in a mock-crime task to appear as “innocents” by slowing down their responses. Their results indicated that a big percentage of the guilty participants not previously exposed to the aIAT succeeded in faking the test, but only when explicitly taught the strategy to counterfeit the test outcome. These results were further refined by Agosta et al. (2011b) who showed that: (i) instructed fakers (explicitly instructed by the experimenter to succeed in altering the test outcome) may alter the test outcome by making a false memory appear true and vice versa and (ii) fakers may be distinguished from non-fakers on the basis of an algorithm that compares response speed in simple blocks with response speed in double blocks. Their results are summarized in the **Table 4**.

In short, a non-trained subject instructed to fake, but using self-discovered strategies, does not often succeed in his/her attempt. By contrast, when previously trained on the best strategy to fake (e.g., speed up incongruent block and slow down the congruent block), examinees can alter their results and beat the “memory detector.” However, these successful fakers may be detected on the basis of their response pattern through

a faking-detection algorithm. This algorithm is based on a comparison of the average speed in double and single blocks. Indeed, participants leave a signature when trying to fake the test: They do not alter their RTs in single blocks and are abnormally slow in double categorizations blocks (Agosta et al., 2011b). This feature has been used with high accuracy (83%) to detect fakers. The more efficient algorithm for detecting fakers consists of three steps: (i) remove all responses below 150 and above 10,000 ms, (ii) replace errors with the average RT of the block with a penalty of 600 ms, and (iii) calculate the ratio between the average RT of the fastest block (between 3 or 5) and single tasks that are directly connected to the fastest task in terms of motor response (1 and 2 or 1 and 4, respectively). If the result exceeds 1.08, then the respondent is faking. This cut-off was identified as the one yielding the maximal classification accuracy in our sample.

Hu et al. (2012) investigated this same issue. The authors confirmed that specific instructions given to the subjects might be effective in altering the aIAT results. Furthermore, they showed that this pattern of results might be further enhanced with a specific training in the incongruent trial. Thus, they reported that instructions and training together are more effective than instructions alone in reversing the results compared with a pre-test. In their experiment, they reported failing to find a significant difference between fakers and non-fakers using the previously described indexes. Those results highlight the need for an in-depth investigation of this important issue. Only two studies have been published so far on the possibility of identifying fakers with non-consistent results.

Table 3 | Correlation and agreement of D-IAT values and IAT effect for normal (congruent block in 3rd position and incongruent block in 5th position) and inverted (congruent block in 5th position and incongruent block in the 3rd position) orders.

Experiment	3–5	Agreement D-IAT	Correlation IAT effect	Correlation D-IAT
Mock crime (Agosta et al., 2011b)	Order counterbalanced across participants	85%	$r = 0.25$ $p < 0.11$	$r = 0.63$ $p < 0.001$
White lies (Agosta et al., 2013)	Order fixed with the first aIAT administered having the congruent as block 3 and the second aIAT administered having the congruent block as block 5	95%	$r = 0.25$ $p < 0.28$	$r = 0.15$ $p < 0.53$

Table 4 | Data from four experiments comparing control non-fakers, naïve fakers, and instructed fakers are reported (Agosta et al., 2011b).

Experiment	Non-fakers D-IAT	Non-fakers correct classifications	Naïve fakers D-IAT	Naïve fakers correct classifications	Instructed fakers D-IAT	Instructed fakers correct classifications
1	1.06	14/14	0.78	14/14	-0.45	5/14
2	0.64	19/20	0.24	6/10	-0.42	7/20
3	1.13	20/20	0.82	18/18	-0.81	4/34
4	0.45	11/12	0.15	6/12	0.06	7/12

Control non-fakers were administered the test without specific instructions, and naïve fakers were instructed to alter the results but were not taught the more efficient strategy. Instructed fakers were instructed to alter the results by speeding up in the incongruent trial and slowing down in the incongruent trial.

EFFECTS OF NEGATIVE SENTENCES AS DESCRIPTORS OF AUTOBIOGRAPHICAL EVENTS ON aIAT ACCURACY

False memories may be described by using a negative description of the true memory. Agosta et al. (2011c) have shown that, when affirmative sentences and reminder labels are used to describe the true and false autobiographical events, accuracy is very high at up to 90% (Agosta et al., 2011c). By contrast, in four studies, the authors (Agosta et al., 2011c) showed that when negative sentences and labels are used, there is a reduction of about the 30%, in the accuracy of the aIAT in identifying the true autobiographical event. The accuracy of the aIAT is reduced not only by negative sentences, but also by affirmative sentences describing counter-events. The affirmative counter-event sentences were stated with expressions such as *different place from* instead of the negative (e.g., "I have been to Rome," vs. "I have been to a different place than Rome"). Negative and affirmative counter-event sentences can be considered as equivalent from this point of view. Counter-event sentences show a more difficult grammatical structure than simple negative sentences and, at the same time, have a negative inner meaning (e.g., having been in a different place than Rome means not having been in Rome). Those might be plausible reasons for the aIAT's low accuracy when using counter-event sentences. The use of negatives renders the test highly inaccurate and should therefore be avoided.

aIAT APPLICATION TO FLASHBULB AND FALSE MEMORIES

FLASHBULB MEMORIES

For many years, researchers have debated whether flashbulb memories (FBMs) can be considered either as a special class of accurate emotional memories that are exceptionally vivid and resistant to decay (Pillemer, 1984; Bohannon, 1988; Conway et al., 1994) or as memories affected by reconstructive factors such as ordinary autobiographical memories. The controversial debate concerning the real existence of this special class of memories reflects the difficulty in establishing the accuracy of these autobiographical formations.

FBMs are usually recalled with a higher degree of confidence than other autobiographical memories (Brown and Kulik, 1977; Weaver, 1993; Talarico and Rubin, 2003, 2007, 2009). It is interesting to note that the participants' confidence does not decrease even when it is clear that the recalled event had not occurred in the same way as it is remembered (Neisser and Harsch, 1992). Indeed, according to some authors, what makes FBMs so unique and special is the individual's sense of confidence in his/her accuracy, which is preserved for a long time after the occurrence of the original eliciting event (Weaver, 1993; Talarico and Rubin, 2003, 2007, 2009).

Lanciano et al. (2012) have investigated the specific characteristic of FBM by asking 14 participants to fill out a questionnaire concerning the death of Pope Johannes Paulus II. On average, subjects were tested 2235 days after the pope's death. The questionnaire investigated seven FBM attributes: (1) date when the individuals learned of the pope's death, (2) day, (3) time of the day, (4) informant (family, friends, colleagues, media), (5) location (country, city, room, or other kind of location, i.e., the car), (6) presence of other people, and (7) ongoing activity. An aIAT contrasting the true memory with a fabricated false memory was

administered to participants 1 week later. All 14 participants were correctly classified using the D-IAT values. Average D was 3.85, which is a very high value compared to other typical values as reported in Table 1. Consistency among repeated measures of FBM is a typical parameter describing the quality of this sort of memory. The authors reported a high correlation of 0.85 between consistency value and D-IAT values at the aIAT. In short, the more consistent the FBM is among repetitions, the higher the D-IAT value is observed.

FALSE MEMORIES

It is known that human memory is prone to various kinds of distortions and illusions (Roediger, 1996; Schacter, 1999; Loftus, 2003). It has been shown that, in contrast to deception, memory illusions are often not accompanied by a subjective feeling that people are responding untruthfully. Quite the contrary, memory illusions like those produced in the Deese-Roediger-McDermott paradigm (DRM; Deese, 1959; Roediger and McDermott, 1995) are accompanied by a sense of recollection that, at the conscious phenomenological level, makes them indistinguishable from true memories. DRM false memories are obtained by presenting lists of words related to a non-presented critical lure. The probability of recalling and recognizing the critical lure is usually quite high (Roediger and McDermott, 1995; Balota et al., 1999; Stadler et al., 1999; Budson et al., 2002). Previous findings have shown that critical lures seem to elicit the same quality (i.e., remember judgments) of presented items (e.g., Roediger and McDermott, 1995), and participants are even able to state in which voice they heard the non-presented critical lure when half of the list items had been presented by a female voice and half by a male voice (Payne et al., 1996).

In this study, Marini et al. (2012) used a standard DRM task to induce false memories, followed by the two aIATs. By comparing the results of the two aIATs, one could observe whether participants were responding differently to true and false DRM memories. One aIAT compared presented items with non-presented distractors (aIAT true memories), whereas the second aIAT (aIAT false memories) compared critical lures with non-presented distractors. Specifically, the aIAT true memories evaluated the association of the presented items with the true logical dimension, while the aIAT false memories evaluated the association of the critical lures with the true logical dimension. Therefore, if true memories (presented items in the aIAT true memories) and false memories (critical lures in the aIAT false memories) were encoded differently, as suggested by neuroimaging studies (Cabeza et al., 2001; Slotnick and Schacter, 2004), they would have a different strength in their association with the true logical dimension. If, however, the aIAT is based on the individual's "aware" belief that the critical lure is indeed present, then the aIAT would be ineffective in detecting any difference between presented items and critical lures. Results indicated that false memories are strongly associated with true sentences (36/36 participants), giving rise to similar associations as true memories with true sentences.

This result indicates that the aIAT reflects exactly what is stored in our memory, and if a memory is strongly believed to be true, then the aIAT would identify it as a true memory. An interesting issue that stems from the false-memory work concerns its applied

implications: Does the aIAT always identify a true memory when this is strongly believed to be true? Does the self-persuasion of a false memory as true influence the result of the aIAT? All these issues have to be investigated in more detail in future studies.

It has been shown that false memories may stem from “source confusion” (Takarangi et al., 2013), which is defined as “the attribution of a specific memory to a particular source using heuristics that may lead to errors.” Takarangi et al. (2013) reported an experiment aimed at verifying the aIAT diagnostic abilities in detecting whether an action was performed or not. After asking their participants to perform or not to perform an action, the authors further required them to imagine both performed and non-performed actions.

They reported an overall aIAT accuracy of 97.5% in detecting whether the action was performed or not, confirming the efficiency of aIAT in identifying memories of performed actions when contrasted to memories of non-performed actions.

Importantly, the experimental design allowed the computation of a source discrimination score derived by subtracting ratings for non-performed actions from ratings of performed actions (the authors asked the participants to rate how much they believed that they had performed the action, and then rated how much they remembered performing the action). They found that imagining an action increased the subjective trend of believing and remembering actions as performed rather than non-performed actions. They also found that the D-IAT value diminishes with the source discrimination score. In short, the more the memories are subjectively confused (acted vs. not acted) by the subject, the lower the D-IAT. The authors claim that this is a limitation of the aIAT when it is required to identify false memories. However, a close inspection of Figure 2 in their paper shows that only two of 79 subjects were misclassified and this indicates that, even if D-IAT is affected by source confusion, this did not increase misclassification in their study.

DETECTION OF INTENTIONS

Deliberation of a future action is called prior intention in one terminology (Searle, 1983). Prior intentions include goal-related processing and deliberative conscious intentions that are intuitively believed to be the leading cause of our future behaviors (Bratman, 1987; Cohen and Levesque, 1990). In other words, these are mental representations that occur prior to the action itself and are typically believed to cause the action subjectively. Searle (1983) refers to prior intentions as the initial representation of the goal of an action prior to the initiation of the action: a type of intention that is formed in advance of a deliberate plan

for a future action. In contrast, an intention in action (also termed motor intention) is the proximal cause of the physiological chain leading to an overt behavior.

Other scholars have addressed a possible distinction between long-term antecedents of action (prior intentions; Searle, 1983) and short-term antecedents of actions (intentions in action; Searle, 1983; Becchio et al., 2008, 2010). Long-term antecedents have also been named “prospective intentions” (Pacherie and Haggard, 2010), “distal intentions” (Pacherie, 2008), or “future-directed” intentions (Bratman, 1987).

An experiment showing that intentions to act may be identified reliably with the aIAT will be summarized here (please refer to Table 5). Agosta et al. (2011a) have investigated whether real intentions could be distinguished from false intentions using the aIAT, finding that both short-term intentions (where to sleep the upcoming night) and long-term intentions (professional career) could be distinguished from plausible, but false intentions.

They further showed that the basis of such discrimination was related to intentions *per se* rather than hopes. In fact, when contrasted with hope sentences, intentions with or without pleasant outcomes were strongly associated with true sentences (Agosta et al., 2011a; Experiment 2).

DETECTION OF REASONS UNDERLYING LIES

According to De Paulo et al. (1996) and Vrij (2007), the reasons to lie may differ in terms of (1) the person who benefits from the lie (whether self or other-oriented), (2) the consequences of lying (in order to gain advantage or to avoid costs), and (3) the type of lying (whether for materialistic or psychological reasons). Self and other oriented lies are told either to protect oneself or others psychologically (e.g., protect from embarrassment or loss of face). According to Feldman (2009), standards of tact and politeness and expectations can make deception, to some degree, almost inevitable. Agosta et al. (2013) showed that the aIAT might be used to distinguish true from false reasons underlying other oriented lies (white lies) and that 20/20 (direct order) and 19/20 (reversed order) participants were correctly classified, with a D-IAT average value of 0.46 (direct order) and 0.50 (reversed order), respectively.

CONCLUSIONS

We have reviewed the validation experiments conducted so far that use the autobiographical IAT. The aIAT is a variant of the IAT (Greenwald et al., 1998) that might be used to establish the association of an event with the true/false logical dimension. In other words, the aIAT reveals which one of two contrasting events is more associated with the truth.

Table 5 | The data for the intention experiment.

Experiment	Participants	RT congruent (ms)	RT incongruent (ms)	D	% correct classification
Short term—sleep	N = 11	1011	1975	1.30	11/11
Long term—job	N = 11	1047	1507	1.02	11/11

The first condition refers to a short-term intention, where to sleep the coming night, while the second condition refers to a long-term intention of future work. Classification reaches 100% accuracy for both conditions.

Validation experiments have highlighted high classification accuracy over a series of tests with average accuracy over 90%. The average effect sizes were moderate: 0.57 for first-administration experiments and 0.67 for second-administration experiments. The previous results refer to a wide range of type of memories for a total of 578 subjects. Results from Experiment 1 in Lanciano et al. (2012) were excluded because the D-IAT was abnormally high, presumably due to the outstanding features of flashbulb memories.

It is worth noting that the same research group has carried out most of the studies conducted so far. Only a few experiments were conducted outside our laboratory (e.g., Hu and Rosenfeld, 2012; Hu et al., 2012; Takarangi et al., 2013) or in one associated laboratory (Lanciano et al., 2012). More studies from other laboratories are needed in order to better validate the technique and to determine a more reliable effect size, as some of the independent replications revealed lower effect sizes (Hu and Rosenfeld, 2012; Hu et al., 2012).

The validity of other lie detection techniques such as the CIT has usually been calculated using Cohen's d . For example, the meta-analysis by Ben-Shakhar and Elaad (2003) reported an overall average effect size of $d = 1.55$. Comparison of aIAT and CIT effect sizes test might be difficult, given substantial differences in calculating Cohen's d (calculated as the difference between the means of the detection score distributions of the guilty and innocent samples; Ben-Shakhar and Elaad, 2003) and the D-IAT values (calculated as the difference between the incongruent and congruent blocks of the same aIAT). The D-IAT algorithm takes into account the phenomenon of speed-accuracy trade-off, which is not an issue in CIT experiments.

In the future, the validation pipeline should include test-retest reliability over longer time frames and all other issues addressed in the CIT/GKT literature such as the modulating effects of personality and the full investigation of countermeasures. The CIT is the major memory-detection technique and has a much longer history and has been tested on a wider variety of conditions. The aIAT validation studies, compared with the CIT validation studies, lack extensive field studies. As is frequently reported, in the lie detection literature, studies carried out in the laboratory tend to overestimate accuracy and for this reason it will be critical for the aIAT to collect data in more ecological high-stake conditions (Elaad, 2011).

We have also identified a series of conditions that reduce the validity of the test and therefore should be avoided. Such conditions include the use of negative sentences in describing the events as well as using negative reminder labels. We have derived a D/accuracy function that permits us to estimate at the single subject level the probability of a given result in terms of accuracy, showing that classification accuracy for D-IAT values in the range of 0–0.2 is very poor, while D-IAT values above 0.6 are high and values between 0.2 and 0.6 are above 80%. In practical uses of the aIAT, attention should be paid to the level of D-IAT size as an indirect index of result reliability.

Here we summaries the guidelines for building an effective aIAT on the basis of the validation experiments reported above:

- Sentences related to true and false categories should always be true and false for the respondent (examples of *true* sentences are “I am in front of a computer,” or “I am sitting on a chair”; examples of *false* sentences are “I am climbing a mountain,” or “I am skiing”).
- Only one of the two events used to build an aIAT should be true and the other should be false. Two contrasting events should always be used; for example, “I left the door open” and “I closed the door” are good examples of sentences, as only one of the two is true for the respondent. The aIAT is supposed to uncover which one of the two is true.
- Do not use negative reminder labels or sentences. Use two contrasting autobiographical events.
- Before proceeding to the interpretation of the results, check whether the examinee has faked the test or not. The only available index up to now has been published in Agosta et al. (2011b). It compares response speed in single blocks (blocks 1, 2, 4) with response speed in double blocks (blocks 3 and 5).
- Evaluate which of the two autobiographical sentences is associated with *true* sentences using the D-IAT value. Compare blocks 3 and block 5. Identify the fastest block. The target autobiographical memory as the one that is more associated with the *true* logical dimension (on the basis of the faster reaction times in blocks 3 or 5).
- In single case studies, the use of a window of uncertainty is recommended. We suggest 0–0.2 as the uncertainty window. In such a range, probabilities of correct classification range from 50 to 75%. Reliability, as measured by agreement, is good for D-IAT values between 0.2 and 0.6 and very good for D-IAT values above 0.6.

Here, we add suggestions for aIAT users resulting from our own experience and highlighting the need for new studies that deeply investigate these issues:

- Memories should be encoded in sentences limited to a single line and about half a screen.
- Sentences describing autobiographical events should give a clear-cut description of the event. Fuzzy descriptions should be avoided, as valid discriminations have not been proven for such types of descriptions.
- Only two single, specific events (one true and one false) should be investigated; no more than two events, even if grouped in the same categories, should be used.
- In single case studies, confidence in the final results could be enhanced by using a design that includes: (i) build an aIAT on known autobiographical data (e.g., date of birth, names of sons, etc.) and check that this personal information is correctly identified and (ii) when testing the event of interest, the central fact and peripheral details should be used in different aIATs. The administration of these different aIATs should be delayed in time at least one week from one administration to the other in order to avoid any reduced effect on the D-IAT value.

REFERENCES

- Agosta, S., Castielo, U., Rigoni, D., Lionetti, S., and Sartori, G. (2011a). Behavioral Prior intentions: a technique for detection and its neural bases. *J. Cogn. Neurosci.* 23, 3888–3902. doi: 10.1162/jcnn_a_00039
- Agosta, S., Ghirardi, V., Zogmaister, C., Castielo, U., and Sartori, G. (2011b). Detecting fakers of the autobiographical IAT. *Appl. Cogn. Psychol.* 25, 299–306. doi: 10.1002/acp.1691
- Agosta, S., Mega, A., and Sartori, G. (2011c). Detrimental effects of using negative sentences in the autobiographical aIAT. *Acta Psychol.* 136, 296–306. doi: 10.1016/j.actpsy.2010.05.011
- Agosta, S., Pezzoli, P., and Sartori, G. (2013). How to detect deception in everyday life and reasons underlying it. *Appl. Cogn. Psychol.* 27, 256–262. doi: 10.1002/acp.2902
- Balota, D. A., Cortese, M. J., Duchek, J. M., Adams, D., Roediger, H. L., McDermott, K. B., et al. (1999). Veridical and false memories in healthy older adults and in dementia of the Alzheimer's type. *Cogn. Neuropsychol.* 16, 361–384. doi: 10.1080/026432999380834
- Becchio, C., Sartori, L., Bulgheroni, M., and Castello, U. (2008). The case of Dr. Jekyll and Mr. Hyde: a kinematic study on social intention. *Conscious. Cogn.* 17, 557–564. doi: 10.1016/j.concog.2007.03.003
- Becchio, C., Sartori, L., and Castielo, U. (2010). Toward you: the social side of action. *Curr. Dir. Psychol. Sci.* 19, 183–188. doi: 10.1177/0963721410370131
- Ben-Shakhar, G., and Elaad, E. (2003). The validity of psychophysiological detection of information with the Guilty Knowledge Test: a meta-analytic review. *J. Appl. Psychol.* 88, 131–135. doi: 10.1037/0021-9010.88.1.131
- Bohannon, J. N. (1988). Flashbulb memories of the space shuttle disaster: a tale of two theories. *Cognition* 29, 179–196. doi: 10.1016/0010-0277(88)90036-4
- Bratman, M. (1987). *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.
- Brown, R., and Kulik, J. (1977). Flashbulb memories. *Cognition* 5, 73–99. doi: 10.1016/0010-0277(77)90018-X
- Budson, A. E., Sullivan, A. L., Mayer, E., Daffner, K. R., Black, P. M., and Schacter, D. L. (2002). Suppression of false recognition in Alzheimer's disease and in patients with frontal lobe lesions. *Brain* 125, 2750–2765. doi: 10.1093/brain/awf277
- Cabeza, R., Rao, S. M., Wagner, A. D., Mayer, A. R., and Schacter, D. L. (2001). Can medial temporal lobe regions distinguish true from false? An event-related functional MRI study of veridical and illusory recognition memory. *Proc. Natl. Acad. Sci. U.S.A.* 98, 4805–4810. doi: 10.1073/pnas.081082698
- Cohen, P. R., and Levesque, H. J. (1990). Intention is choice with commitment. *Artif. Intell.* 42, 213–261. doi: 10.1016/0004-3702(90)90055-5
- Conway, M. A., Anderson, S. J., Larsen, S. F., Donnelly, C. M., McDaniel, M. A., McClelland, A. G. R., et al. (1994). The formation of flashbulb memories. *Mem. Cogn.* 22, 326–343. doi: 10.3758/BF03200860
- Deese, J. (1959). Influence of inter-item associative strength upon immediate free recall. *Psychol. Rep.* 5, 305–312. doi: 10.2466/pr.0.1959.5.3.305
- De Paulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., and Epstein, J. A. (1996). Lying in everyday life. *J. Pers. Soc. Psychol.* 70, 5979–5995.
- Elaad, E. (2011). "Validity of the concealed information test in realistic contexts," in *Memory Detection. Theory and Application of the Concealed Information Test*, eds B. Verschueren, G. Ben-Shakhar, and E. Meijer (New York; London: Cambridge University Press), 171–186. doi: 10.1017/CBO9780511975196.010
- Feldman, R. S. (2009). *Liar: The Truth About Lying*. Croydon: Virgin Books.
- Freng, S., and Kehn, A. (2012). Determining true and false witnessed events: can an eyewitness implicit association test distinguish between the seen and the unseen? *Psychiatry Psychol. Law iFirst*, 1–20. doi: 10.1080/13218719.2012.735885
- Gray, N. S., Brown, A. S., MacCulloch, M. G., Smith, J., and Snowden, R. J. (2005). An implicit test of the associations between children and sex in pedophiles. *J. Abnorm. Psychol.* 114, 304–308. doi: 10.1037/0021-843X.114.2.304
- Greenwald, G. A., McGhee, E. D., and Schwartz, K. L. J. (1998). Measuring individual differences in implicit cognition: the implicit association test. *J. Pers. Soc. Psychol.* 74, 1464–1480. doi: 10.1037/0022-3514.74.6.1464
- Greenwald, G. A., Nosek, B. A., and Banaji, M. R. (2003). Understanding and using the implicit association test: an improved score algorithm. *J. Pers. Soc. Psychol.* 85, 197–216. doi: 10.1037/0022-3514.85.2.197
- Higgins, J. P., Thompson, S. G., Deeks, J. J., and Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ* 327, 557–560. doi: 10.1136/bmj.327.7414.557
- Hu, X., and Rosenfeld, P. J. (2012). Combining the P300-complex trial-based Concealed Information test and the reaction time-based autobiographical Implicit Association Test in concealed memory detection. *Psychophysiology* 49, 1090–1100. doi: 10.1111/j.1469-8986.2012.01389.x
- Hu, X., Rosenfeld, J. P., and Bodenhausen, G. V. (2012). Combating automatic autobiographical associations: the effect of instruction and training in strategically concealing information in the autobiographical implicit association test. *Psychol. Sci.* 23, 1079–1085. doi: 10.1177/0956797610364762
- Nosek, B. A., Greenwald, A. G., and Banaji, M. R. (2007). "The Implicit Association Test at age 7: a methodological and conceptual review," in *Automatic Processes in Social Thinking and Behavior*, ed J. A. Bargh (Hove: Psychology Press), 265–292.
- Pacherie, E. (2008). The phenomenology of action: A conceptual framework. *Cognition* 107, 179–217. doi: 10.1016/j.cognition.2007.09.003
- Pacherie, E., and Haggard, P. (2010). "What are Intentions?" in *Conscious Will and Responsibility. A tribute to Benjamin Libet*, eds L. Nadel and W. Sinnott-Armstrong (Oxford: Oxford University Press), 70–84
- Payne, D. G., Elie, C. J., Blackwell, J. M., and Neuschatz, J. S. (1996). Memory illusions: recalling, recognizing, and recollecting events that never occurred. *J. Mem. Lang.* 35, 261–285. doi: 10.1006/jmla.1996.0015
- Petroczi, A., Aidman, E. V., Hussain, I., Deshmukh, N., Nepusz, T., Uvacsek, M., et al. (2010). Virtue or pretense? looking behind self-declared innocence in doping. *PLoS ONE* 5:e10457. doi: 10.1371/journal.pone.0010457
- Pillemer, D. B. (1984). Flashbulb memories of the assassination attempt on President Reagan. *Cognition* 16, 63–80. doi: 10.1016/0010-0277(84)90036-2
- Roediger, H. L. (1996). Memory illusions. *J. Mem. Lang.* 35, 76–100. doi: 10.1006/jmla.1996.0005
- Roediger, H. L., and McDermott, K. B. (1995). Creating false memories: remembering words not presented in lists. *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 803–814. doi: 10.1037/0278-7393.21.4.803
- Sartori, G., Agosta, S., Zogmaister, C., Ferrara, S. D., and Castielo, U. (2008). How to accurately detect autobiographical events. *Psychol. Sci.* 19, 772–780. doi: 10.1111/j.1467-9280.2008.02156.x

- Schacter, D. L. (1999). The seven sins of memory: insights from psychology and cognitive neuroscience. *Am. Psychol.* 54, 182–203. doi: 10.1037/0003-066X.54.3.182
- Schmukle, S., Stefan, C., Back, M. D., and Egloff, B. (2008). Validity of the five-factor model for the implicit self-concept of personality. *Eur. J. Psychol. Assess.* 24, 263–272. doi: 10.1027/1015-5759.24.4.263
- Searle, J. (1983). *Intentionality: An Essay in the Philosophy of Mind*. New York, NY: Cambridge University Press. doi: 10.1017/CBO9781139173452
- Seymour, T. L., Seifert, C. M., Shafto, M. G., and Mosmann, A. L. (2000). Using response time measures to assess “guilty knowledge”. *J. Appl. Psychol.* 85, 30–37. doi: 10.1037/0021-9010.85.1.30
- Slotnick, S. D., Schacter, D. L. (2004). A sensory signature that distinguishes true from false memories. *Nat. Neurosci.* 7, 664–672. doi: 10.1038/nn1252
- Stadler, A. M., Roediger, H. L., and McDermott, K. B. (1999). Norms for word lists that create false memories. *Mem. Cogn.* 27, 494–500. doi: 10.3758/BF03211543
- Talarico, J. M., and Rubin, D. C. (2009). “Flashbulb memories result from ordinary memory processes and extraordinary event characteristics,” in *Flashbulb Memories: New Issues and New Perspectives*, eds O. Luminet and A. Curci (Hove: Psychology Press), 79–97.
- Talarico, J. M., and Rubin, D. C. (2003). Confidence, not consistency, characterizes flashbulb memories. *Psychol. Sci.* 14, 455–461. doi: 10.1111/1467-9280.02453
- Talarico, J. M., and Rubin, D. C. (2007). Flashbulb memories are special after all: in phenomenology, not accuracy. *App. Cogn. Psychol.* 21, 557–578. doi: 10.1002/acp.1293
- Takarangi, M. K., Strabge, D., Shortland, A. E., and James, H. E. (2013). Source confusion influences the effectiveness of the autobiographical IAT. *Psychon. Bull. Rev.* doi: 10.3758/s13423-013-0430-3. [Epub ahead of print].
- Tulving, E. (1983). *Elements of Episodic Memory*. Oxford: Oxford University Press.
- Verschueren, B., Ben-Shakar, G., and Meijer, E. (2011). *Memory Detection. Theory and Application of the Concealed Information Test*. New York; London: Cambridge University Press. doi: 10.1017/CBO9780511975196
- Verschueren, B., Prati, V., and De Houwer, J. (2009). Cheating the lie detector: faking the autobiographical IAT. *Psychol. Sci.* 20, 410–413. doi: 10.1111/j.1467-9280.2009.02308.x
- Vrij, A. (2007). *Detecting Lies and Deceit: Pitfalls and Opportunities*. Chichester: John Wiley and Sons Ltd.
- Weaver, C. A. (1993). Do you need a “flash” to form a flashbulb memory? *J. Exp. Psychol. Gen.* 122, 39–46. doi: 10.1037/0096-3445.122.1.39

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 25 July 2012; accepted: 23 July 2013; published online: 13 August 2013.
Citation: Agosta S and Sartori G (2013) The autobiographical IAT: a review. Front. Psychol. 4:519. doi: 10.3389/fpsyg.2013.00519

This article was submitted to Frontiers in Cognitive Science, a specialty of Frontiers in Psychology.

Copyright © 2013 Agosta and Sartori. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Learning to lie: effects of practice on the cognitive cost of lying

B. Van Bockstaele^{1*}, B. Verschueren^{1,2,3}, T. Moens¹, Kristina Suchotzki¹, Evelyne Debey¹ and Adriaan Spruyt¹

¹ Faculty of Psychology and Educational Sciences, Department of Experimental Clinical and Health Psychology, Ghent University, Ghent, Belgium

² Faculty of Social and Behavioural Sciences, Department of Clinical Psychology, University of Amsterdam, Amsterdam, Netherlands

³ Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, Netherlands

Edited by:

Matthias Gamer, University Medical Center Hamburg-Eppendorf, Germany

Reviewed by:

Catherine Kaylor-Hughes, University of Nottingham, UK

George Visu-Petru, Babes-Bolyai University, Romania

Christopher Baker, Saint Xavier University, USA

***Correspondence:**

B. Van Bockstaele, Faculty of Psychology and Educational Sciences, Department of Experimental Clinical and Health Psychology, Ghent University, Henri Dunantlaan 2, B-9000 Ghent, Belgium.

e-mail: bram.vanbockstaele@ugent.be;
bram.vanbockstaele@gmail.com

Cognitive theories on deception posit that lying requires more cognitive resources than telling the truth. In line with this idea, it has been demonstrated that deceptive responses are typically associated with increased response times and higher error rates compared to truthful responses. Although the cognitive cost of lying has been assumed to be resistant to practice, it has recently been shown that people who are trained to lie can reduce this cost. In the present study ($n = 42$), we further explored the effects of practice on one's ability to lie by manipulating the proportions of lie and truth-trials in a Sheffield lie test across three phases: Baseline (50% lie, 50% truth), Training (frequent-lie group: 75% lie, 25% truth; control group: 50% lie, 50% truth; and frequent-truth group: 25% lie, 75% truth), and Test (50% lie, 50% truth). The results showed that lying became easier while participants were trained to lie more often and that lying became more difficult while participants were trained to tell the truth more often. Furthermore, these effects did carry over to the test phase, but only for the specific items that were used for the training manipulation. Hence, our study confirms that relatively little practice is enough to alter the cognitive cost of lying, although this effect does not persist over time for non-practiced items.

Keywords: **deception, cognitive training, response inhibition, lie detection, intentionality**

INTRODUCTION

Cognitive theories on deception posit that deliberate and successful lying requires more cognitive resources than telling the truth (Vrij et al., 2006, 2011). Liars have to fabricate a story, monitor the reactions of the interaction partner, make sure that their story remains coherent and consistent, control behaviors that may signal lying or stress, and inhibit or conceal the truth. Several neuroimaging studies have provided evidence in line with this idea, showing that prefrontal brain regions which are involved in cognitive control (i.e., the anterior cingulate, dorsolateral prefrontal, and inferior frontal regions) are more active when participants are lying compared to when they are telling the truth (for reviews, see Christ et al., 2009; Gamer, 2011). The higher activation of brain regions involved in cognitive control suggests that individuals who are lying are engaged in a cognitively demanding task. Although it is generally agreed that lying comes at a cognitive cost and telling the truth is the default, dominant response, there is less agreement as to whether this cognitive cost is invariable or whether it is malleable through practice. For instance, pathological liars lie so often that lying becomes an automatism rather than an exception (Dike et al., 2005). One can thus expect that such people experience less cognitive difficulty when lying. The same holds for crime suspects who face interrogation and who have thoroughly practiced their story (Spence et al., 2008) or for people who have told the same lies so often that they believe their lies to be the truth (Polage, 2012).

To date, however, evidence concerning the effect of practice on the cognitive cost of lying is both scarce and mixed. Johnson et al. (2005) asked participants to memorize a list of words,

and later used these and other words in an old/new recognition task. Over different blocks of the word recognition task, participants were instructed to either respond truthfully or deceptively. Crucially, they found that both behavioral and neurological measures of cognitive control were unaffected by practice in deceptive responding. These findings led the authors to conclude that lying *always* comes at a cognitive cost, and thus that the cognitive complexity of lying is resistant to practice. It should be noted, however, that Johnson et al. adopted a blocked within-subjects design with a random succession of truthful and deceptive blocks. Such an approach may have been suboptimal to study the impact of practice on the cognitive cost of lying as participants' ability to lie in deceptive blocks may have been counteracted by intermediate truthful blocks, and vice versa. Vendemia et al. (2005) used autobiographical statements about their participants which were either true or false. In three sessions, participants were required to respond truthfully on half of the trials and deceptively on the other half of the trials, depending on the color of the statements. Although reaction time data revealed no practice effect whatsoever the errors data did show that the difference between deceptive and truthful responses diminished following practice. This latter finding illustrates that practice may have had some effect on the cognitive cost of lying. Furthermore, the training manipulation in this experiment was relatively weak, as participants were required to respond both truthfully and deceptively 50% of the time. As such, participants were not explicitly trained to either lie or tell the truth. Finally, using a Sheffield lie test (Spence et al., 2001; a variant of the differentiation of deception paradigm by Furedy et al.,

1988), Verschuere et al. (2011b) recently challenged the idea that the cognitive cost of lying is resistant to practice. In the standard version of this task, autobiographical questions are presented on a computer screen, and participants provide yes/no-answers using one of two different response keys. The questions can appear in two different colors, and participants are instructed to lie if the sentence is presented in the one color (lie-trials) and to tell the truth if the sentence is presented in the other color (truth-trials). In a control group, with 50% lie-trials and 50% truth-trials, they found that lying requires more cognitive resources than telling the truth, as illustrated by slower response times and more errors on lie-trials compared to truth-trials (i.e., the lie-effect; see also Sartori et al., 2008). In two other groups, a number of filler trials were added in order to manipulate the numbers of lie and truth-trials. In the frequent-lie group, all these filler trials required a deceptive response. In contrast, all the filler trials required a truthful response in the frequent-truth group. As such, participants in the frequent-lie group were required to lie on 75% of the trials whereas participants in the frequent-truth group only lied on 25% of the trials. Both the response latency data and error rates indicated that lying became easier while people were lying more often, and lying became more difficult while people gave more truthful responses.

In the present experiment, we further examined the influence of practice on the cognitive cost of lying. More specifically, we investigated (1) whether practice has an effect on participants' initial cognitive cost of lying, and (2) whether such effects continue to exist after the training, and thus whether practice really changes the dominance of the truth response. The experimental design used by Verschuere et al. (2011b) did not allow firm conclusions concerning this important matter because they only assessed the lie-effects while participants were being exposed to either a high proportion of lie-trials or a high proportion of truth-trials. As such, their results indicate only that *while* participants are lying often, lying becomes easier, and *while* participants are often telling the truth, lying becomes more difficult. Therefore, we attempted to replicate and extend the findings of Verschuere et al. by prolonging the training phase and by adding a baseline phase and a test phase in which participants were required to respond deceptively and truthfully equally often. We expected that (1) the cognitive cost of lying would change as a result of our training manipulation, and (2) if practice does genuinely change the dominance of the truth response, the change in the cognitive cost of lying would persist over time. In other words, we expected a linear trend in the size of the lie-effect in the training phase as a function of the proportion of lie-trials, with a smaller lie-effect in the frequent-lie group, a medium lie-effect in the control group, and a larger lie-effect in the frequent-truth group. If training genuinely alters the dominance of the truth response, this linear trend should extend to the test phase.

MATERIALS AND METHODS

PARTICIPANTS

Forty-five undergraduate students (23 men) of Ghent University participated in exchange for course credits. The data of three participants were not analyzed because of poor accuracy on test trials (see below; participants' accuracy scores = 54, 58, and 62%; deviating more than 2.5SDs from the group average = 89%, SD = 10%).

Hence, our results are based on the data of 42 participants. All participants provided written informed consent prior to the experiment.

MODIFIED SHEFFIELD LIE TEST

In the Sheffield lie test, we used 108 different questions. Thirty-six of these questions were yes-or-no questions about basic semantic knowledge, and were used to allow us to give performance feedback during the acquaintance phase. Half of these practice questions required a "no"-response (e.g., "Is London a Belgian city?"), and the other half required a "yes" -response (e.g., "Is a stone hard?"). The remaining 72 questions were autobiographical yes-or-no questions related to actions that participants had or had not performed on the day of testing (see Table A1 in Appendix). Before the start of the experiment, participants were asked to give a truthful response to these questions. Some of these questions were more likely to elicit an affirmative response (e.g., "Did you drink water?") than others (e.g., "Did you greet a police officer?"). In this way we tried to establish a yes-no ratio of approximately 50%. Analyses of the yes/no ratio revealed that participants gave more no-answers (67%, SD = 6.92) than yes-answers (33%), $t(41) = 15.82$, $p < 0.001$. However, this was the case for all three groups (see below), all $ts > 9.27$, all $ps < 0.001$, and there was no difference between the groups, $F(2, 39) = 1.65$, $p = 0.20$. Half of these questions were used in filler trials, and the other half in test trials (counterbalanced). The general appearance of test trials and filler trials was identical. On each trial, a sentence was presented in white bold Arial font in the center of the black screen, together with the response labels "YES" and "NO" at the sides of the screen. The response labels could either appear in blue or in yellow, and, depending on this color, participants were required to give a truthful or a deceptive yes-no response (i.e., blue = truth, yellow = lie, or vice versa) by pressing either the "4" or the "6" key on the numeric pad of a standard AZERTY keyboard. The assignment of the two response buttons to either yes-or-no responses was counterbalanced across participants, as was the assignment of the two colors to either truthful or deceptive responding. There was no response deadline. In order to prevent strategic recoding of the task, we also included catch trials (Johnson et al., 2003, 2005; Verschuere et al., 2011b). On these trials, either the word "yes" or the word "no" appeared in the center of the screen and participants were required to respond according to their meaning (i.e., press the yes-button for the word "yes", and the no-button for the word "no"), irrespective of the color of the response labels.

Our modified version of the Sheffield lie test consisted of 924 trials, presented in an acquaintance phase (24 trials), a baseline phase (180 trials), a training phase (540 trials), and a test phase (180 trials). The acquaintance phase consisted of only semantic trials with performance feedback allowing participants to familiarize with the task at hand. The data of these trials were not analyzed. In the *baseline* phase, we presented 72 test trials (36 truth, 36 lie), 72 filler trials (36 truth, 36 lie), and 36 catch trials (18 yes, 18 no) in an intermixed, random fashion. In the *training* phase, participants were randomly assigned to one of three different groups. In all three groups, we presented three identical blocks, each consisting of 72 test trials (36 truth, 36 lie), 72 filler trials, and 36 catch trials. For the test trials, the proportion of truthful and deceptive

responses remained 50/50. However, the proportion of filler trials requiring a deceptive or a truthful response differed across the three groups in the training phase. In the frequent-lie group, all the filler trials required a deceptive response, whereas in the frequent-truth group, all the filler trials required the participants to respond truthfully. Finally, in the control group, half of the filler trials required a truthful response, and half required a deceptive response. As such, due to the manipulation of the lie-truth proportion of the filler trials, participants in the frequent-lie group lied on 75% of the trials in the training phase, whereas participants in the frequent-truth group only lied on 25% of the trials. Participants were not informed about this manipulation. The last phase was a *test* phase, which was identical to the baseline phase. Participants were allowed to take a short break after each block.

DATA PROCESSING

For the analyses of the response latency data, trials with erroneous responses were discarded. In order to reduce the impact of extreme reaction times, we recoded response latencies faster than 300 ms and slower than 3000–300 and 3000 ms respectively (Greenwald et al., 1998)¹. Using this procedure, we recoded a total of 10.43% of the correct trials (for latencies larger than 3000 ms: 5.43% lie-trials and 4.77% truth-trials; for latencies smaller than 300 ms: 0.11% lie-trials and 0.12% truth-trials). Next, for each participant and each experimental condition, we calculated the average response latencies (ms) and accuracy scores (%). Finally, we calculated “lie-effect” scores by subtracting the response latencies and accuracy scores on lie-trials from the response latencies and accuracy scores on truth-trials. A large positive lie-effect score reflects greater difficulty in lying, and a negative lie-effect score reflects greater ease in lying. For all analyses, the alpha level was set to 0.05.

RESULTS

TEST TRIALS

Test trials were analyzed in order to investigate (1) whether practice in truthful and deceptive responding on the filler trials affected truthful and deceptive responding on the test trials during the training phase, and (2) whether these effects transferred to the test phase. To do so, we subjected the lie-effect scores of both the reaction time data and the errors to 3 (Group: frequent-lie vs. control vs. frequent-truth) \times 3 (Experiment Phase: baseline vs. practice vs. test) repeated measures ANOVAs with Group as a between subjects factor and Experiment Phase as a within subject factor. We expected no group differences in the baseline phase, and a linear effect of Group (i.e., a gradual increase in the magnitude of the lie-effect from the frequent-lie group over the control group to the frequent-truth group) in the training phase. Finally, we expected this training effect to generalize to the test phase.

Reaction times

Neither of the main effects reached significance, both $F_s < 1$. However, the interaction between the linear effect of Group

and Experiment Phase followed a significant quadratic course, $F(1, 39) = 18.56$, $p < 0.0005$ (see **Figure 1A**), indicating that the linear effect of Group varied across the three Experiment Phases. Whereas the lie-effect was the same for the three groups during the baseline phase, $F < 1$ (frequent-lie group: $M = 222$, $SD = 217$; control group: $M = 148$, $SD = 242$; frequent-truth group: $M = 163$, $SD = 176$), there was a significant linear effect of Group during the training phase, $F(1, 39) = 7.56$, $p < 0.01$, $f = 0.44^2$. This linear course illustrates that the size of the lie-effect gradually increased from the frequent-lie group ($M = 53$, $SD = 245$) over the control group ($M = 145$, $SD = 123$) to the frequent-truth group ($M = 239$, $SD = 146$). In the test phase, there was no effect of Group, $F < 1$, indicating that the lie-effect scores of the three groups no longer differed significantly (frequent-lie group: $M = 180$, $SD = 238$; control group: $M = 171$, $SD = 195$; frequent-truth group: $M = 127$, $SD = 231$).

Errors

As for the reaction time data, neither of the main effects reached significance, both $F_s < 1.45$, both $p_s > 0.24$. However, the interaction between Experiment Phase and the linear effect of Group again followed a significant quadratic course, $F(1, 39) = 4.73$, $p < 0.05$ (see **Figure 1B**), illustrating that group differences in the lie-effect varied across the three Experiment Phases. The lie-effect was the same for the three groups in the baseline phase, $F < 1$ (frequent-lie group: $M = 4.56$, $SD = 4.44$; control group: $M = 4.37$, $SD = 7.68$; frequent-truth group: $M = 4.76$, $SD = 9.27$). However, in the training phase, there was a significant linear effect of Group, $F(1, 39) = 7.21$, $p < 0.05$, $f = 0.42$. As can be seen in **Figure 1B**, the lie-effect in the training phase was smaller in the frequent-lie group ($M = 0.60$, $SD = 3.90$), intermediate in the control group ($M = 2.32$, $SD = 3.01$), and larger in the frequent-truth group ($M = 5.27$, $SD = 6.32$). These group differences were no longer significant in the test phase, $F < 1$ (frequent-lie group: $M = 4.76$, $SD = 5.49$; control group: $M = 4.17$, $SD = 10.77$; frequent-truth group: $M = 3.39$, $SD = 9.26$).

FILLER TRIALS

Filler trials were analyzed in order to investigate whether practice with specific items influences the cognitive cost of lying on these specific items. For these analyses, we discarded the data of the training phase because these trials were all either lie-trials or truth-trials for the two experimental groups, and hence do not allow to calculate the crucial difference between truthful and deceptive responses. For both reaction times and error rates, the lie-effect scores were subjected to a 3(Group) \times 2(Experiment Phase: baseline vs. test) repeated measures ANOVA. We expected no group differences in the baseline phase, and a linear effect of Group (i.e., a gradual increase in the magnitude of the lie-effect from the frequent-lie group over the control group to the frequent-truth group) in the test phase.

¹Another outlier analysis in which we first removed all reaction times faster than 200 ms and slower than 5000 ms and then removed all reaction times that deviated more than 3SDs from the individual's mean yielded the same overall pattern of results.

²We calculated the effect size f using the following formula: $f = \sqrt{[\eta_p^2 / (1 - \eta_p^2)]}$. According to Cohen (1992), values from 0.10 represent small effects, values from 0.25 represent medium effects and values from 0.40 represent large effects.

Reaction times

Analysis of the lie-effect scores on filler trials yielded a significant main effect of Group, $F(2, 39) = 4.07, p < 0.05$. Follow-up between-group comparisons showed that neither the frequent-truth group nor the frequent-lie group differed significantly from the control group, $F(1, 26) = 0.95, p = 0.10$, and $F(1, 26) = 1.04, p = 0.32$, respectively. However, the lie-effect was significantly larger in the frequent-truth group compared to the frequent-lie group, $F(1, 26) = 7.64, p < 0.05$. The interaction between Experiment Phase and the linear effect of Group was not significant, $F(1, 39) = 2.65, p = 0.11$. Exploratory analyses on each experiment phase separately showed, however, that there was no linear effect of Group in the baseline phase, $F(1, 39) = 1.29, p = 0.26$ (frequent-lie group: $M = 111, SD = 188$; control group: $M = 226, SD = 230$; frequent-truth group: $M = 189, SD = 102$), but a clear linear effect in the test phase, $F(1, 39) = 8.28, p < 0.005, f = 0.46$. **Figure 2A** illustrates that in the test phase, the lie-effect in the test phase gradually increased from the frequent-lie group ($M = 24, SD = 296$) over the control group ($M = 131, SD = 156$) to the frequent-truth group ($M = 269, SD = 199$).

Errors

Neither of the main effects reached significance, both $F_s < 1$, but the interaction between Experiment Phase and the linear effect of Group was significant, $F(1, 39) = 6.27, p < 0.05$ (see

Figure 2B). A one-way ANOVA on the lie-effect scores revealed no linear effect of Group in the baseline phase, $F(1, 39) = 2.88, p = 0.10$ (frequent-lie group: $M = 4.17, SD = 4.84$; control group: $M = 4.36, SD = 4.84$; frequent-truth group: $M = 0.00, SD = 8.92$). In contrast, the main effect of Group showed a significant linear course in the test phase, $F(1, 39) = 4.11, p < 0.05, f = 0.33$. As can be seen in **Figure 2B**, the lie-effect again gradually increased from the frequent-lie group ($M = -0.59, SD = 4.89$) to the control group ($M = 1.19, SD = 6.86$), and from the control group to the frequent-truth group ($M = 3.97, SD = 5.94$).

DISCUSSION

In the present experiment, we investigated whether practice in lying or telling the truth influences the cognitive cost of lying. Like Verschueren et al. (2011b), we found that during the training phase, lying became more difficult for participants in the frequent-truth group than for participants in the frequent-lie group. As such, our present results are in conflict with the conclusions of Johnson et al. (2005) and Vendemia et al. (2005), who both argued that the cognitive cost of lying is resistant against practice. However, the experimental designs of both Johnson et al. and Vendemia et al. may have been suboptimal to investigate the effects of practice on the cognitive cost of lying. As mentioned earlier, Johnson et al. randomly intermixed blocks of truth-trials and blocks of lie-trials *within* participants, which may

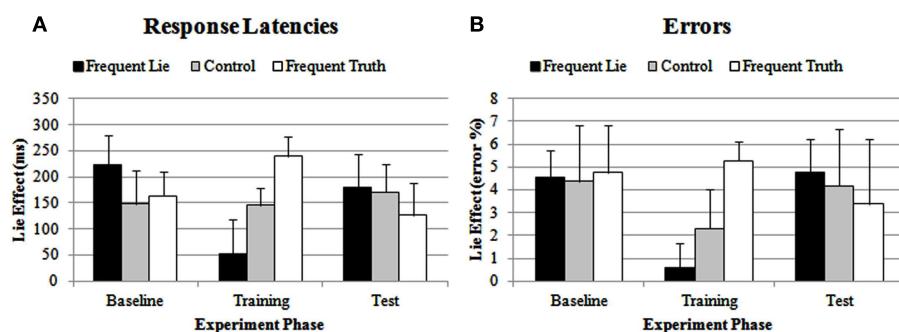


FIGURE 1 | Lie-effect of response latencies (A) and errors (B) on test trials for the frequent-lie, control, and frequent-truth group, during the baseline phase, the training phase and the test phase. Error bars represent the standard error of the mean.

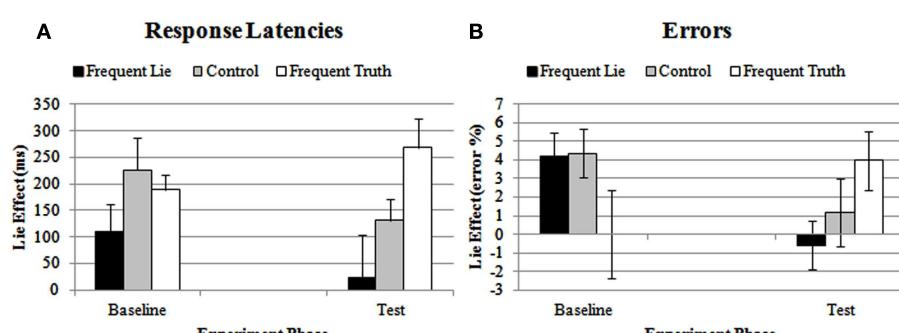


FIGURE 2 | Lie-effect of response latencies (A) and errors (B) on filler trials for the frequent-lie, control, and frequent-truth group, during the baseline phase and the test phase. Error bars represent the standard error of the mean.

have prevented consistent changes in their participants' abilities to lie. Likewise, participants in the experiment of Vendemia et al. were manipulated to lie on only 50% of the trials, mimicking the design that we used for our control group. It could therefore be argued that participants in the study of Vendemia et al. were not consistently trained to respond either truthfully or deceptively, making consistent changes in their cognitive ability to lie less likely.

Furthermore, we found that practice can have some enduring effects on the cognitive cost of lying in the test phase (see also Hu et al., 2012). During the test phase, lying was easier for participants in the frequent-lie group and lying was more difficult for participants in the frequent-truth group. However, these effects were limited to the data of the filler questions (i.e., the specific questions which were used for the training manipulation). The finding that the training effect did not carry over to the test questions (i.e., the questions which required 50% truthful and 50% deceptive responses throughout the entire experiment) indicates that our training manipulation was not sufficient to genuinely alter the dominance of the truth response. Although unexpected, it is not uncommon to find that cognitive training manipulations do not generalize to non-trained stimuli (e.g., Schoenmakers et al., 2007). Further research is needed to investigate whether a more intensive training (e.g., over several days) would have such an enduring effect on new items. Nevertheless, our finding that changes in the lie-effect on the filler trials persisted over time further challenges the assumption that "... even after thousands of trials of practice, it is unlikely that the increased difficulty associated with making deceptive responses will be erased entirely" (Johnson et al., 2005, p. 402).

An interesting remaining issue concerns the mechanism underlying the changes in the lie-effect on test trials during the training phase. As this change did not persist in the test phase, the effects may be caused by specific properties of the task or the design. In our opinion, there are at least three – not necessarily mutually exclusive – possible explanations. A first explanation stems from research on task switching (e.g., see Monsell, 2003; Kiesel et al., 2010; Vandierendonck et al., 2010). In a typical task switching design, participants are on each trial required to perform one out of two different tasks. Participants are generally faster and more accurate on trials that are preceded by a trial in which the same task was performed compared to trials that are preceded by a trial in which the other task was performed. The drop in performance on trials that require a task switch is known as the switch cost. Our present experiment bears some resemblance with such dual task paradigms in the sense that our participants were also required to perform one out of two possible tasks, namely lying or telling the truth. Crucially, during the training phase, the switching between lying and telling the truth differed between the three groups. In the frequent-truth group, the overall larger proportion of truth-trials increased the probability that truth-trials involved repetitions and that lie-trials involved a switch (e.g., Truth-Truth-Truth-Lie-Truth). In a similar fashion, in the frequent-lie group, truth-trials were more likely to involve switches and lie-trials were more likely to involve repetitions (e.g., Lie-Lie-Truth-Lie-Lie). Thus, task switch costs may have increased the lie-effect in the frequent-truth group, and reduced it in the frequent-lie group. A

second possible mechanism behind the group differences on the test trials during the training phase is the oddball-effect (e.g., see Squires et al., 1975; Stevens et al., 2000; Goldstein et al., 2002). In a typical oddball task, participants are required to respond differently to two types of stimuli. Crucially, one of the stimuli is highly frequent, and the other is less frequent. Participants are typically fast to respond to the frequent stimuli, but slower to respond to the less frequent stimuli. In the present study, participants in the frequent-lie group encountered many lie-trials, and truth-trials were relatively rare. As a result, these participants were more likely to respond fast on lie-trials and slow on truth-trials, resulting in a decreased lie-effect. Inversely, in the frequent-truth group, the truth-trials were highly frequent and the lie-trials were relatively rare, resulting in fast responses on truth-trials and slower responses on lie-trials, and thus leading to a stronger lie-effect. A third possible explanation for the group differences during the training phase is goal neglect (e.g., see De Jong et al., 1999; Kane and Engle, 2003; Debey et al., 2012). According to the goal neglect theory (Duncan, 1995), the selection of an appropriate response is guided by task goals. The more active such a task goal is, the more accurate and fast a response will be, while responses will be slower and less accurate if they are guided by a more neglected task goal. It is possible that our manipulation of the proportions of lie and truth-trials resulted in our three groups having different dominant task goals. As a result, in the frequent-lie group, the most active task goal may have been to respond deceptively and inhibit truthful responses, resulting in fast and accurate responses on lie-trials and slower and less accurate responses on truth-trials, resulting in a smaller lie-effect. Inversely, if the main task goal in the frequent-truth group was to respond truthfully and avoid lying, this would result in fast and accurate responses on truth-trials, and slower and less accurate responses on lie-trials, hence resulting in a larger lie-effect.

In future research, it may be possible to differentiate between these different mechanisms. For instance, presenting lie- and truth-trials in a predictable order should reduce the impact of a switch cost or oddball-effect, while such a manipulation is unlikely to influence the effect of goal neglect. Another possibility is to manipulate the duration of the response-stimulus interval (RSI). While longer RSIs provide more preparation time and should hence decrease switch costs (Monsell, 2003), longer RSIs have also been shown to hamper goal maintenance and induce goal neglect (De Jong et al., 1999; Debey et al., 2012). Thus, if the difference in the lie-effects during the training phase is driven by a switch cost, then an increased RSI should reduce this difference. Alternatively, if the difference is driven by goal neglect, then an increased RSI should further inflate the effect.

The results of our study may also have implications for the detection of deception in forensic settings (Granagh and Strömwall, 2004; Verschueren et al., 2011a). Given the fact that lying becomes more difficult while people often tell the truth, the accuracy of lie detection tests may be improved by adding a large number of verifiable questions to the interrogation. Our results suggest that if suspects are obliged to respond truthfully on these verifiable questions, they may experience greater difficulty when lying on a crucial incriminating question. Our results of the filler trials suggest that the cognitive cost of lying can be reduced

for specific well-trained lies. Translated to the forensic context, this may mean that a guilty suspect who has repeated the same lies over and over again (e.g., to the police, to lawyers, to judge, etc.) may experience less cognitive load when lying. Our results suggest that, with repeated lying, deceptive responses may cognitively mirror truth telling, thus hampering lie detection. However, more research is needed to investigate these forensic implications in detail. For instance, it is uncertain whether our results would be replicated in a context where participants have something to gain or lose or where arousal and emotional distress are high, or whether certain strategies or counter-measures can influence our pattern of results.

Our study also has a number of limitations. First, our sample consisted only of 42 participants, resulting in 14 participants per group. As a result, our experiment may have lacked the statistical power that is needed to uncover smaller effects. A second limitation is more inherent to our specific methodology, namely our use of autobiographical questions as stimuli in the Sheffield lie test. As mentioned earlier, these specific questions were not emotionally salient, nor were they related to crime. As such, the possible forensic implications of our present results need to be addressed in an ecologically more valid context, for instance by using questions related to a mock crime that participants have or have not committed. Also, although the autobiographical questions that we used were related to actions that participants had or had not

performed on the day of testing, some participants may have been uncertain of their initial answers as well as their responses during the subsequent Sheffield lie test. Such uncertainty may have arisen especially on questions concerning habitual behaviors (e.g., buying a newspaper), or on actions that were performed on the day before testing. In our follow-up research, we now ask participants to perform specific actions in the laboratory, prior to testing (e.g., see Debey et al., 2012). This new methodology has advantages over the methodology that we used in the present experiment, as it allows us to control the yes/no ratio of answers and it reduces possible uncertainty with the participants.

In sum, the data of our present experiment suggest that lying becomes cognitively less demanding while participants are often lying, and that lying becomes cognitively more difficult while participants are often telling the truth. Furthermore, these effects were not due to baseline differences between the three groups, and practice on specific lies had enduring effects over time, suggesting that the detection of well-trained lies may prove to be a thorny issue.

ACKNOWLEDGMENTS

Kristina Suchotzki is supported by an ECRP Grant (09-ECRP-025; FWO Grant ESF 3G099310). Evelyne Debey is a fellow of the Special Research Fund (Aspirant BOF UGent) at Ghent University. Adriaan Spruyt is Postdoctoral Fellow of the Flemish Research Foundation (FWO – Vlaanderen).

REFERENCES

- Christ, S. E., van Essen, D. C., Watson, J. M., Brubaker, L. E., and McDermott, K. B. (2009). The contributions of prefrontal cortex and executive control to deception: evidence from activation likelihood estimate meta-analysis. *Cereb. Cortex* 19, 1557–1566.
- Cohen, J. (1992). A power primer. *Psychol. Bull.* 112, 155–159.
- De Jong, R. D., Berendsen, E., and Cools, R. (1999). Goal neglect and inhibitory limitations: dissociable causes of interference effects in conflict situations. *Acta Psychol. (Amst.)* 101, 379–394.
- Debey, E., Verschueren, B., and Crombez, G. (2012). Lying and executive control: an experimental investigation using ego depletion and goal neglect. *Acta Psychol. (Amst.)* 140, 133–141.
- Dike, C. C., Baranoski, M., and Griffith, E. E. H. (2005). Pathological lying revisited. *J. Am. Acad. Psychiatry Law* 33, 342–349.
- Duncan, J. (1995). “Attention, intelligence, and the frontal lobes,” in *The Cognitive Neurosciences*, ed. M. S. Gazzaniga (Cambridge: MIT Press), 721–733.
- Furedy, J. J., Davis, C., and Gurevich, M. (1988). Differentiation of deception as a psychological process – A psychophysiological approach. *Psychophysiology* 25, 683–688.
- Gamer, M. (2011). “Detection of deception and concealed information using neuroimaging techniques,” in *Memory Detection: Theory and Application of the Concealed Information Test*, eds B. Verschueren, G. Ben-Shakhar, and E. Meijer (Cambridge: Cambridge University Press), 90–113.
- Goldstein, A., Spencer, K. M., and Donchin, E. (2002). The influence of stimulus deviance and novelty on the P300 and Novelty P3. *Psychophysiology* 39, 781–790.
- Granhag, P. A., and Strömwall, L. A. (2004). *The Detection of Deception in Forensic Contexts*. Cambridge: Cambridge University Press.
- Greenwald, A. G., McGhee, D. E., and Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: the implicit association test. *J. Pers. Soc. Psychol.* 74, 1464–1480.
- Hu, X., Chen, H., and Fu, G. (2012). A repeated lie becomes a truth? The effect of intentional control and training on deception. *Front. Psychol.* 3:488. doi:10.3389/fpsyg.2012.00488
- Johnson, R., Barnhardt, J., and Zhu, J. (2003). The deceptive response: effects of response conflict and strategic monitoring on the late positive component and episodic memory-related brain activity. *Biol. Psychol.* 64, 217–253.
- Johnson, R., Barnhardt, J., and Zhu, J. (2005). Differential effects of practice on the executive processes used for truthful and deceptive responses: an event-related brain potential study. *Brain Res. Cogn. Brain Res.* 24, 386–404.
- Kane, M. J., and Engle, R. W. (2003). Working memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to stroop interference. *J. Exp. Psychol. Gen.* 132, 47–70.
- Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., et al. (2010). Control and interference in task switching: a review. *Psychol. Bull.* 136, 849–874.
- Monsell, S. (2003). Task switching. *Trends Cogn. Sci. (Regul. Ed.)* 7, 134–140.
- Polage, D. C. (2012). Fabrication inflation increases as source monitoring ability decreases. *Acta Psychol. (Amst.)* 139, 335–342.
- Sartori, G., Agosta, S., Zogmaister, C., Ferrara, S. D., and Castiello, U. (2008). How to accurately detect autobiographical events. *Psychol. Sci.* 19, 772–780.
- Schoenmakers, T., Wiers, R. W., Jones, B. T., Bruce, G., and Jansen, A. T. M. (2007). Attentional re-training decreases attentional bias in heavy drinkers without generalization. *Addiction* 102, 399–405.
- Spence, S. A., Farrow, T. F., Herford, A. E., Wilkinson, I. D., Zheng, Y., and Woodruff, P. W. R. (2001). Behavioral and functional anatomical correlates of deception in humans. *Neuroreport* 12, 2849–2853.
- Spence, S. A., Kaylor-Hughes, C. J., Brook, M. L., Lankappa, S. T., and Wilkinson, I. D. (2008). “Munchausen’s syndrome by proxy” or a “miscarriage of justice?” An initial application of functional neuroimaging to the question of guilt versus innocence. *Eur. Psychiatry* 23, 309–314.
- Squires, N. K., Squires, K. C., and Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalogr. Clin. Neurophysiol.* 38, 387–401.
- Stevens, A. A., Skudlarski, P., Gatenby, J. C., and Gore, J. C. (2000). Event-related fMRI of auditory and visual oddball tasks. *Magn. Reson. Imaging* 18, 495–502.
- Vandierendonck, A., Lefooghe, B., and Verbrugge, F. (2010). Task

- switching: interplay of reconfiguration and interference control. *Psychol. Bull.* 136, 601–626.
- Vendemia, J. M. C., Buzan, R. F., and Green, E. P. (2005). Practice effects, workload, and reaction time in deception. *Am. J. Psychol.* 5, 413–429.
- Verschueren, B., Ben-Shakhar, G., and Meijer, E. (2011a). *Memory Detection: Theory and Application of the Concealed Information Test*. Cambridge: Cambridge University Press.

- Verschueren, B., Spruyt, A., Meijer, E. H., and Otgaar, H. (2011b). The ease of lying. *Conscious. Cogn.* 20, 908–911.
- Vrij, A., Granhag, P. A., Mann, S., and Leal, S. (2011). Outsmarting the liars: toward a cognitive lie detection approach. *Psychol. Sci.* 20, 28–32.
- Vrij, A., Visser, R., Mann, S., and Leal, S. (2006). Detecting deception by manipulating cognitive load. *Trends Cogn. Sci. (Regul. Ed.)* 10, 141–142.

Conflict of Interest Statement: The authors declare that the research was

conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 24 July 2012; accepted: 06 November 2012; published online: 30 November 2012.

*Citation: Van Bockstaele B, Verschueren B, Moens T, Suchotzki K, Debey E and Spruyt A (2012) Learning to lie: effects of practice on the cognitive cost of lying. *Front. Psychology* 3:526. doi: 10.3389/fpsyg.2012.00526*

This article was submitted to Frontiers in Cognitive Science, a specialty of Frontiers in Psychology.

Copyright © 2012 Van Bockstaele, Verschueren, Moens, Suchotzki, Debey and Spruyt. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.

APPENDIX

Table A1 | List of the two lists of autobiographical questions that were used as either filler or test items (counterbalanced) in the experiment.

List 1	List 2
Did you go for a run?	Did you stop at a traffic light?
Did you go down a staircase?	Did you go to a supermarket?
Did you go up a staircase?	Did you buy some flowers?
Did you buy petrol?	Did you do the dishes?
Did you eat chocolate?	Did you take an elevator?
Did you take a bus?	Did you clean a window?
Did you take a train?	Did you reschedule an appointment?
Did you open a dustbin?	Did you read a book?
Did you take a bath?	Did you park a moped?
Did you make a sandwich?	Did you squeeze a lemon?
Did you post a letter?	Did you send an e-mail?
Did you close a door?	Did you stroke a pet?
Did you take a shower?	Did you wear a coat?
Did you buy a newspaper?	Did you open a fridge?
Did you buy a magazine?	Did you switch on a computer?
Did you use a knife?	Did you smoke a cigarette?
Did you use an umbrella?	Did you look at a watch?
Did you take a pill?	Did you open a water tap?
Did you speak to a police officer?	Did you lift a toilet seat?
Did you eat a grapefruit?	Did you use a pedestrian crossing?
Did you break a window?	Did you use an ATM?
Did you use a telephone?	Did you change money?
Did you receive a telegram?	Did you vacuum a carpet?
Did you drink fruit juice?	Did you drink cough syrup?
Did you listen to the radio?	Did you greet someone?
Did you use the internet?	Did you clean the house?
Did you stand in a queue?	Did you check your PO box?
Did you sit in a waiting room?	Did you brush your teeth?
Did you make your bed?	Did you listen to an MP3?
Did you wash your hands?	Did you sit on a bicycle?
Did you sign a document?	Did you stand on a ladder?
Did you drink coffee?	Did you sit on a chair?
Did you speak to a child?	Did you rip a piece of paper?
Did you watch television?	Did you water the plants?
Did you eat onions?	Did you use your keys?
Did you drink water?	Did you boil some water?



A repeated lie becomes a truth? The effect of intentional control and training on deception

Xiaoqing Hu^{1,2}, Hao Chen¹ and Genyue Fu^{1*}

¹ Department of Psychology, Zhejiang Normal University, Jinhua, China

² Department of Psychology, Northwestern University, Evanston, IL, USA

Edited by:

Matthias Gamer, University Medical Center Hamburg-Eppendorf, Germany

Reviewed by:

Nobuhito Abe, Kyoto University, Japan

Bruno Verschueren, Ghent University, Belgium

***Correspondence:**

Genyue Fu, Department of Psychology, Zhejiang Normal University, 688 Yingbin Road, Jinhua, China.

e-mail: fugu@zjnu.cn

Deception has been demonstrated as a task that involves executive control such as conflict monitoring and response inhibition. In the present study, we investigated whether or not the controlled processes associated with deception could be trained to be more efficient. Forty-eight participants finished a reaction time-based differentiation of deception paradigm (DDP) task using self- and other-referential information on two occasions. After the first baseline DDP task, participants were randomly assigned to one of three groups: a control group in which participants finished the same task for a second time; an instruction group in which participants were instructed to speed up their deceptive responses in the second DDP; a training group in which participants received training in speeding up their deceptive responses, and then proceeded to the second DDP. Results showed that instruction alone significantly reduced the RTs associated with participants' deceptive responses. However, the differences between deceptive and truthful responses were erased only in the training group. The result suggests that the performance associated with deception is malleable and could be voluntarily controlled with intention or training.

Keywords: training, intentional control, deception, instruction, differentiation of deception paradigm, automaticity

INTRODUCTION

Despite the claim that there is no unique lie-specific characteristic associated with lying or deception, such as Pinocchio's nose (cf. Rosenfeld, 1995), it has been widely accepted that lying requires greater amount of cognitive control than telling the truth. Studies from developmental psychology found that children's ability to tell lies are closely related with their development of executive control functions (Talwar and Lee, 2008). Studies from cognitive psychology similarly demonstrated that lying required more mental operations than truth (e.g., decisions to lie, construction of lying responses), which led to prolonged reaction times (RTs, e.g., Walczyk et al., 2003). Recently, research from cognitive neuroscience adds evidence that also supports the notion that lying is more task-demanding than truth: compared with truth, lying ubiquitously recruits brain regions that are involved in cognitive control such as the dorsolateral prefrontal cortex (DLPFC) and the anterior cingulate cortex (ACC, Spence et al., 2001; Langbehn et al., 2002; Sip et al., 2008). A recent meta-analysis of neuroimaging of deception showed that the brain regions involved during lying are highly overlapped with the brain regions involved in executive functions, especially working memory and response inhibition (Christ et al., 2009).

This attribute of lying was recently utilized in applied research to aid in deception detection. For instance, it has been shown that people are generally not good at spotting liars via behavioral cues (Bond and DePaulo, 2006). However, it was found that people are more accurate in detecting lies when liars' cognitive demand is high than when liars' cognitive demand is low. This is based on the idea that as lying is already task-demanding, liars whose cognitive demand is particularly high would find it more difficult to

manage lying as fewer resources are available, compared to liars whose cognitive demands are low, as relatively more resources can be used for lying (Vrij et al., 2008).

Although there are converging lines of evidence supporting the notion that lying is more task-demanding than truth-telling, this hypothesis should be investigated with more scrutiny given the recent evidences. Like many other complex social behaviors, lying is far from a uniformed homogenous behavior. There are increasing studies aiming to de-couple different sub-types of deceptions. For instance, people may tell lies either about others or about oneself; the event people may lie about could be experienced or not-experienced; the lies could also either be spontaneous or be well-practiced (Ganis et al., 2003, 2009; Abe et al., 2006; Johnson et al., 2008; Walczyk et al., 2009; Hu et al., 2011). These studies have consistently found that different types of lies showed different behavioral patterns, as well as non-overlapping neural activities. Specifically, lying about experienced events was associated with higher level of ACC activity compared to lying about not-experienced events, which is taken to suggest the former is associated with higher conflict (Abe et al., 2006). Moreover, it has been found that rehearsed, previously memorized lies were associated with less conflict compared with spontaneous lies, as evidenced by decreased activities in ACC (Ganis et al., 2003).

Although the abovementioned studies provided evidences that how cognitive demand may vary depending on different types of lies, whether or not the cognitive demand associated with lying can be intentionally reduced remains an open question. So far, only a few of studies investigated this issue. For instance, Johnson et al. (2005) found that although practice reduced the RTs of deceptive

responses generally, the difference between deception and truth still remained. Walczyk et al. (2009) gave participants time to prepare and practice their lies before a cognitive lie detection test. Results showed that participants' practiced deceptive responses were associated with reduced RTs than deceptive responses that had not been prepared nor practiced prior to the test (Walczyk et al., 2009; see also DePaulo et al., 2003, for preparation's influence on liars' behavioral cues). Another recent study manipulated the proportion of questions that required either honest or deceptive responses during a question set. It was found that when participants must deceive frequently in a question set, the lies became less task-demanding than when participants should tell the truth frequently. In other words, the more questions participants lied about, the easier it was to lie (Verschueren et al., 2011).

In the present study, we directly investigated whether or not lying can be trained to be more automatic and less task-demanding. We argued that since in most previous deception studies, participants were instructed to lie immediately after they receive the instruction, the lying can be classified as unpracticed. However, in real-life scenarios, liars may construct and practice lies before the interrogation. Indeed, practice or training may help people to improve the efficiency of knowledge retrieval, response inhibition and even working memory capacity across various task domains (Pirolli and Anderson, 1985; MacLeod and Dunbar, 1988; Milham et al., 2003; Olesen et al., 2004; Walczyk et al., 2009; Brehmer et al., 2012; Hu et al., 2012b). Since deception or lying has been conceptualized to rely on similar executive functions, especially working memory and response inhibition (Christ et al., 2009), it is possible that as these general-purposes processes (e.g., working memory, response inhibition) are malleable upon training, deception can also be trained to be more automatic. Thus, we hypothesized that participants who received training on deception would similarly find lying to be less demanding. Moreover, the post-training deception may not even be distinguished from truth.

In addition to the training condition, we also investigated the effect of instruction on deceptive responses. It has been found that giving participants specific instructions regarding response pattern can have considerable effects over participants' behavioral performance (Verschueren et al., 2009; Hu et al., 2012b). Specifically, participants can significantly reduce their RTs in tasks involving response conflict and control upon mere instruction (Hu et al., 2012b). This instruction group is also necessary for us to examine whether or not behavior changes between pre- and post-test, if any, can be attributed to training or to experimental instructions. For instance, it has been argued that the benefits of training on participants' performance may not necessarily due to the training itself, but can be due to factors such as participants' expectations about improvements (e.g., Brehmer et al., 2012). Thus, the instruction manipulation allows us to investigate the effect of mere instruction over one's deceptive responses. Furthermore, the comparison between the instruction group and the training group enables us to dissociate the behavioral change related to training from the change that is due to instructions. This may also provide us with a more detailed picture of the factors that may influence deceptions.

MATERIALS AND METHODS

PARTICIPANTS

Forty-eight participants (nineteen males, average years = 22.23) from Jinhua, China were recruited via advertisements on campus and received monetary compensation for their time. They were randomly assigned to three groups ($N = 16$ in each group): control group (seven males, mean age = 22.13), instruction group (four males, mean age = 22.5), and training group (eight males, mean age = 22.25). Consent forms were obtained from participants before the experiment.

MATERIALS

Three pieces of personal information from each participant were collected for self-referential information list: full name, birth-date, and hometown. Next, a list of names, dates, and Chinese city/town names were provided to participants, who were instructed to select those with special personal meanings (e.g., a city may become relevant because participants' relatives live there). Then three pieces of information, a name, a date, and a town name, were randomly selected from the list that contained only personal-irrelevant information. These three pieces of information were used as other-referential information.

Stimuli were presented as words using E-prime. Each item was presented for 15 times, resulting in a total of 90 trials [3(name, date, town) \times 2(self-referential vs. other-referential) \times 15] in one block. Stimulus was presented for 300 ms in white font against a black background on a computer monitor. The inter-stimulus-interval was randomly varied between 1500 and 2500 ms.

EXPERIMENTAL PROCEDURAL

The Differentiation of Deception Paradigm (DDP) was constructed following Furedy et al. (1988). The task consisted of two blocks: in the truthful block, participants were asked to respond to all stimuli honestly. They were asked to press one key indicating "self" to their self-referential information; and to press another key indicating "other" to the other-referential information. In the deceptive block, participants were asked to press "self" to the other-referential information and to press "other" to their own information, i.e., to pretend they were someone else while concealing their true identity. The order of the two blocks was counterbalanced across participants. Since participants may develop a response-mapping strategy from the first block to the second block by merely reversing the button press without experiencing being truthful or deceptive, another 30 trials of words "SELF" and "OTHER" were included in each block as catch trials. These catch trials were randomly interspersed among the self- and other-referential information. Specifically, participants were instructed to press the key indicating "self" to "SELF" catch trials and to press the key indicating "other" to "OTHER" catch trials. Importantly, this response-mapping was consistent across both truthful and dishonest blocks (for other types of catch trials, see Johnson et al., 2005; Hu et al., 2011). Thus, participants finished 120 trials (90 response trials and 30 catch trials) in each block. Speed and accuracy were equally emphasized.

Upon the completion of the baseline DDP for all participants, participants in the *control* group performed an irrelevant vision illusory task for 15 min, followed by a second DDP. In the irrelevant

task, participants watched a series of apparent motion pictures and decided whether the dots on the picture were moving or not. The control group aimed to control for possible effect of task familiarity/fatigue over one's behavioral performance across two tests.

For participants in the *instruction* group, their RTs and errors from the truthful and deceptive blocks of the baseline DDP they just finished were calculated and shown to them. Participants were debriefed regarding the meanings of these behavioral measures. They were explicitly told that their deception could be inferred from the increased RT and the decreased accuracy in the deceptive block compared to the RTs and accuracy from the honest block (in fact, every participants in the instruction group and the training group (described below) showed at least one of the two behavioral indicators associated with deception). Next they were instructed to try their best to speed up their RTs and to reduce possible incorrect responses during the deceptive block in the following DDP task. After the instruction was given, participants conducted the second DDP task in the same order as in the baseline DDP task.

For participants in the *training* group, everything was the same as in the instruction group, except that in addition to being instructed to speed up and be more accurate, they were given 360 trials (i.e., three deceptive blocks) that required deceptive response to improve their behavioral performance of the deceptive block. There were two intervals during the training session in which participants took a short break. After the training, participants proceeded to the second DDP in the same order as in the baseline DDP task.

RESULTS

The behavioral data from the baseline DDP and the second DDP from three groups is presented in **Figures 1** and **2**. It can be observed that the baseline deception is associated with longer RT and reduced accuracy. However, the RTs of deception were reduced in both the training group and the instruction group.

To statistically test our hypothesis, we first conducted a mixed-model 3 (group as a between-subject variable: control vs. instruction vs. training) by 2 (response type as a within-subject variable: truth vs. deception) by 2 (time as a within-subject variable: first vs. second DDP) by 2 (stimulus type as a within-subject variable: self- vs. other-referential information) analysis of variance (ANOVA) on RTs of correct responses. This test yielded a significant main effect of response type [$F(1, 45) = 116.19, p < 0.001, \eta_p^2 = 0.72$], indicating that deception took significantly longer time than truth (Mean \pm SE, 596.48 ± 11.76 vs. 520.79 ± 9.99 ms). The main effect of time was also significant, $F(1, 45) = 29.06, p < 0.001, \eta_p^2 = 0.39$. This was due to participants' faster RTs in the second DDP compared to the first DDP task (540.36 ± 9.96 in the second DDP vs. 576.92 ± 11.73 ms in the first DDP). Moreover, stimulus type was also significant, $F(1, 45) = 10.14, p < 0.01, \eta_p^2 = 0.18$, as self-referential information had faster RTs than other-referential information (553.44 ± 10.19 vs. 563.84 ± 10.73 ms). Regarding interactions, a significant two-way, stimulus type by response interaction was significant [$F(1, 45) = 83.56, p < 0.001, \eta_p^2 = 0.65$]. This was because the RTs discrepancy between honest and deceptive responses for self-referential information was larger (497.96 ± 9.73 vs. 608.93 ± 12.19 ms) than for other-referential information (543.64 ± 10.62 vs. 584.04 ± 12.09 ms).

Most importantly, the three-way response \times time \times group interaction was significant, $F(2, 45) = 8.26, p < 0.01, \eta_p^2 = 0.27$. No other effects were found significant. To understand this three-way interaction, we focused on the influence of time over response type by conducting 2 (first vs. second DDP) by 2 (deception vs. truthful responses) ANOVAs in three groups separately.

In the control group (see **Figure 1A**), there was no significant interaction between time and responses [$F(1, 15) < 1, p > 0.5, \eta_p^2 < 0.1$], suggesting that differences between deception and truth did not change across time.

In the instruction group (see **Figure 1B**), however, a significant time by response interaction was found: $F(1, 15) = 12.16, p < 0.01, \eta_p^2 = 0.45$. This suggested that mere instruction would

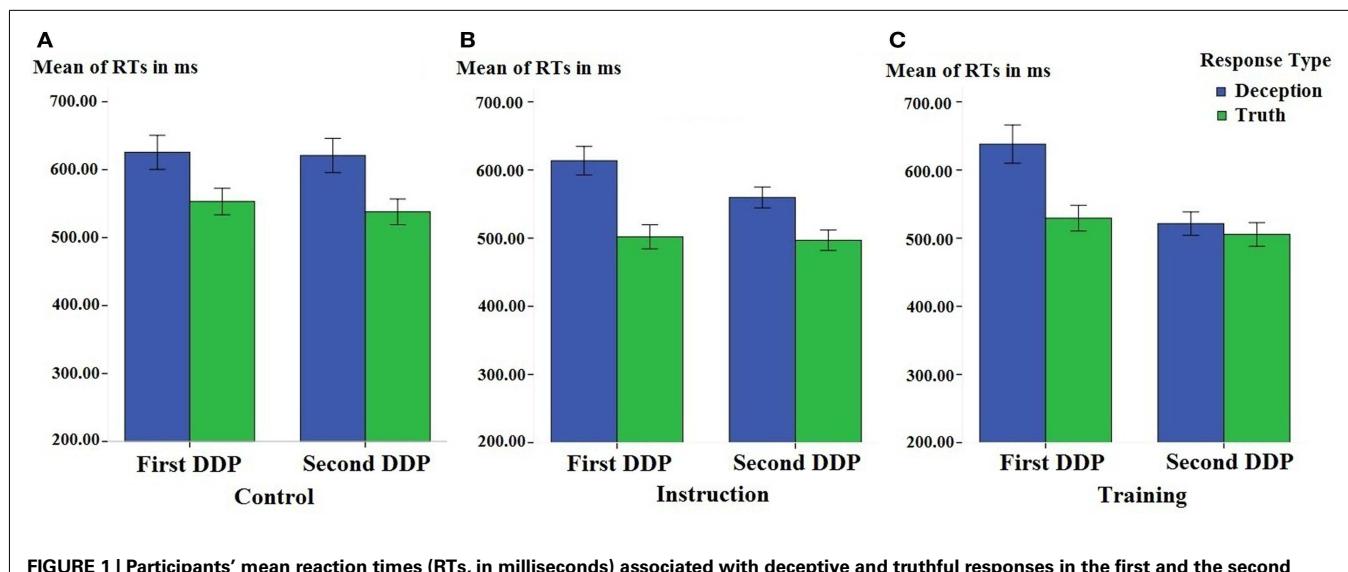


FIGURE 1 | Participants' mean reaction times (RTs, in milliseconds) associated with deceptive and truthful responses in the first and the second differentiation of deception paradigm task, in the control (A), instruction (B), and training group (C), separately. Error bars indicate ± 1 Standard Error.

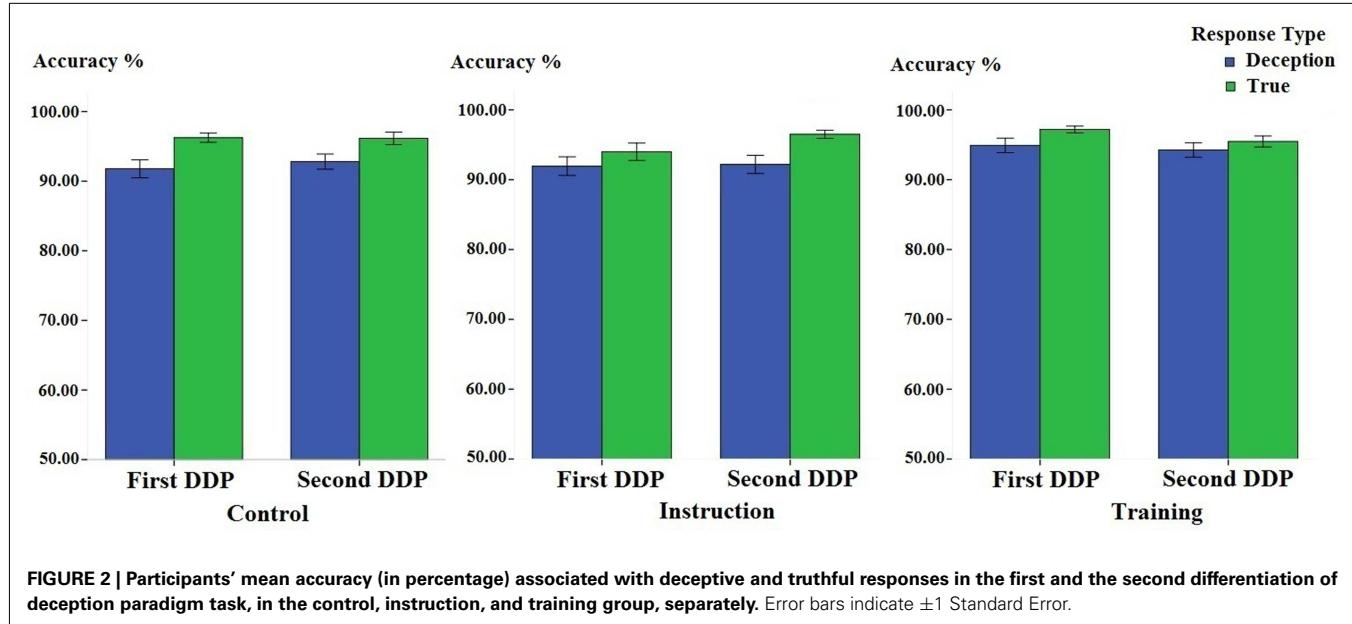


FIGURE 2 | Participants' mean accuracy (in percentage) associated with deceptive and truthful responses in the first and the second differentiation of deception paradigm task, in the control, instruction, and training group, separately. Error bars indicate ± 1 Standard Error.

significantly influence participants' behavioral performance of deception. To understand this interaction and to highlight our main variable of interest (i.e., differences between deceptive and honest responses), we calculated the differences between deception and truth blocks in the first and the second DDP separately. A paired sample *t*-test showed that participants who received the speed up instruction significantly reduced the differences between deceptive and honest responses from the first to the second DDP task (111.54 ± 12.98 vs. 62.73 ± 14.65 ms; $t(15) = 3.49$, $p < 0.01$, Cohen's $d = 0.88$).

Even though instruction did reduce the differences between deceptive and truthful response from the first to the second DDP task, it remained to be investigated whether or not RTs can distinguish deceptive from honest response in the second DDP task. A paired sample *t*-test comparing RTs of deceptive and honest responses found that the RTs associated with deceptive response were still longer than the RTs associated with honest responses (559.65 ± 15.36 vs. 496.92 ± 14.89 ms, $t(15) = 4.28$, $p < 0.001$, Cohen's $d = 1.04$). This pattern of results suggested that even though instruction did influence participants' deceptive responses, it was not sufficient to eliminate the deception-truth differences.

In the training group (see Figure 1C), the same 2(time: first vs. second DDP task) \times 2(response: deception vs. truth) within-subject repeated measure ANOVA resulted in a significant main effect of time: $F(1, 15) = 26.33$, $p < 0.001$, $\eta_p^2 = 0.64$, suggesting that the training significantly reduced the RTs of the DDP task. The same test also revealed a significant main effect of response type [$F(1, 15) = 20.02$, $p < 0.001$, $\eta_p^2 = 0.57$], suggesting that deception and truth was significantly different. Most importantly, the time by response interaction was significant: $F(1, 15) = 17.45$, $p < 0.001$, $\eta_p^2 = 0.54$.

To understand this interaction, we calculated the RTs difference between deceptive and truthful responses ($RT_{deception} - RT_{truth}$) of the first and the second DDP tasks separately. It was found

that this difference was significantly reduced after participants' training from the first to the second DDP (108.67 ± 22.69 vs. 15.82 ± 10.93 ms, $t(15) = 4.18$, $p < 0.001$, Cohen's $d = 1.30$). Moreover, in the second, post-training DDP, RTs for deceptive responses were not different from those of truthful responses (505.44 ± 17.39 for honest responses vs. 521.25 ± 17.13 ms for deceptive responses, $t(15) = 1.45$, $p > 0.1$, Cohen's $d = 0.23$). In other words, training eliminated the difference between deceptive and truthful responses in the second DDP task.

Regarding accuracy (see Figure 2), the same condition by response by time by stimulus type mixed-model ANOVA revealed only a significant main effect of response type [$F(1, 45) = 80.12$, $p < 0.001$, $\eta_p^2 = 0.64$], indicating deceptive responses was less accurate than honest responses (0.93 ± 0.01 vs. $0.96 \pm 0.01\%$). Neither other main effect nor interaction was significant (all $p > 0.05$). The accuracy results suggested that there was no speed-accuracy trade-off.

DISCUSSION

The present data show that behavioral RTs performance associated with deception can be influenced significantly via instruction and training, as evidenced by significantly decreased RTs in the second DDP compared with the baseline performance. This pattern of results also shows that deception is not always associated with higher cognitive demand, as most previous studies suggested.

Results from the baseline DDP task replicated previous findings that lying usually produces prolonged RTs when compared with truth (Furedy et al., 1988; Walczyk et al., 2003; Hu et al., 2011; Verschueren et al., 2011). The prolonged RT and reduced accuracy are usually taken as indicators of high response conflict and cognitive control in tasks such as the Stroop task (MacLeod and Dunbar, 1988). Evidence from neuroimaging studies also demonstrated that when people generate deceptive responses in DDP tasks, the brain regions associated with cognitive control and conflict monitoring processes were more active than when participants give honest

responses (e.g., Abe et al., 2006; for a meta-analysis, see Christ et al., 2009).

The present data, however, suggested that instruction and training could significantly decrease the task demand associated with deception as evidenced by reduced RTs. Specifically, the speed up instruction alone significantly reduced the difference between deceptive and honest responses. This result is also partially consistent with one recent study, in which it was found that instruction alone can result in speeding up RTs in an autobiographical implicit association test (aIAT) that involves response conflict and control (Hu et al., 2012b). Specifically, in an aIAT, participants are asked to perform a RT-based classification task that consists of four types of sentences: (1) true sentences (e.g., *I am in front of a laptop*), (2) false sentences (e.g., *I am climbing a mountain*), (3) crime-relevant sentences (e.g., *I stole a wallet*), and (4) crime-irrelevant sentences (e.g., *I read an article*). It is hypothesized that for criminals, it will be easier to press the same button to both crime-relevant sentences and true sentences given that both have truth values (i.e., congruent responses) than to press the same button to both crime-relevant sentences and false sentences (i.e., incongruent responses that involve conflict). Thus, the aIAT examines the mental associations between criminal events and truth value. Hu et al. (2012b) found that participants who were instructed to speed up their RTs in the incongruent blocks were able to reverse the baseline results pattern, i.e., showing quicker responses in the incongruent blocks than congruent blocks, thus obtaining an innocent diagnosis.

However, in the present study, participants who were similarly instructed to speed up their responses in the deception blocks only reduced, but did not eliminate or reverse, the differences between deceptive and honest responses compared to the baseline results pattern. Given this discrepancy, it is possible that the influence of instruction over one's performance depends on the nature of the specific type of response conflict and control involved in the task: in the autobiographical IAT task, the stimulus-response conflict involves in the incongruent responses concerns *recently established* mental associations (e.g., mock crime that was committed 10 min or weeks before the test, see also Hu and Rosenfeld, 2012); whereas in the self-other DDP task, however, the stimulus-response conflict involved in the deceptive responses concerns *long-term* mental associations (e.g., one's self-referential information is always true). Indeed, De Houwer and colleagues found that participants could successfully fake their performance of an IAT that assessed one's attitudes toward *novel* social groups (i.e., recently established mental associations, De Houwer et al., 2007).

Based on the discussion above, it is thus possible that the influence of instruction over performances in deception/response conflict tasks depends on the strength of mental associations: if the mental associations are newly acquired or recently established, it is likely that instruction alone will effectively help the participants to control the response conflict and behavioral performance. If the mental associations are established via long-term practice or socialization, however, it is likely that instruction itself is not sufficient for participants to overcome the response conflict involved in the task.

In addition to instruction, training here played an additive role in helping participants control their deceptive performance.

Specifically, after participants were trained to speed up their responses in the deceptive block, the honest-deception differences in the baseline were eliminated in the post-training DDP task. As discussed above, controlling response conflicts that are generated from long-term associations (here self-referential information refers to "self" instead of "other") may require training. Another example of response conflict that is generated from well-established association is the Stroop task (Stroop, 1935). In the Stroop task, people usually take longer to name the color of the incongruent word font (e.g., press the button indicating red color in response to the word "GREEN" printed in red color) than to name the word meanings since people (at least adults) are automatic in processing the meaning of the words (for a review, see MacLeod, 1991). Regarding whether or not training can reduce the Stroop effect, MacLeod and Dunbar (1988) employed a variant of the Stroop task and found that the stroop effect can be reversed only with extensive training as long as 20 h, rather than with relatively short training that last for 2 or 4 h.

Thus, though instruction alone is effective in reducing responses conflict that from recently established mental associations, training seems to be necessary in reducing conflicts that are from long established, well-practiced associations. Our results also extended another recent study, in which the deceptive responses could be made easier when the frequency of deceptive responses was increased in a question set (Verschuere et al., 2011). Together with these results, the current study, with an emphasis on training conducted within participants, supports the view that deception is malleable and its performance index can be voluntarily controlled to be more automatic.

One question arising here concerns the fact that deception seems to be more malleable than previous studies suggested (e.g., Johnson et al., 2005). Two possible reasons may be relevant: (1) previous studies showed that people may lie frequently in daily interaction (DePaulo and Bell, 1996). In other words, people may already "practice" lies in daily life, which makes lying more malleable; (2) more importantly, unlike previous studies in which participants merely repeated the tasks without an intention to improve (e.g. Johnson et al., 2005), participants in the present study practiced the deceptive responses with a conscious goal to speed up. Since mere practice (without an intention to improve) did not significantly change participants' task performance (Hu et al., 2012b), it seems that instruction is a necessary element in training-induced behavioral change.

The present data may also shed light on deception detection studies. Specifically, if a deception detection study involves comparisons between unpracticed lying and truth-telling, then the results may not generalize to situations where well-practiced lies are involved (see also Walczyk et al., 2009). Thus, some preparations of lying may be profitably included in deception detection studies so as to increase ecological validity.

A related question is how to better detect prepared lies. Although we did not directly investigate this question here, some recent findings may be helpful: (1) as in Verschuere et al. (2011)'s study, adding filler questions that required honest responses may increase the lie-honest differences. This is based on the premise that increasing the predominance of one response mode (e.g., honest) should make the competing response mode more difficult

(e.g., lies). Although participants may practice their lies before the test, some build-in filler questions that required honest responses during the test may make the prepared lies more difficult; similarly, Hu et al. (2012a) recently found that in a concealed information test, a higher number of irrelevant stimuli may make countermeasures or deliberate faking more difficult. (2) Although suspects can prepare some lies in anticipating certain questions, asking unanticipated questions for which liars may not be able to prepare may be helpful (see Vrij et al., 2009). A third strategy is to use certain algorithm to detect fakers based on their response patterns. For instance, Agosta et al. (2011) recently developed an algorithm to detect fakers in the aIAT contexts. Because of different task structures were used in the aIAT and the DDP, the algorithm cannot be directly applied here. Moreover, participants in the Agosta et al. (2011)'s study were asked to slow down their RTs to fake the test. A recent study showed that the algorithm based on slowing down RTs cannot be used in detecting fakers when they used the speeding up strategy, which was adopted in the present study (Hu et al., 2012b). Notwithstanding, future research should directly investigate whether or not prepared liars can be detected using certain abovementioned strategies.

A possible limitation of the present study is the relatively small sample size ($N=16$ in each group) we used here. However, it should be noted that as the effect sizes we obtained here were large (given the effect is considered as large when Cohen's $d > 0.8$), and because large sample is usually required to observe small effect, we reasoned that the present sample size would not render our results unstable (see also Hu et al., 2012b). Nevertheless, future studies using large sample are necessary to replicate the

effect we obtained here. Another possible caveat is the demand and expectancy effect may play a role here. However, it should be mentioned that unlike many psychological research in which the rationale/hypothesis of the study is concealed from participants, researchers in deception detection are usually interested in examining the extent to which participants can intentionally control their behavior during the test. This required participants to understand the rationale of the tests. This procedure is similar to many previous studies in deception detection studies involving countermeasures or deliberate faking strategies (e.g., Rosenfeld et al., 2004; Verschueren et al., 2009; Agosta et al., 2011; Hu et al., 2012b). Finally, it should be noted that although we obtained the instruction/training effect regarding RTs, we failed to find the similar pattern with accuracy. This may be due to the ceiling effect for accuracy results: in each group/condition, the accuracy was around 95%.

To conclude, this study showed that the performance of deception is malleable and becomes more automatic upon training. Meanwhile, instruction itself plays a significant role in inducing behavioral changes associated with deception. The results imply that future deception detection studies should take this variation of deception into account to better understand the complexity of lying and the corresponding behavioral/neural patterns, and to better identify liars.

ACKNOWLEDGMENTS

This study was supported by the National Natural Science Foundation of China (No. 31070894) and Program for Innovative Research Team in Zhejiang Normal University.

REFERENCES

- Abe, N., Suzuki, M., Tsukiura, T., Mori, E., Yamaguchi, K., Itoh, M., et al. (2006). Dissociable roles of pre-frontal and anterior cingulate cortices in deception. *Cereb. Cortex* 16, 192–199.
- Agosta, S., Ghirardi, V., Zogmaister, C., Castiello, U., and Sartori, G. (2011). Detecting fakers of the autobiographical IAT. *Appl. Cogn. Psychol.* 25, 299–306.
- Bond, C. F., and DePaulo, B. M. (2006). Accuracy of deception judgments. *Pers. Soc. Psychol. Rev.* 10, 214–234.
- Brehmer, Y., Westerberg, H., and Backman, L. (2012). Working-memory training in younger and older adults: training gains, transfer, and maintenance. *Front. Hum. Neurosci.* 6:63. doi:10.3389/fnhum.2012.00063
- Christ, S. E., Essen, D. C., Watson, J. M., Brubaker, L. E., and McDermott, K. B. (2009). The contributions of prefrontal cortex and executive control to deception: evidence from activation likelihood estimate meta-analyses. *Cereb. Cortex* 19, 1557–1566.
- De Houwer, J., Beckers, T., and Moors, A. (2007). Novel attitudes can be faked on the Implicit Association Test. *J. Exp. Soc. Psychol.* 43, 972–978.
- DePaulo, B. M., and Bell, K. L. (1996). Truth and investment: lies are told to those who care. *J. Pers. Soc. Psychol.* 71, 703–716.
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., and Cooper, H. (2003). Cues to deception. *Psychol. Bull.* 129, 74–118.
- Furedy, J. J., Davis, C., and Gurevich, M. (1988). Differentiation of deception as a psychological process – a psychophysiological approach. *Psychophysiology* 25, 683–688.
- Ganis, G., Kosslyn, S. M., Stose, S., Thompson, W. L., and Yurgelun-Todd, D. A. (2003). Neural correlates of different types of deception: an fMRI investigation. *Cereb. Cortex* 13, 830–836.
- Ganis, G., Morris, R. R., and Kosslyn, S. M. (2009). Neural processes underlying self- and other-related lies: an individual difference approach using fMRI. *Soc. Neurosci.* 4, 539–553.
- Hu, X., Hegeman, D. J., Landry, E., and Rosenfeld, J. P. (2012a). Increasing the number of irrelevant stimuli increases ability to detect countermeasures to the P300-based Complex Trial Protocol for concealed information detection. *Psychophysiology* 49, 85–95.
- Hu, X., Rosenfeld, J. P., and Bodenhausen, G. V. (2012b). Combating automatic autobiographical associations: the effect of instruction and training in strategically concealing information in the autobiographical Implicit Association Test. *Psychol. Sci.* 23, 1079–1085.
- Hu, X., and Rosenfeld, J. P. (2012). Combining the P300-complex trial-based Concealed Information Test and the reaction time-based autobiographical Implicit Association Test in concealed memory detection. *Psychophysiology* 49, 1090–1100.
- Hu, X., Wu, H., and Fu, G. (2011). Temporal course of executive control when lying about self- and other-referential information: an ERP study. *Brain Res.* 1369, 149–157.
- Johnson, R., Barnhardt, J., and Zhu, J. (2005). Differential effects of practice on the executive processes used for truthful and deceptive responses: an event-related brain potential study. *Brain Res. Cogn. Brain Res.* 24, 386–404.
- Johnson, R., Henkell, H., Simon, E., and Zhu, J. (2008). The self in conflict: the role of executive processes during truthful and deceptive responses about attitudes. *Neuroimage* 39, 469–482.
- Langleben, D. D., Schroeder, L., Maldjian, J. A., Gur, R. C., McDonald, S., Ragland, J. D., et al. (2002). Brain activity during simulated deception: an event-related functional magnetic resonance study. *Neuroimage* 15, 727–732.
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect – an integrative review. *Psychol. Bull.* 109, 163–203.
- MacLeod, C. M., and Dunbar, K. (1988). Training and Stroop-like interference – evidence for a continuum of automaticity. *J. Exp. Psychol. Learn. Mem. Cogn.* 14, 126–135.
- Milham, M. P., Banich, M. T., Claus, E. D., and Cohen, N. J. (2003). Practice-related effects demonstrate complementary roles of anterior cingulate and prefrontal cortices in attentional control. *Neuroimage* 18, 483–493.
- Olesen, P. J., Westerberg, H., and Klingberg, T. (2004). Increased prefrontal and parietal activity after training of working memory. *Nat. Neurosci.* 7, 75–79.

- Pirolli, P. L., and Anderson, J. R. (1985). The role of practice in fact retrieval. *J. Exp. Psychol. Learn. Mem. Cogn.* 11, 136–153.
- Rosenfeld, J. P. (1995). Alternative views of Bashore and Rapp (1993) alternatives to traditional polygraphy – a critique. *Psychol. Bull.* 117, 159–166.
- Rosenfeld, J. P., Soskins, M., Bosh, G., and Ryan, A. (2004). Simple effective countermeasures to P300-based tests of detection of concealed information. *Psychophysiology* 41, 205–219.
- Sip, K. E., Roepstorff, A., McGregor, W., and Frith, C. D. (2008). Detecting deception: the scope and limits. *Trends Cogn. Sci. (Regul. Ed.)* 12, 48–53.
- Spence, S. A., Farrow, T. F. D., Herford, A. E., Wilkinson, I. D., Zheng, Y., and Woodruff, P. W. R. (2001). Behavioural and functional anatomical correlates of deception in humans. *Neuroreport* 12, 2849–2853.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *J. Exp. Psychol.* 18, 643–662.
- Talwar, V., and Lee, K. (2008). Social and cognitive correlates of children's lying behavior. *Child Dev.* 79, 866–881.
- Verschueren, B., Prati, V., and De Houwer, J. (2009). Cheating the lie detector: faking in the autobiographical Implicit Association Test. *Psychol. Sci.* 20, 410–413.
- Verschueren, B., Spruyt, A., Meijer, E. H., and Otgaar, H. (2011). The ease of lying. *Conscious. Cogn.* 20, 908–911.
- Vrij, A., Leal, S., Granhag, P. A., Mann, S., Fisher, R. P., Hillman, J., et al. (2009). Outsmarting the liars: the benefit of asking unanticipated questions. *Law Hum. Behav.* 33, 159–166.
- Vrij, A., Mann, S. A., Fisher, R. P., Leal, S., Milne, R., and Bull, R. (2008). Increasing cognitive load to facilitate lie detection: the benefit of recalling an event in reverse order. *Law Hum. Behav.* 32, 253–265.
- Walczuk, J. J., Mahoney, K. T., Doverpike, D., and Griffith-Ross, D. A. (2009). Cognitive lie detection: response time and consistency of answers as cues to deception. *J. Bus. Psychol.* 24, 33–49.
- Walczuk, J. J., Roper, K. S., Seemann, E., and Humphrey, A. M. (2003). Cognitive mechanisms underlying lying to questions: response time as a cue to deception. *Appl. Cogn. Psychol.* 17, 755–774.
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- Received: 28 July 2012; accepted: 23 October 2012; published online: 12 November 2012.*
- Citation: Hu X, Chen H and Fu G (2012) A repeated lie becomes a truth? The effect of intentional control and training on deception. *Front. Psychology* 3:488. doi: 10.3389/fpsyg.2012.00488*
- This article was submitted to Frontiers in Cognitive Science, a specialty of Frontiers in Psychology.*
- Copyright © 2012 Hu, Chen and Fu. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.*



When Pinocchio's nose does not grow: belief regarding lie-detectability modulates production of deception

Kamila E. Sip^{1,2,3*}, David Carmel⁴, Jennifer L. Marchant^{5,6}, Jian Li⁷, Predrag Petrovic⁸, Andreas Roepstorff^{1,2}, William B. McGregor² and Christopher D. Frith^{1,6}

¹ Center of Functionally Integrative Neuroscience, Aarhus University Hospital, Aarhus, Denmark

² Department of Aesthetics and Communication - Linguistics, Aarhus University, Aarhus, Denmark

³ Department of Psychology, Rutgers University - Newark, NJ, USA

⁴ Department of Psychology, The University of Edinburgh, Edinburgh, UK

⁵ Institute of Cognitive Neuroscience, University College London, London, UK

⁶ Wellcome Trust Centre for Neuroimaging, University College London, London, UK

⁷ Department of Psychology, IDG/McGovern Institute for Brain Research Peking University, Beijing, China

⁸ Department of Clinical Neuroscience, Karolinska Institutet Stockholm, A Medical University, Stockholm, Sweden

Edited by:

Matthias Gamer, University Medical Center Hamburg-Eppendorf, Germany

Reviewed by:

Thomas Baumgartner, University of Basel, Switzerland

Nobuhito Abe, Kyoto University, Japan

***Correspondence:**

Kamila E. Sip, Social and Affective Neuroscience Lab, Department of Psychology, Rutgers University - Newark, Smith Hall, Room 301, 101 Warren Street, Newark, NJ 07102, USA.

e-mail: ksip@psychology.rutgers.edu

Does the brain activity underlying the production of deception differ depending on whether or not one believes their deception can be detected? To address this question, we had participants commit a mock theft in a laboratory setting, and then interrogated them while they underwent functional MRI (fMRI) scanning. Crucially, during some parts of the interrogation participants believed a lie-detector was activated, whereas in other parts they were told it was switched-off. We were thus able to examine the neural activity associated with the contrast between producing true vs. false claims, as well as the independent contrast between believing that deception could and could not be detected. We found increased activation in the right amygdala and inferior frontal gyrus (IFG), as well as the left posterior cingulate cortex (PCC), during the production of false (compared to true) claims. Importantly, there was a significant interaction between the effects of deception and belief in the left temporal pole and right hippocampus/parahippocampal gyrus, where activity increased during the production of deception when participants believed their false claims could be detected, but not when they believed the lie-detector was switched-off. As these regions are associated with binding socially complex perceptual input and memory retrieval, we conclude that producing deceptive behavior in a context in which one believes this deception can be detected is associated with a cognitively taxing effort to reconcile contradictions between one's actions and recollections.

Keywords: mock-crime, deception, beliefs, lie-detection, fMRI

INTRODUCTION

Deception is inherently social. Deceptive behavior involves not only the creation of a representation that is at odds with physical reality, but also the manipulation of another person's beliefs in a particular context (Sip et al., 2008a). This, in turn, means that deceivers must hold a belief about whether their deception is likely to be detected because a high likelihood of detection may lead to anxiety, altering the deceiver's emotional state and arousal level. Although several recent studies have attempted to elucidate the neural underpinnings of producing (e.g., Abe et al., 2007; Baumgartner et al., 2009; Kozel et al., 2009; Sip et al., 2010) and detecting (Grèzes et al., 2004, 2006) deceptive behavior, the role of beliefs about the detectability of deception remains poorly understood.

Behavioral research has shown that neither deceivers nor truthful people respond in the same way to all situations, as their behavior depends on their emotional state (Ekman and Friesen, 1969; Ekman, 1992), the complexity of what is said (Vrij, 2000; Vrij et al., 2001), and their need to control the impression

they make on others (Vrij, 1993). From a behavioral standpoint, therefore, there is no diagnostic cue that serves as a unique indication of deception (Vrij, 2004; Vrij et al., 2007; Vrij, 2008). This may be due to the complex nature of the demands that deceptive behavior places on the deceiver: it requires a series of conjectures about the deceived person's knowledge, the gap between this knowledge and the truth, the feasible manipulations this gap leaves room for, and the chances of getting caught. Deception is thus a sophisticated activity, involving a host of cognitive processes including memory, reasoning, and theory of mind. Furthermore, producing deception is emotionally taxing, and causes anxiety and physiological arousal that require effortful self-regulation (e.g., Abe et al., 2007; Baumgartner et al., 2009).

The multi-faceted act of attempting to deceive is therefore likely to require the concerted activity of several neural mechanisms, with activity in different, widely distributed brain regions mediating the various processes underlying deceptive behavior. Recently, a great deal of interest has centered on neuroimaging to test whether this technology could prove to be a useful

and reliable tool for lie-detection [for review see Greely and Illes (2007); Sip et al. (2008a,b)]. Several physiological (e.g., Bell et al., 2008; Gamer et al., 2010, 2012) and functional MRI (fMRI) studies (see e.g., Kozel et al., 2005, 2009; Mohamed et al., 2006) of mock-crimes have investigated the neural correlates of information inhibition and suppression that are associated with deceptive behavior. This previous research, however, has focused almost entirely on comparing deceptive vs. truthful behavior, neglecting the potential effects of participants' belief in the efficacy of lie-detection, and how such belief may modulate the neural activity underlying deception. Peoples' beliefs about whether or not their deception can be detected may affect activity in all brain regions that are involved in the production of this behavior. Alternatively, such belief may only modulate activity in a subset of these regions—for example, the belief that deception may be detected might alter activity in regions whose activity mediates the emotional aspects of deceptive behavior, but not those mediating aspects related to memory and reasoning. Clarifying this issue has both theoretical implications for understanding the systems underlying deception, and practical implications for the use of neuroimaging in forensic contexts.

In the current study we used a mock-theft paradigm to investigate whether people's beliefs about lie-detectability affect the brain activity that underlies the production of deception. Instead of focusing primarily on comparing the neural activity evoked by participants' false and true claims, we investigated whether people's beliefs regarding whether or not their false responses can be detected affect the brain activity underlying the production of these responses. By analogy to the well-known story of Pinocchio's growing nose, we asked: would Pinocchio's nose only grow when he *believed* his lies could be detected?

Subjective beliefs about the world and other people underlie most social and socio-economic decisions (e.g., Frith and Frith, 2003). Specifically, our beliefs and expectations modulate our emotional and physiological states, the way we interact with others, and how we make and evaluate choices (e.g., Pollina et al., 2004; Petrovic et al., 2005; De Martino et al., 2006; Mobb et al., 2006; Sip et al., 2010, 2012). Deception is an instance of belief manipulation, and is likely to rely on the deceiver's own beliefs.

Previous studies of deception have found increased activation in the amygdala—a region known to be involved in the processing of emotional information—when participants produce (Abe et al., 2007; Baumgartner et al., 2009) and detect deception (Grèzes et al., 2004, 2006). Additionally, several other regions known to mediate cognitive processes involving memory and reasoning, such as the inferior frontal gyrus (IFG), the anterior and posterior cingulate cortex (ACC, PCC, respectively) have been associated with producing false responses (e.g., Spence et al., 2001; Ganis et al., 2003; Langleben et al., 2005; Nuñez et al., 2005; Gamer et al., 2007; Sip et al., 2012). The amygdala, in particular, seems to be a likely candidate for modulation by production of deception and belief due to its central role in emotional processing and its ubiquitous involvement in belief-related tasks (Grèzes et al., 2004, 2006; Abe et al., 2007; Baumgartner et al., 2009). Grèzes et al. (2004, 2006) conducted two studies on non-verbal

deception in which participants either judged whether a third party deceived them (2004), or witnessed deception which they were not the target of themselves (2006). Increased activation in the amygdala was found in both studies only when participants detected that they were being deceived by a third party. More recently, other groups found amygdala activation to be associated with breaking a previously made promise (Abe et al., 2007; Baumgartner et al., 2009). Taken together, these studies suggest that the amygdala plays an important role in processing deception, regardless of whether one is personally engaged in producing it or is a target of deceit.

Inhibiting a choice of a risky option has been shown to be associated with risk evaluation and risk aversion in cases where no deception was involved (Aron et al., 2004; Christopoulos et al., 2009). The IFG has been implicated in production of deception where participants needed to inhibit their true responses (Langleben et al., 2005; Gamer et al., 2007). In a recent study by Sip et al. (2012), activation in the right IFG was observed when participants were deciding whether or not to produce a falsehood. This activation occurred regardless of which response, true or false, was made, which suggests that the IFG integrates contextual information about a risky choice rather than the value of a claim itself. It remains unknown, however, whether this region mediates any belief-related activity.

Here, we had participants commit a mock-crime (stealing a gadget they were motivated to keep) and then undergo a realistic interrogation, designed to induce increased anxiety, while undergoing fMRI scanning. Importantly, we manipulated their belief about the detectability of their deception by notifying them that a (fictitious) lie-detector was either active or inactive during different parts of the interrogation. We expected this manipulation to modulate activity in the network of brain regions previously associated with producing deception—the amygdala, IFG, ACC, and PCC. Our main goal was to find out whether belief would alter activity in all these areas, only some of them, or an entirely separate set of neural regions.

MATERIALS AND METHODS

PARTICIPANTS

Nineteenth healthy, right-handed participants with no reported neurological or psychiatric disorders, and from diverse social and professional backgrounds, took part in the experiment. Data from two participants were removed from the analysis. One admitted to stealing an object in the first few questions; the other fell asleep during the functional scans. The remaining 17 participants (7 females) were between 20 and 45 years old. Participants gave written informed consent to take part in the study, which was approved by the Joint Ethics Committee of the National Hospital for Neurology and Neuroscience (UCL NHS Trust) and Institute of Neurology (UCL).

STIMULI AND PROCEDURE

Upon arrival at the laboratory, each participant was given both written and verbal instructions. Participants were told that they would steal an item and that afterwards, as they were interrogated in the scanner while connected to a lie-detector, their brain activity would be monitored. Unknown to the participants, the

"lie-detector" was not real, and comprised two mock electrodes and a finger grip to imitate a polygraph test.

Two rooms were used in the mock-theft stage. The rooms were marked "red" and "blue" by pieces of appropriately colored paper placed on the inside and outside of each door. Each room contained typical office furniture and items, among which were a pair of earphones and a USB memory stick. The earphones and memory sticks were placed out of immediate view, in specific locations known to the researchers.

Each participant was escorted by the experimenter (author KES) to the corridor outside the red and blue rooms. Participants were informed that they had the right to refrain from taking part in the study, if it conflicted with their morals, and they would still be paid for participation. No participant took this option.

The participants were asked to enter each room and search it carefully in order to locate the earphones and the USB memory stick. They were asked to select one room and "steal" a single object from it. Participants could enter the rooms as many times as they wanted, but were asked to go into each room at least once in order to become familiar with both rooms and locate all the objects. After taking an object, they put it into an opaque bag provided by the experimenter, and hid it in a locker before going into the scanner.

In the scanner control room, the participant met the interrogator, who was introduced as an expert in the field of criminal investigation, with a specialty in polygraph tests (the interrogator was actually either author DC or PP, who are not, in fact, such specialists; one was assigned to each participant randomly). Before entering the control room they were told the interrogator did not know whether or not they had stolen anything, but only that they had been inside both rooms and had searched them. They were also told that if, by the end of the interrogation, the

interrogator could not tell whether they had taken an object, then they would get to keep the object they took (in fact, the interrogators were aware that all participants had taken an object, and half of the participants were selected at random and allowed to keep the stolen object). The interrogator explained the procedure of the interrogation, and presented the equipment that would ostensibly be used to measure skin conductance responses (or GSR, for galvanic skin responses, the acronym used during the interrogation). To illustrate "typical" skin conductance readings, the interrogator presented computer-generated graphs to the participants. These graphs were unrelated to real polygraph readings; one showed a relatively smooth line and, according to the interrogator, indicated "telling the truth," while the other was very spiky and indicated "lying." The aim of this presentation was to persuade the participants that the lie-detection device works reliably.

Participants were told that the "lie-detector" would enable the interrogator to discriminate between honest and deceptive responses. However, it would only be turned on for half the time during each scanning session, and they would be informed when this was happening.

The questions used during scanning were pre-recorded and played in a randomized order. Pre-recorded comments, such as "I see you're finding this difficult," were also used to maintain a realistic atmosphere. Depending on its content, each question was accompanied by a picture of either the red or blue room, or by a picture of one of the objects on an appropriately colored background (see **Figure 1**). Participants were asked to answer the questions by pressing keys marked yes/no on a response pad (two-specific keys on a four-key pad), as well as to mouth their response with a pre-specified noise—[mm]/əməm/for "no" and [mhmm]/əmhmm/for "yes"—to verify they were attending to

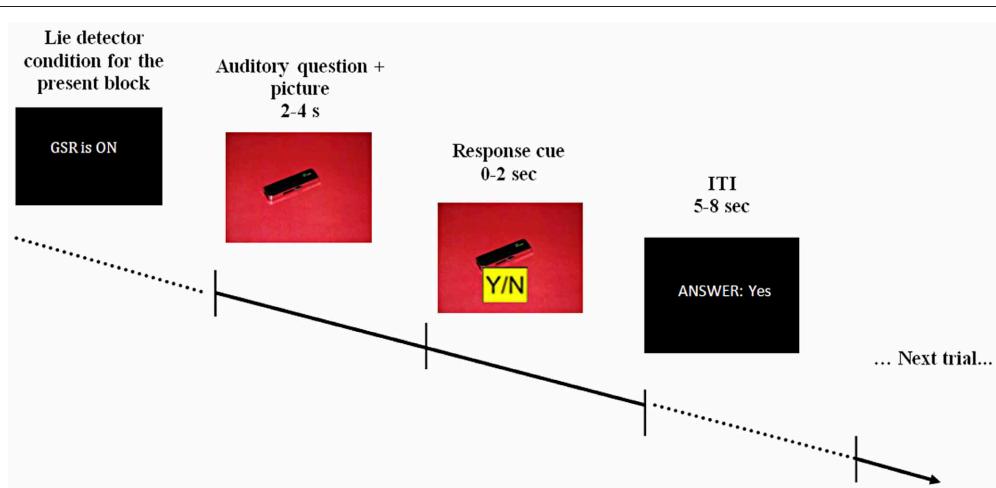


FIGURE 1 | A schematic example of the stimulus display. At the beginning of each block participants were told that the lie-detector (represented by the acronym GSR, for galvanic skin response) was either on or off. During the interrogation, pre-recorded auditory questions were read out over earphones, accompanied by appropriate visual presentations (question presentation took 2–4 s). After the question was completed, a response cue appeared on the screen for 2 s, during which participants

had to provide a response. The response cue (0–2 s) was randomly assigned on each trial (Y/N or N/Y) to prevent participants from pressing only one button as a default response. Participants' response (which could be either "yes," "no," "no response" if no response was given within the allotted time or "wrong button" if a button without an assigned meaning was pressed) was displayed on the screen for the duration of the 5–8 s inter-trial interval (ITI).

the task. Participants were informed both auditory and with written text each time the “lie-detector” was supposedly turned on or off.

Participants were not explicitly instructed to produce false statements, but merely motivated to try and keep the object they took. The questions used during the interrogation fell into several categories. A set of 12 personal questions, such as “Is your name John?” or “Are you British?” was used to acquaint participants with the procedure; they were told that such questions were used to establish a “baseline” for their skin conductance responses. A further set of 11 general questions allowed true responses to non-incriminating aspects of participants’ behavior, e.g., “*Did you go into the Red-room?*” which would always elicit a true response of “yes,” because all the participants were asked to perform the same set of actions. Together, the personal and general questions helped establish a realistic atmosphere.

The crucial part of the interrogation consisted of 35 theft-related questions, divided between truth- and falsehood-eliciting based on each participant’s choice of stolen object: (1) theft-related falsehood-eliciting questions (14 related to the theft of an object that the participant actually took); and (2) 21 theft-related truth-eliciting questions (related to the theft of an object that the participant did not take). The nature of a specific theft-related question (truth- or falsehood-eliciting), as well as the number of questions of each type, was determined by context. For example, a question such as “*Did you take earphones from the Red-Room?*” would be a theft-related falsehood-eliciting question, to which the participant would respond “no,” if they took the object from that room. The same question would be a theft-related truth-eliciting question (again evoking the response “no”), if the participant took the earphones from the other room. The experiment thus had a 2 (belief: lie-detector on, lie-detector off) \times 2 (behavior: true, false) factorial design. Each participant was scanned twice, with each of the two scanning sessions divided into one half with the lie-detector “on” and the other with the lie-detector “off.” The order of the “on” and “off” conditions was randomly assigned and counterbalanced across participants.

An eye-tracker (ASL E-5000) was used to make sure participants did not fall asleep or close their eyes to avoid looking at the visual stimuli. The participants used a plastic box with four push buttons to register their responses.

In a written post-scan questionnaire, participants rated emotions they may have experienced during the interrogation (e.g., upset, anxious, nervous); whether they felt guilty about stealing the object in question; their confidence in getting away with lying (and whether this differed when the lie-detector was active or not), and their motivation to keep the stolen object. Participants responded using a 0–5 scale where 0 means “not at all” and 5 means “a lot.” Additionally, participants were asked whether they had tried to use any strategy to deceive the interrogator, and if they had, to describe this strategy.

fMRI SCANNING AND PREPROCESSING

A 1.5 Tesla Siemens Sonata MRI scanner (Siemens, Erlangen, Germany) was used to acquire T1-weighted anatomical images and T2*-weighted echo-planar functional images with blood oxygenation level-dependent (BOLD) contrast (35 axial slices,

2 mm slice thickness with 1 mm gap, 3 \times 3 resolution in plane, slice TE = 50 ms, volume TR = 3.15 s, 64 \times 64 matrix, 192 \times 192 mm FOV, 90° flip angle). During two functional EPI sessions, an average of 221 whole brain volumes (range 214–225 depending on participants’ response speed) were acquired. The first 4 volumes were discarded to allow for T1 contrast to reach equilibrium.

Image processing was carried out using SPM8 (Statistical Parametric Mapping software, Wellcome Trust Centre for Neuroimaging, UCL; www.fil.ion.ucl.ac.uk/spm) implemented in MATLAB (The Mathworks Inc., Massachusetts, USA; www.mathworks.com). EPI images were realigned to correct for movements by aligning the functional (T2*-weighted EPI) images of each run to the first volume using a six-parameter rigid body transformation. Mean functional images were then coregistered to the T1-weighted anatomical image and normalized into Montreal Neurological Institute (MNI) template space using a 12-parameter affine transformation (parameters were estimated from segmentation and normalization of anatomical images to MNI template using SPM8). Normalized functional images were resampled into 2 \times 2 \times 2 voxel resolution. A Gaussian kernel with a full width at half maximum of 6 mm was applied for spatial smoothing.

fMRI ANALYSIS

In a statistical model that included all events in the scanning run, each event was convolved with the standard haemodynamic response function of SPM8 (Holmes and Friston, 1998). The design matrix comprised a column for each experimental condition, with separate events defined by their onset time and duration (based on participants’ response times). The fit to the data was estimated for each participant using a general linear model (Friston et al., 1995) with a 128 s high-pass filter, global scaling, and modeling of serial autocorrelations.

Individual T-contrasts related to the different conditions within a factorial design comprising the conditions of interest (2 factors: lie-detector on vs. off, and true vs. false response) were created from the parameter estimates (beta weights). T-contrasts were computed within subjects for the main effects and interaction between belief about whether the lie-detection device was active and the type of response (true or false) to theft-related questions. These were then used in separate second level random effects analyses in order to facilitate inferences about group effects (Friston et al., 1995). Results are reported for clusters with at least 10 voxels and a significance threshold of $p < 0.001$ (uncorrected for multiple comparisons; Wager et al., 2007). Missed trials were modeled by a regressor of no interest in the GLM analysis. All brain loci are reported in MNI coordinates. Anatomical loci were determined using the Wake Forest University PicAtlas and were double checked against the Harvard-Oxford probabilistic atlas using a 50% probability threshold (Desikan et al., 2006).

RESULTS

DEBRIEFING

All participants claimed to have been highly motivated to keep the object they took. Interestingly, 14 of the 17 participants chose

to take the memory stick rather than the earphones, claiming in debriefing that they found it more appealing; the fact that the choice was not random confirms that the task was engaging and personally relevant.

Eight of the 17 participants reported that they had tried to use strategies to avoid detection. Strategies included attempting to control their breathing, focusing on something else, silently repeating in their heads *I didn't steal anything*, or trying to prolong their response times when giving truthful answers in an attempt to confuse the interrogator (e.g., one participant said "*I would delay giving a response when asked about the object I didn't steal to create confusion*").

All the participants reported that they found the interrogation realistic (i.e., none of them suspected that the questions they were asked were actually pre-recorded), though unsurprisingly, some of them noted that they would have been more nervous if the interrogation had not taken place in the context of an experiment. The majority of the participants (12 out of 17) reported that they found it easier to lie when they were told that the lie-detector was inactive.

BEHAVIORAL RESULTS

To examine whether the belief that a "lie-detector" was active affected participants' production of deceptive responses, we examined reaction time (RT) data (Nuñez et al., 2005; Abe et al., 2007; Kozel et al., 2009). RTs were calculated as the duration from the end of a question to the participant's button response. A 2 (belief: "lie-detector" on, "lie-detector" off) \times 2 (question type: theft-related truth-eliciting, theft-related falsehood-eliciting) repeated-measures ANOVA revealed no main effects [belief: $F_{(1, 16)} = 0.169, p = 0.69$; question type: $F_{(1, 16)} = 0.00, p = 0.97$], and no interaction between belief and question type [$F_{(1, 16)} = 2.381, p = 0.142$; see **Figure 2**]. The similarity between the RTs evoked by questions in the different conditions calls into question previous

reports (e.g., Nuñez et al., 2005; Abe et al., 2007; Kozel et al., 2009), which suggested that RTs could be used to distinguish deceptive and truthful behavior (but see the Discussion, where we note the limitations of using RTs in the present context).

Interestingly, examination of the general questions indicated that they evoked longer RTs than theft-related ones. Indeed, including them in the statistical analysis, by running a 2 (belief: "Lie-detector" on, "Lie-detector" off) \times 3 (question type: theft-related truth-eliciting, theft-related falsehood-eliciting, general truth-eliciting) repeated-measures ANOVA revealed a main effect of question type [$F_{(2, 32)} = 10.1, p < 0.05$], but no main effect of belief [$F_{(1, 16)} = 0.71, p = 0.41$] nor an interaction between belief and question type [$F_{(2, 32)} = 1.78, p = 0.19$] (**Figure 2**). To investigate the main effect further, *post-hoc* paired *t*-tests [corrected for multiple comparisons using the sequential Bonferroni method (Holm, 1979; Rice, 1989) and collapsed across the belief conditions, as there was no main effect of belief] were conducted. The tests indicated that participants' responses to general questions were slower than to either the theft-related falsehood-eliciting [$t_{(16)} = 3.45, p < 0.05$] or theft-related truth-eliciting questions [$t_{(16)} = 3.31, p < 0.05$]. RTs to theft-related truth-eliciting and theft-related falsehood-eliciting questions did not differ [$t_{(16)} = 0.02, p = 0.99$]. These findings suggest that the increased arousal caused by being asked theft-related questions may have increased the speed with which participants responded to such questions, but the specific content of the questions—whether or not they referred to the object the participant had stolen—did not modulate response times. A different possibility that must be acknowledged is that the pre-recorded theft-related questions were easier to discern while they were still being read out, leading to uniformly faster responses than general questions did.

Three participants explicitly stated in the post-scan questionnaire that they tried to slow their truthful responses in order to mislead the interrogator. However, the behavioral data show that although these three participants made slower responses overall, the patterns of their RTs did not differ from the rest of the group. Despite their claims, their response times were actually slightly faster for true compared to false claims. Excluding these participants did not alter the pattern or significance of any of the analyses reported.

Participants missed an average of 3.62 trials ($SD = 4.1$) out of a total of 104 trials. One participant missed 14 trials and was the only outlier in terms of missed responses (>3 standard deviations from the mean). This participant's behavioral responses were otherwise within 3 standard deviations from the mean on all measures, and excluding this participant did not alter the pattern or significance of any of the analyses reported.

IMAGING RESULTS

To examine the effect of belief on the brain activity underlying the production of deception, we examined BOLD responses evoked by questions in a factorial design with the factors belief (lie-detector on or off) and behavior (true or false responses). Investigations comparing the neural activity associated with true and false responses have been carried out before, and we expected to find increased activation for false (compared

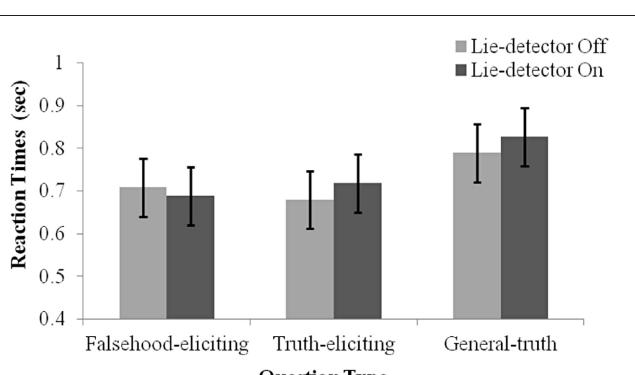


FIGURE 2 | Mean RT under the different conditions. Separate means are given for false, true and general responses with the lie-detector "on" and "off". Error bars represent one standard error of the mean. Participants' responses were slower for general questions than for theft-related questions. RTs to truth- and falsehood-eliciting theft-related questions did not differ, and RTs were not modulated by whether the lie-detector was "on" or "off".

to true) responses in similar regions to those found in those previous studies (Ganis et al., 2003; Langleben et al., 2005; Abe et al., 2007; Baumgartner et al., 2009; Kozel et al., 2009; Sip et al., 2012): amygdala, IFG, and PCC. Our main question, however, was whether the difference between the neural activation evoked by false and true responses would be modulated by participants' beliefs about whether their deception could be detected, and whether such modulation would occur in all or only in a subset of the regions that process deception production.

Significantly activated regions identified in the second level analysis are detailed in **Table 1**. The tests revealed a main effect of response type, whereby producing deceptive responses was associated with higher BOLD activation, in the right amygdala and IFG, and in the left PCC (**Figure 3**). There were no regions in which a main effect in the opposite direction (true > false) was observed, and no regions showed a main effect of belief in either direction (lie-detector on > off or off > on).

In addition to the main effects reported above, we found a significant interaction between belief and behavior in two regions: the right hippocampus/parahippocampal gyrus (**Figures 4A,B**) and the left temporal pole (**Figures 4C,D**), regions that have both been previously associated with social processes such as theory of mind and face recognition (Olson et al., 2007), and deceptive decision-making (Ganis et al., 2003; Mohamed et al., 2006). Examination of the patterns of responses in these regions reveals that the interaction was due to greater activation when producing deceptive, compared to truthful, responses when the lie-detector was believed to be on, and a reversed pattern when the lie-detector was believed to be off.

To further investigate the effects underlying the interaction, we analyzed the BOLD responses associated with the simple effects of deceptive vs. truthful responses in each belief condition. For the right hippocampus/parahippocampal gyrus, we found that when participants believed the lie-detector was on, activation when producing deceptive responses was significantly greater than when producing truthful responses [$t_{(16)} = 5.397, p < 0.001$]. However, when the lie-detector was believed to be off both kinds of response were reduced and were not significantly different from each other [$t_{(16)} = 1.6, p = 0.14$]. Belief that lies could be detected thus led to differential responses in this region. For the left temporal pole, there was again significantly greater activation

when producing deceptive (compared to truthful) responses in the "lie-detector on" condition [$t_{(16)} = 2.54, p < 0.05$]. However, this difference was reversed in the "lie-detector off" condition, in which truthful responses led to significantly greater responses than deceptive responses [$t_{(16)} = 3.643, p < 0.01$].

DISCUSSION

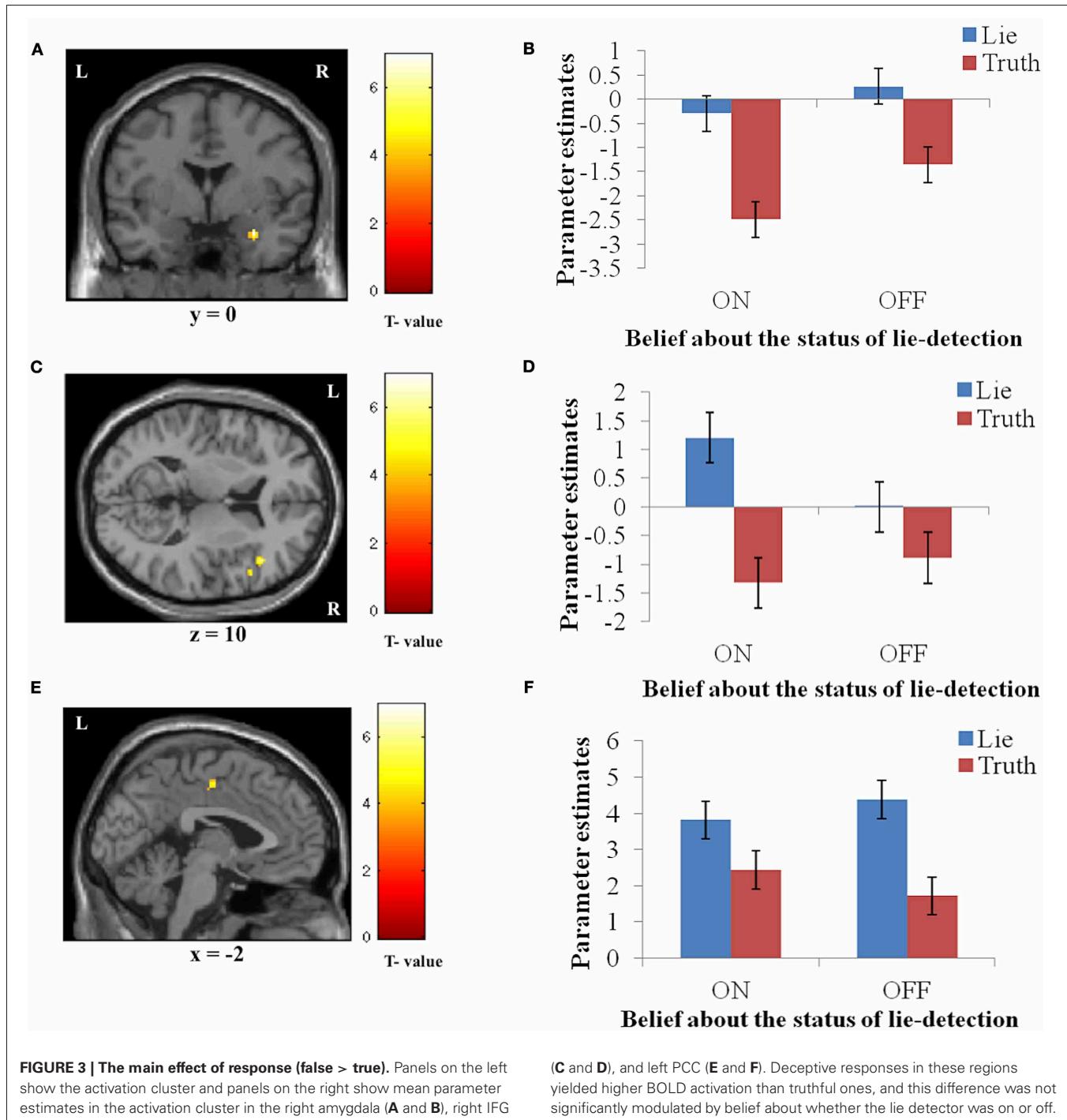
We conducted an fMRI investigation to test whether beliefs about how detectable deception was would affect the neural activity involved in producing it. Specifically, we studied the effect of a belief that a lie-detector was on or off on the neural processing underlying deception. Our results show that a belief in the assumed efficacy of lie-detection does indeed modulate the neural activity in a subset of the regions—the right hippocampus/parahippocampal gyrus and left temporal pole—involved in producing false claims, such that false responses led to greater activity than true responses when participants believed lie-detector was inactive. This difference was not present (and in the left temporal pole was reversed) when the lie-detector was believed to be off. Other regions—right amygdala, right IFG, and left PCC—were more active when producing a false claim, but this difference was not modulated by belief in lie-detectability.

Replicating our previous findings (Sip et al., 2010, 2012), our behavioral results showed that in an ecologically valid scenario there was no difference in RTs for producing true and false statements in a context in which both can be used deceptively. These findings are at odds with other deception studies which have found faster responses when participants were being truthful, compared to when they were producing a false claim (e.g., Kozel et al., 2005; Langleben et al., 2005; Spence et al., 2008; Seymour and Fraynt, 2009). This discrepancy, however, might stem from the realistic experimental paradigm we employed, which may have encouraged some of the participants to attempt to use strategies that would mislead the interrogator. Indeed, during debriefing we learned that some participants had tried to use response timing as a countermeasure to detection. This suggests that people produce deception in various ways if they are allowed to use their own deceptive strategy. However, it must also be noted that in the current study, we used auditory questions combined with visual presentation of relevant items. The visual stimuli may have interfered with auditory processing, or facilitated response preparation such that

Table 1 | Brain regions showing activation during response production.

Brain region	Hemisphere	x	y	z	t-value	Cluster size
MAIN EFFECT OF RESPONSE (FALSE > TRUE)						
Amygdala	R	30	0	-24	6.98	17
Inferior frontal gyrus (IFG)	R	44	26	10	6.24	25
Posterior cingulate cortex (PCC)	L	-2	-12	50	4.83	10
INTERACTION (ON FALSE-TRUE > OFF FALSE-TRUE)						
Hippocampus/parahippocampal gyrus	R	36	-18	-18	4.98	26
Temporal pole	L	-44	14	-22	4.89	17

Peak activation coordinates in standard MNI space and their associated t-scores. Regions shown were significantly activated at a threshold of $p < 0.001$ (uncorrected) with a cluster extent threshold of 10 voxels.



participants could have decided what response to provide before the question was fully articulated. However, the actual response could only be provided after the question was posed, so calculating RTs as the elapsed time from the end of the question was the only way to avoid additional assumptions regarding the point in time at which participants decided what answer to give. This calculation also avoided false-positive difference in RTs that might be caused by differences in the lengths of the posed questions.

Our neuroimaging results demonstrate that the assumption that the same brain regions would always be either active or inactive when one tells a lie or the truth, respectively (Mohamed et al., 2006) is an oversimplification. Neural activity in various regions, including the ACC, DLPFC, IFG, the caudate nucleus, and the amygdala (e.g., Kozel et al., 2005; Baumgartner et al., 2009; Greene and Paxton, 2009; Sip et al., 2010, 2012; Gamer et al., 2012) has been implicated in the production of deception. The present findings involve a smaller set of areas than

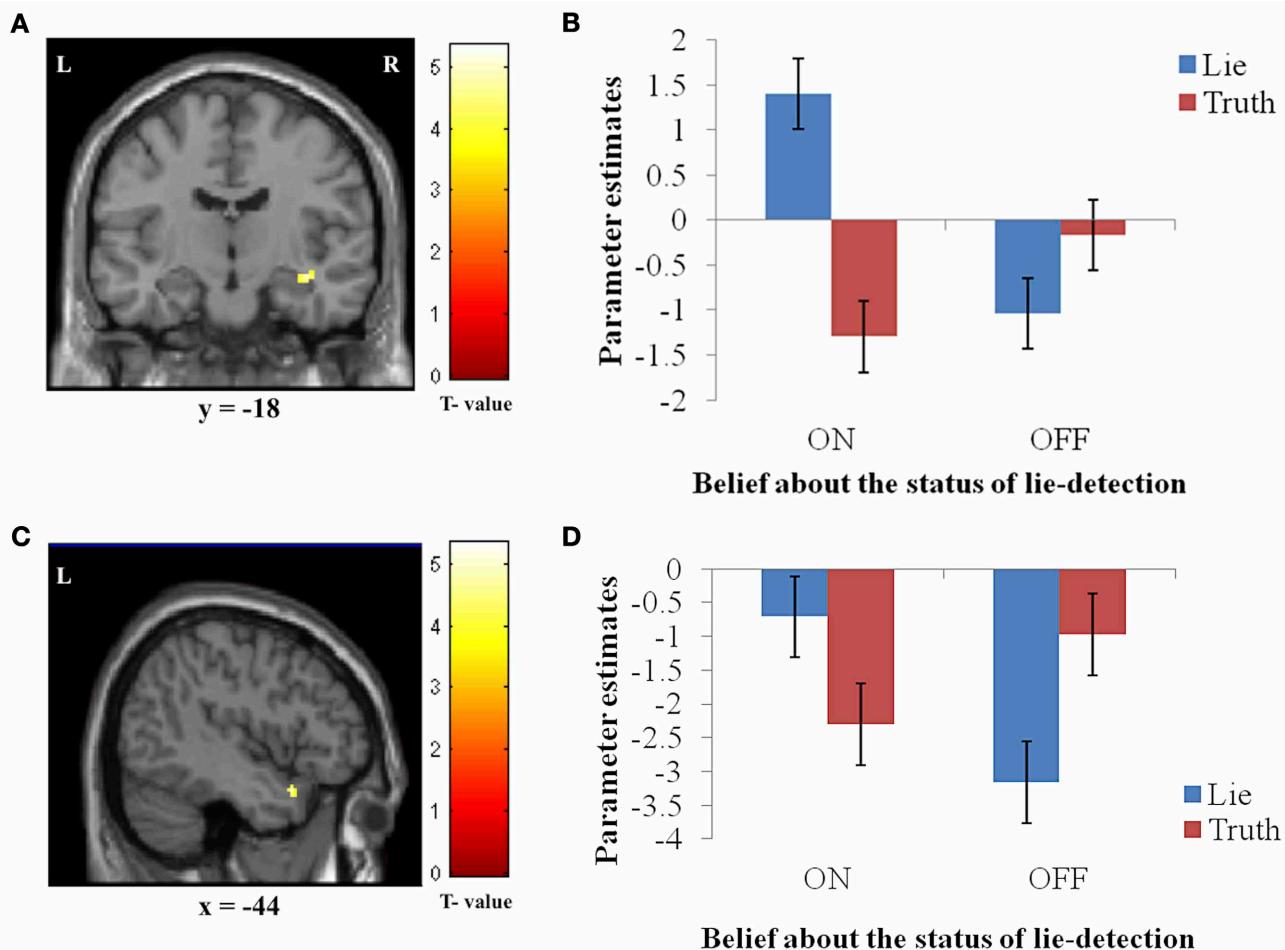


FIGURE 4 | The interaction between response type (true or false) and belief about the lie-detector (on or off). Panels on the left show the activation cluster and panels on the right show parameter estimates in the activation cluster in the right hippocampus/parahippocampal gyrus (**A** and **B**)

and left temporal pole (**C** and **D**). In these loci, the difference between the BOLD activation caused by false vs. true responses was abolished (and for the hippocampus/parahippocampal gyrus, reversed) when participants believed the lie detector was off.

reported in previous neuroimaging studies of deception [for a review see Sip et al. (2008a)]. Unlike these previous studies, we did not observe activation in dorsolateral prefrontal cortex (DLPFC), ACC, or the caudate nucleus. The fact that we found activation in a smaller set of regions than previously reported could be due to several factors that are not substantive to the issue of deception, such as the specific statistical model and significance thresholds employed in different studies, specific characteristics of the participant cohort, or the visual and auditory stimuli used in the course of the interrogation. We speculate, however, that a substantive factor—the realistic nature of the mock-theft scenario used in the present study—might also potentially be at play. Such scenarios have been shown previously to reduce participants' physiological arousal (indicated by skin conductance) during interrogation, compared to more standard experimental procedures (though it must be noted that this was observed in the context of a different method for lie-detection, and may have been modulated by reduced memory for crime-related items; Carmel et al., 2003). Although negative findings (the absence of

activation in particular brain regions) must always be interpreted with extreme caution, further work may benefit from attempting to address the relation between how realistic a mock-crime scenario is and how widespread neural activation across the brain is during interrogation.

MAIN EFFECTS: DECEPTIVE vs. TRUTHFUL RESPONSES

Deceptive responses produced greater BOLD responses than truthful responses, regardless of the belief condition, in three regions: the right amygdala, right IFG, and left PCC. The amygdala and IFG have been implicated in recent ecologically valid examinations of deception (Abe et al., 2007; Baumgartner et al., 2009; Sip et al., 2012). Here, the observed activation in the amygdala, which is known to be involved in processing emotionally relevant information [for a review see Dolan (2007); Olson et al. (2007)], suggests that participants experienced an emotional conflict resulting from making false claims while risking a potential confrontation, and that this occurred regardless the believed status of the lie-detector device. Abe et al. (2007) were the

first to report amygdala involvement in producing verbal deception, employing a realistic scenario in which participants underwent interrogation. They speculated that emotional processing, reflected in the increased amygdala activation they observed, was associated with attempts to deceive the interrogator. In a different study, Baumgartner et al. (2009) showed that breaking a previously expressed promise and consequently deceiving others in a social context appears to create anxiety associated with social consequences of the act rather than with producing false claims *per se*.

In previous studies, the PCC has been implicated in processing the emotional aspects of context and in integrating emotion- and memory-related processes (Mohamed et al., 2006). Here we observe increased activation for producing false vs. true claims, suggesting that the cognitive load associated with deception places demands on emotional processing. This specific processing, however, was not modulated by belief in lie-detectability, indicating that it is largely independent of those processes that mediate the emotion and anxiety engendered by the context of such belief. Previous studies have also shown right IFG involvement in deception (Gamer et al., 2007; Sip et al., 2012) as well as in response inhibition (Aron et al., 2004) and risk aversion (Christopoulos et al., 2009). Interestingly, right IFG was previously involved in production of deceptive responses in a social context where participants had to first comprehend the question, and then choose to whether to inhibit a true response and claim falsehood instead (Sip et al., 2012). The present findings thus suggest that the right IFG plays a generalized role in deception that is related to monitoring response release, and that this process is unlikely to be modulated by belief about lie-detectability.

INTERACTION OF DECEPTIVE/TRUTHFUL RESPONSE AND BELIEF ABOUT LIE-DETECTABILITY

We found two regions, the right hippocampus/parahippocampal gyrus and left temporal pole, in which response and belief interacted significantly to produce greater BOLD activation for deceptive responses when the lie-detector was believed to be on, but not when it was believed to be off. The temporal pole has been implicated in various socio-emotional processes involved in broadly construed theory of mind (Carr et al., 2003; Frith and Frith, 2003; Völlm et al., 2006), moral judgments (Moll et al., 2002; Heekeren et al., 2003), and deception detection (Grèzes et al., 2004, 2006). Olson et al. (2007) suggested that this region thus combines emotional responses with highly processed sensory stimuli. In our study, the increased temporal pole activity we observed when the lie-detector was “on” may be due to participants attempting not only to regulate their own emotional responses but also to infer the emotional states and beliefs of their interrogator. The realistic interrogation scenario, involving an ostensible “real-life interrogator,” may have increased participants’ anxiety and contributed to the modulation found in the activity of this region (which is known to have reciprocal anatomical connections to the amygdala; Dolan, 2007; Olson et al., 2007). The pattern of responses in the temporal pole was reversed in the “lie-detector off” condition, in which truthful responses led to greater activation than deceptive ones. The functional

significance of this reversal remains unclear and requires further elucidation.

We also observed a differential activation pattern in the right hippocampus/parahippocampal gyrus, where BOLD activity differed for deceptive and truthful responses (and was greater for deceptive ones), but only when the lie-detector was believed to be on. We had not originally included these areas amongst those in which we expected to find differential activation—although the hippocampus has been previously associated with producing deceptive response (Mohamed et al., 2006), and the parahippocampal gyrus has been associated with reporting autobiographical memories (which participants must draw on to produce truthful and deceptive responses; Ganis et al., 2003), neither region has been reported as consistently as other regions in the context of deception [for an overview, see e.g., Sip et al. (2008a)]. The differential activation we find here suggests that these areas may play a role related to belief, which had not been tapped into by previous studies where this factor was not manipulated.

The hippocampus is known to play a central role in memory (e.g., Burgess et al., 2002) as well as predictions about upcoming events related to past experiences [for a review see e.g., Buckner (2010)]. A previous investigation of neural connectivity (Smith et al., 2006) has shown that not only the content of a memory but also the context in which a memory was created have a measurable impact on episodic retrieval and interpersonal communication. It is thus noteworthy that the cluster of activation that included the hippocampus also extended to the parahippocampal gyrus. In a social context the parahippocampal gyrus (as well as temporal pole) allows for a proper identification of communicational intent, as demonstrated in a previous study of sarcasm (Rankin et al., 2009). A seemingly insincere communication, such as sarcasm, shares certain characteristics with deception, as in both the communicated content is at odds with reality. However, in contrast to deception, sarcasm lacks the deceptive intent; listener is meant to realize the true meaning of what is communicated. Importantly, to distinguish sarcasm from deception, one needs to identify the meaning based on contextual cues. Similarly, in the current study, participants interacted with another person and based on prosody cues obtained from the interrogator, had to monitor whether their denial of an action they did remember performing (e.g., stealing a pair of ear phones) could be successful. Their belief regarding whether the lie-detector is active was thus directly relevant to this process of inference. The right parahippocampal gyrus may therefore perform a similar role, mediating social interaction, and its underlying intent, in both the contextualized production of deception of the present study and in processing sarcasm (Rankin et al., 2009).

Emotionally charged experiences involve the hippocampus, parahippocampal gyrus, and amygdala in the process of encoding and consolidating these events into memories (Richter-Levin and Akirav, 2001). The hippocampus is known to play a crucial role in associative learning, as well as encoding and representing the value of reward (e.g., Richter-Levin and Akirav, 2001; Smith et al., 2006; Wimmer and Shohamy, 2012). Recently, Wimmer and Shohamy (2012) offered novel neural evidence indicating that the hippocampus may play an important role in value-based

decision-making. They showed that the hippocampus not only encodes reward value but also spreads it across items that were not previously considered rewarding. In light of the present findings, we propose that the neural connectivity between the hippocampus/parahippocampal gyrus and amygdala (Phelps, 2004; Smith et al., 2006) may facilitate a similar role for the hippocampus/parahippocampal gyrus in context-dependent social interactions, where social value must be flexibly assigned.

Interestingly, although activity in the amygdala was significantly modulated by response (deceptive vs. truthful), this modulation did not interact with belief about the status of the lie-detector. Our original hypothesis that the amygdala would be a prime candidate for belief-related modulation was therefore not borne out. Importantly, previous studies reporting deception-related amygdala activation (Abe et al., 2007; Baumgartner et al., 2009) did not have the immediate confrontation element that was present in the interrogation scenario of the current study. The absence of a significant interaction in the amygdala could thus be due either to belief modulating other functions than the emotional processes associated with amygdala activity, or to a ceiling effect—the interrogation context may have been sufficient to induce differential deception-related activity regardless of belief about lie-detectability.

REFERENCES

- Abe, N., Suzuki, M., Mori, E., Itoh, M., and Fujii, T. (2007). Deceiving others: distinct neural responses of the prefrontal cortex and amygdala in simple fabrication and deception with social interactions. *J. Cogn. Neurosci.* 19, 287–295.
- Aron, A. R., Robbins, T. W., and Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends Cogn. Sci.* 8, 170–177.
- Baumgartner, T., Fischbacher, U., Feierabend, A., Lutz, K., and Fehr, E. (2009). The neural circuitry of a broken promise. *Neuron* 64, 756–770.
- Bell, B. G., Kircher, J. C., and Bernhardt, P. C. (2008). New measures improve the accuracy of the directed-lie test when detecting deception using a mock crime. *Physiol. Behav.* 94, 331–340.
- Buckner, R. L. (2010). The role of the hippocampus in prediction and imagination. *Annu. Rev. Psychol.* 61, 27–48.
- Burgess, N., Maguire, E. A., and O'Keefe, J. (2002). The human hippocampus and spatial and episodic memory. *Neuron* 35, 625–641.
- Carmel, D., Dayan, Y., Naveh, A., Raveh, O., and Ben-Shakhar, G. (2003). Estimating the validity of the guilty knowledge test from simulated experiments: the external validity of mock crime studies. *J. Exp. Psychol. Appl.* 9, 261–269.
- Carr, L., Iacoboni, M., Dubeau, M. C., Mazziotta, J. C., and Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. *Proc. Natl. Acad. Sci. U.S.A.* 100, 5497–5502.
- Christopoulos, G. I., Tobler, P. N., Bossaerts, P., Dolan, R. J., and Schultz, W. (2009). Neural correlates of value, risk, and risk aversion contributing to decision making under risk. *J. Neurosci.* 29, 6469–6472.
- De Martino, B., Kumaran, D., Seymour, B., and Dolan, R. J. (2006). Frames, biases, and rational decision-making in the human brain. *Science* 313, 684–687.
- Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31, 968–980.
- Dolan, R. J. (2007). The human amygdala and orbital prefrontal cortex in behavioural regulation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362, 787–799.
- Ekman, P. (1992). *Telling Lies: Clues to Deceit in the Marketplace, Politics and Marriage*. New York, NY: W W Norton.
- Ekman, P., and Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry* 32, 88–106.
- Friston, K. J., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C., Frackowiak, R. S., et al. (1995). Analysis of fMRI time-series revisited. *Neuroimage* 2, 45–53.
- Frith, U., and Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 459–473.
- Gamer, M., Bauermann, T., Stoeter, P., and Vossel, G. (2007). Covariations among fMRI, skin conductance, and behavioral data during processing of concealed information. *Hum. Brain Mapp.* 28, 1287–1301.
- Gamer, M., Klimecki, O., Bauermann, T., Stoeter, P., and Vossel, G. (2012). fMRI-activation patterns in the detection of concealed information rely on memory-related effects. *Soc. Cogn. Affect. Neurosci.* 7, 506–515.
- Gamer, M., Kosiol, D., and Vossel, G. (2010). Strength of memory encoding affects physiological responses in the Guilty Actions Test. *Biol. Psychol.* 83, 101–107.
- Ganis, G., Kosslyn, S. M., Stose, S., Thompson, W. L., and Yurgelun-Todd, D. A. (2003). Neural correlates of different types of deception: an fMRI investigation. *Cereb. Cortex* 13, 830–836.
- Greely, H. T., and Illes, J. (2007). Neuroscience-based lie detection: the urgent need for regulation. *Am. J. Law Med.* 33, 377–431.
- Greene, J. D., and Paxton, J. M. (2009). Patterns of neural activity associated with honest and dishonest moral decisions. *Proc. Natl. Acad. Sci. U.S.A.* 106, 12506–12511.
- Grèzes, J., Berthoz, S., and Passingham, R. E. (2006). Amygdala activation when one is the target of deceit: did he lie to you or to someone else? *Neuroimage* 30, 601–608.
- Grèzes, J., Frith, C. D., and Passingham, R. E. (2004). Brain mechanisms for inferring deceit in the actions of others. *J. Neurosci.* 24, 5500–5505.
- Heekeen, H. R., Wartenburger, I., Schmidt, H., Schwintowski, H. P., and Villringer, A. (2003). An fMRI study of simple ethical decision-making. *Neuroreport* 14, 1215–1219.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* 6, 65–70.
- Holmes, A. P., and Friston, K. J. (1998). Generalisability, random effects and population inference. *Neuroimage* 7, 754.
- Kozel, F. A., Johnson, K. A., Grenesko, E. L., Laken, S. J., Kose, S., Lu, X., et al. (2009). Functional MRI detection of deception after committing a mock sabotage crime. *J. Forensic Sci.* 54, 220–231.
- Kozel, F. A., Johnson, K. A., Mu, Q., Grenesko, E. L., Laken, S. J., and George, M. S. (2005). Detecting deception using functional magnetic resonance imaging. *Biol. Psychiatry* 58, 605–613.
- Langleben, D. D., Loughead, J. W., Bilker, W. B., Ruparel, K., Childress,

CONCLUSIONS

Overall, our findings suggest that belief in lie-detection efficacy modulates a subset of the processes involved in producing deception. Cognitive processes involving reasoning and theory of mind, mediated by the IFG and PCC, as well as emotional processes mediated by the amygdala, are involved in the production of deception—but the absence of modulation by belief in these regions suggests that the processes they mediate are functionally separate from those involving belief. However, belief about the detectability of lies does modulate activity in the temporal pole and hippocampus/parahippocampal gyrus, suggesting that the social context and memory-related processing known to be mediated by these regions are the aspects of deception that are affected by belief. We, therefore, conclude that belief in the efficacy of a lie-detection device matters, emphasizing the importance of such beliefs in both basic research and applied (forensic) settings.

ACKNOWLEDGMENTS

This work was supported by University of Aarhus, The Danish Research Council for Culture and Communication, The Danish National Research Foundation's grant to Center of Functionally Integrative Neuroscience and the Wellcome Trust.

- A. R., Busch, S. I., et al. (2005). Telling truth from lie in individual subjects with fast event-related fMRI. *Hum. Brain Mapp.* 26, 262–272.
- Mobbs, D., Weiskopf, N., Lau, H. C., Eric Featherstone, E., Dolan, R. J., and Frith, C. D. (2006). The kuleshov effect: the influence of contextual framing on emotional attributions. *Soc. Cogn. Affect. Neurosci.* 1, 95–106.
- Mohamed, F. B., Faro, S. H., Gordon, N. J., Platek, S. M., Ahmad, H., and Williams, J. M. (2006). Brain mapping of deception and truth telling about an ecologically valid situation: functional MR imaging and polygraph investigation—initial experience. *Radiology* 238, 679–688.
- Moll, J., de Oliveira-Souza, R., Eslinger, P. J., Bramati, I. E., Mourao-Miranda, J., Andreiuolo, P. A., et al. (2002). The neural correlates of moral sensitivity: a functional magnetic resonance imaging investigation of basic and moral emotions. *J. Neurosci.* 22, 2730–2736.
- Nuñez, J. M., Casey, B. J., Egner, T., Hare, T., and Hirsch, J. (2005). Intentional false responding shares neural substrates with response conflict and cognitive control. *Neuroimage* 25, 267–277.
- Olson, I. R., Plotzker, A., and Ezzyat, Y. (2007). The Enigmatic temporal pole: a review of findings on social and emotional processing. *Brain* 130, 1718–1731.
- Petrovic, P., Dietrich, T., Fransson, P., Andersson, J., Carlsson, K., and Ingvar, M. (2005). Placebo in emotional processing—induced expectations on anxiety relief activate a generalized modulatory network. *Neuron* 46, 957–969.
- Phelps, E. A. (2004). Human emotion and memory: interactions of the amygdala and hippocampal complex. *Curr. Opin. Neurobiol.* 14, 198–202.
- Pollina, D. A., Dollins, A. B., Senter, S. M., Krapohl, D. J., and Ryan, A. H. (2004). Comparison of polygraph data obtained from individuals involved in mock crimes and actual criminal investigations. *J. Appl. Psychol.* 89, 1099–1105.
- Rankin, K. P., Salazar, A., Gorno-Tempini, M. L., Sollberger, M., Wilson, S. M., Pavlic, D., et al. (2009). Detecting sarcasm from paralinguistic cues: anatomic and cognitive correlates in neurodegenerative disease. *Neuroimage* 47, 2005–2015.
- Rice, W. R. (1989). Analyzing tables of statistical tests. *Evolution* 6, 223–225.
- Richter-Levin, G., and Akirav, I. (2001). Amygdala-hippocampus dynamic interaction in relation to memory. *Mol. Neurobiol.* 22, 11–20.
- Seymour, T. L., and Fraynt, B. R. (2009). Time and encoding effects in the concealed knowledge test. *Appl. Psychophysiol. Biofeedback* 34, 177–187.
- Sip, K. E., Lynge, M., Wallentin, M., McGregor, W. B., Frith, C. D., and Roepstorff, A. (2010). The production and detection of deception in an interactive game. *Neuropsychologia* 48, 3619–3626.
- Sip, K. E., Roepstorff, A., McGregor, W. B., and Frith, C. D. (2008a). Detecting deception: the scope and limits. *Trends Cogn. Sci.* 12, 48–53.
- Sip, K. E., Roepstorff, A., McGregor, W. B., and Frith, C. D. (2008b). Response to Haynes: there's more to deception than brain activity. *Trends Cogn. Sci.* 12, 126–127.
- Sip, K. E., Skewes, J. C., Marchant, J. L., McGregor, W. B., Roepstorff, A., and Frith, C. D. (2012). What if I get busted? Deception, choice, and decision-making in social interaction. *Front. Neurosci.* 6:58. doi: 10.3389/fnins.2012.00058
- Smith, A. P., Stephan, K. E., Rugg, M. D., and Dolan, R. J. (2006). Task and content modulate amygdala-hippocampal connectivity in emotional retrieval. *Neuron* 49, 631–638.
- Spence, S. A., Farrow, T. F. D., Herford, A. E., Wilkinson, I. D., Zheng, Y., and Woodruff, P. W. R. (2001). Behavioural and functional anatomical correlates of deception in humans. *Neuroreport* 12, 2849–2853.
- Spence, S. A., Taylor-Hughes, C., Farrow, T. F. D., and Wilkinson, I. D. (2008). Speaking of secrets and lies: the contribution of ventrolateral prefrontal cortex to vocal deception. *Neuroimage* 40, 1411–1418.
- Völlm, B. A., Taylor, A. N. W., Richardson, P., Rhiannon, C., Stirling, J., McKie, S., et al. (2006). Neuronal correlates of theory of mind and empathy: a functional magnetic resonance imaging study in a nonverbal task. *Neuroimage* 29, 90–98.
- Vrij, A. (1993). Credibility judgments of detectives: the impact of non-verbal behavior, social skills, and physical characteristics on impression formation. *J. Soc. Psychol.* 133, 601–610.
- Vrij, A. (2000). *Detection Lies and Deceit. The Psychology of Lying and the Implications for Professional Practice*. Chichester: John Wiley & Sons, Ltd.
- Vrij, A. (2004). Why professionals fail to catch liars and how they can improve. *Leg. Criminol. Psychol.* 9, 159–181.
- Vrij, A. (2008). Nonverbal dominance versus verbal accuracy in lie detection. *Crim. Just. Behav.* 35, 1323–1336.
- Vrij, A., Edward, K., and Bull, R. (2001). People's insight into their own behaviour and speech content while lying. *Br. J. Psychol.* 92, (Pt 2), 373–389.
- Vrij, A., Mann, S., Kristen, S., and Fisher, R. P. (2007). Cues to deception and ability to detect lies as a function of police interview styles. *Law Hum. Behav.* 31, 499–518.
- Wager, T. D., Lindquist, M., and Kaplan, L. (2007). Meta-analysis of functional neuroimaging data: current and future directions. *Soc. Cogn. Affect. Neurosci.* 2, 150–158.
- Wimmer, G. E., and Shohamy, D. (2012). Preference by association: how memory mechanisms in the hippocampus bias decisions. *Science* 338, 270–273.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 28 August 2012; accepted: 14 January 2013; published online: 04 February 2013.

Citation: Sip KE, Carmel D, Marchant JL, Li J, Petrovic P, Roepstorff A, McGregor WB and Frith CD (2013) When Pinocchio's nose does not grow: belief regarding lie-detectability modulates production of deception. Front. Hum. Neurosci. 7:16. doi: 10.3389/fnhum.2013.00016

Copyright © 2013 Sip, Carmel, Marchant, Li, Petrovic, Roepstorff, McGregor and Frith. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.



Does the inferior frontal sulcus play a functional role in deception? A neuronavigated theta-burst transcranial magnetic stimulation study

Bruno Verschueren^{1,2,3*†}, Teresa Schuhmann^{4,5†} and Alexander T. Sack^{4,5†}

¹ Department of Clinical Psychology, University of Amsterdam, Amsterdam, Netherlands

² Experimental-Clinical and Health Psychology, Ghent University, Ghent, Belgium

³ Clinical Psychology Science, Maastricht University, Maastricht, Netherlands

⁴ Cognitive Neuroscience, Maastricht University, Maastricht, Netherlands

⁵ Maastricht Brain Imaging Center, Maastricht, Netherlands

Edited by:

Matthias Gamer, University Medical Center Hamburg-Eppendorf, Germany

Reviewed by:

Alberto Priori, Università di Milano, Italy

Ayahito Ito, Tohoku University, Japan

***Correspondence:**

Bruno Verschueren, Department of Clinical Psychology, University of Amsterdam, Weesperplein 4, 1018 XA Amsterdam, Netherlands.
e-mail: b.j.verschueren@uva.nl

[†]These authors equally contributed to this work.

By definition, lying involves withholding the truth. Response inhibition may therefore be the cognitive function at the heart of deception. Neuroimaging research has shown that the same brain region that is activated during response inhibition tasks, namely the inferior frontal region, is also activated during deception paradigms. This led to the hypothesis that the inferior frontal region is the neural substrate critically involved in withholding the truth. In the present study, we critically examine the functional necessity of the inferior frontal region in withholding the truth during deception. We experimentally manipulated the neural activity level in right inferior frontal sulcus (IFS) by means of neuronavigated continuous theta-burst stimulation (cTBS). Individual structural magnetic resonance brain images (MRI) were used to allow precise stimulation in each participant. Twenty-six participants answered autobiographical questions truthfully or deceptively before and after sham and real cTBS. Deception was reliably associated with more errors, longer and more variable response times than truth telling. Despite the potential role of IFS in deception as suggested by neuroimaging data, the cTBS-induced disruption of right IFS did not affect response times or error rates, when compared to sham stimulation. The present findings do not support the hypothesis that the right IFS is critically involved in deception.

Keywords: **deception, response inhibition, transcranial magnetic stimulation, theta-burst, inferior frontal sulcus**

INTRODUCTION

In recent years, deception researchers have focused upon the cognitive processes involved in deception (Vrij, 2008). Formulated broadly, the cognitive perspective on deception holds that deception is cognitively more demanding than truth telling. Deception often involves one or more of the following mental operations: the decision to lie, withholding the truth, fabrication of the lie, monitoring whether the receiver believes the lie and, if necessary, adjusting the fabricated story, and keeping the lying consistent. These operations make lying a cognitively demanding task. Evidence supports the cognitive perspective on deception. For example, lying participants were judged by observers to think harder than truthful participants, and participants subjectively reported more cognitive load when lying compared to truth telling (Vrij et al., 2006). Furthermore, compared to truth telling, lying is associated with more errors, increased and more variable response times (Spence et al., 2001; Johnson et al., 2005; Verschueren et al., 2011). Recently, several studies used brain imaging techniques such as fMRI (Spence et al., 2001; Langleben et al., 2002; Ganis et al., 2003; Kozel et al., 2005; Phan et al., 2005; Monteleone et al., 2006; Abe et al., 2008), PET (Abe et al., 2006), and fNIRS (Tian et al., 2009) to identify which brain regions are associated with deception. Common across these studies is

the greater activation in the prefrontal cortex during lying compared to truth telling (Christ et al., 2009), thereby supporting the idea that deception requires greater executive control than truth telling.

Since deception by definition involves withholding the truth, response inhibition may be crucial for deception. Indeed, liars may or may not overtly express a deceitful answer, but they definitely need to refrain from telling the truth. Response inhibition can be defined as the cognitive function that allows one to intentionally inhibit a dominant, automatic or prepotent response (Miyake et al., 2000). The truth, then, is regarded as the dominant response that needs to be actively inhibited in order to lie (Spence et al., 2008a). Noteworthy from the perspective of the association between response inhibition and deception, is the observation that the same brain regions are critically involved in response inhibition and in deception. Examining the neural correlates of response inhibition, imaging studies have examined brain activity during tasks that require active suppression of a dominant response such as the Go/No-Go task and the Stop-signal task. The Go/No-Go task requires a speeded response to frequently presented Go trials (e.g., the letter Q), but inhibition of responding to the rarely presented No-Go trials (e.g., the letter O). In the Stop-signal, responding to the go task (e.g., press left for circle

and right for square) has to be inhibited when an auditory signal is presented. A particular region in the prefrontal cortex, the right inferior frontal region, is consistently and most strongly activated during such tasks (Garavan et al., 1999; Konishi et al., 1999; Aron et al., 2004; Brass et al., 2005). In 18 patients with right frontal lobe damage, it was found that the greater the damage to the inferior frontal gyrus (IFG), the worse response inhibition performance in the Stop-signal task (Aron et al., 2003). Further support for the functional necessity of the IFG in response inhibition comes from recent work using repetitive transcranial magnetic stimulation (rTMS). rTMS is a non-invasive brain stimulation technique that allows to induce a transient and reversible “virtual lesion” in healthy conscious volunteers. rTMS to the IFG, but not to mid frontal or parietal regions, impaired response inhibition capacity in healthy volunteers (Chambers et al., 2006). As the inferior frontal region is also consistently activated in deception paradigms (Spence et al., 2001; Kozel et al., 2005; Phan et al., 2005; Gamer et al., 2007; Christ et al., 2009), it may be this region that is crucial for inhibiting the truth during deception (Spence et al., 2004).

In sum, brain imaging studies suggest that the inferior frontal region may exert a functional role in withholding the truth during deception. However, since imaging studies are in essence correlation studies, they do not allow conclusions with regard to the functional necessity of brain regions. In order to investigate the functional necessity of this region for deception, one would need to experimentally manipulate its activity level and investigate the impact on deception (Sack, 2006; Luber et al., 2009). Here, we present the first study that used rTMS to unravel the functional relevance of the inferior frontal region for deception. Following recent imaging data (Brass et al., 2005), we focused upon the right inferior frontal sulcus (IFS). We collected structural images of the brain using magnetic resonance imaging (MRI). These individual anatomical brain images were used as a basis for a frameless stereotaxic TMS neuronavigation system, allowing us to precisely map and target the IFS with TMS in each individual participant. Furthermore, we used an innovative TMS protocol, continuous theta-burst rTMS (cTBS), that requires a much shorter stimulation time yet leads to more robust inhibitory after-effects than conventional rTMS protocols (Huang et al., 2005; Thut and Pascual-Leone, 2010). Disruption of the right IFS using cTBS impairs stopping performance in a stop-signal task (Verbruggen et al., 2010). This MRI-guided cTBS neuronavigation approach was used here to transiently disrupt neural processing in the right IFG to examine whether it is causally related to deception.

MATERIALS AND METHODS

PARTICIPANTS

Thirty-one participants were paid €15/h for participation. All participants had normal or corrected-to-normal vision and had no history of neurological or psychiatric disorders. They received medical approval for participation and gave their written informed consent after being introduced to the procedure. The study was approved by the local Medical Ethical Commission, and written informed consent was obtained from all participants.

Due to experimenter error, data from three participants were lost. Furthermore, the data from one participant for whom rTMS

was stopped after a startle response were excluded. Finally, data from one participant were excluded because of an excessive error percentage (18%; >2.5 SDs from the M).

The final sample consisted of 26 participants (15 women, 11 men; $M_{age} = 26.11$ years, $SD = 7.53$; 96% right-handed). Participants were tested in their preferred language (19 Dutch, 6 English, and 1 French).

PROCEDURE

Participants were tested in three separate sessions. In session 1, we obtained anatomical brain measurements of all participants using MRI. In session 2, participants were informed about the experiment and rTMS, filled in the autobiographical questionnaire, and performed the deception test a first time. Next, the active motor threshold (AMT) for each participant was determined. We then used frameless stereotaxy for MRI-guided TMS neuronavigation to the previously defined target region, and applied either a cTBS protocol that has shown to inhibit the stimulated areas for up to 1 h following the TBS itself (Huang et al., 2005; Thut and Pascual-Leone, 2010), or sham TBS using a placebo TMS coil. The second deception test followed immediately after the rTMS/sham stimulation. The procedure was identical for session 3, except that stimulation type differed and that motor threshold was not determined again. Real rTMS stimulation was on day 1 for 15 participants and on day 2 for 11 participants.

This study design and methodological approach enabled us to first define the target brain area based on the individual anatomical data and to subsequently neuronavigate the TMS coil to the anatomically defined stimulation site in each participant. The MRI-guided TMS neuronavigation was monitored online throughout the whole stimulation time, allowing for a precise determination of the actual stimulation site also during stimulation.

Deception paradigm: the Sheffield lie test

The Sheffield lie test is a “differentiation of deception” paradigm (Furedy et al., 1988) that was developed by Spence and colleagues from Sheffield University (Spence et al., 2001, 2008a,b), and has been successfully replicated by our group (Verschueren et al., 2011) and others (Fullam et al., 2009). Participants first completed a questionnaire that listed 72 specific behaviors (e.g., “Bought a newspaper”), and were asked to indicate whether or not they had performed those actions that day. Half of these questions came from the study by Spence et al. (2001) the remaining half were developed for the present study. Trials in the Sheffield lie test consisted of statements from the autobiographical questionnaire presented for 5 s. Participants answer the statements with a right-hand Yes or No response. The Yes and No reminder labels remained on the screen throughout the test. Crucially, their color varied after every six trials. One color (e.g., yellow) indicated the participant to answer truthfully, whereas the other (e.g., blue) was the signal to lie, with colors counterbalanced across participants. Meaning of the colors was assigned in the instructions, and checked in a practice phase with statements for which ground truth was known (e.g., “Are you in France?”). The test consisted of 72 trials, with each of 36 statements appearing once with blue and once with yellow reminder labels. After a 5 min break, participants took the deception test again, this time without practice at

the beginning. One set of 36 questions was used in the first test, and one set in the second test, with sets counterbalanced across participants. These sets were tested beforehand to result in a deception effect of similar magnitude. Statements were presented by a PC using Inquisit 3.0 software (Inquisit, 2009).

MRI measurements

A high-resolution anatomical image was obtained from each participant in a 3-T magnetic resonance scanner (Siemens Allegra MR Tomograph; Siemens AG, Erlangen, Germany) at the Faculty

of Psychology and Neuroscience, Maastricht University, The Netherlands. The data set was acquired with the help of a T1-weighted structural scan with an isotropic resolution of 1 mm using a modified driven equilibrium Fourier transform (MDEFT) sequence with optimized contrast for GM and WM and imaging parameters.

Cortical-surface reconstruction

Data were analyzed using the BrainVoyager QX 2.0 software package (BrainInnovation, Maastricht, The Netherlands). The high-resolution anatomical recordings were used for surface reconstruction of the right hemisphere of each participant (Kriegeskorte and Goebel, 2001). The surface reconstruction was performed in order to recover the exact spatial structure of the cortical sheet and to improve the visualization of the anatomical gyration. The white-gray-matter boundary was segmented with a region growing method preceded by inhomogeneity correction of signal intensity across space. The borders of the two resulting segmented subvolumes were tessellated to produce a surface reconstruction of the right hemisphere.

TMS apparatus and stimulation parameters

Biphasic TMS pulses were applied using the MagProX100 stimulator (Medtronic Functional Diagnostics A/S, Skovlunde, Denmark) and a figure-of-8 coil (MC-B70, inner radius 10 mm, and outer radius 50 mm) for real stimulation. The maximum output of this coil and stimulator combination is approximately 1.9 Tesla and 150 A/ μ s. A specific figure-of-8 placebo coil (MC-P-B70) was also employed in order to reproduce the same acoustic stimulation as the active coil while not inducing the magnetic field (sham stimulation). The coil was manually held tangentially to the skull with the coil handle oriented perpendicular to the posterior part of the IFS using the online visualization function of the BrainVoyager TMS Neuronavigator. Following Huang et al. (2005), continuous theta-burst TMS was applied at 80% AMT. A detailed description of this rTMS paradigm can be found in Huang et al. (2005). In brief, in TBS protocols, short bursts of 50 Hz rTMS are repeated at a rate in the theta range (5 Hz) as a continuous (cTBS) or intermittent (iTBS) train (Huang et al., 2005; Di Lazzaro et al., 2008). Depending on the train intervals, TBS can either have longer-lasting facilitatory or inhibitory after effects. The after effects of TBS were

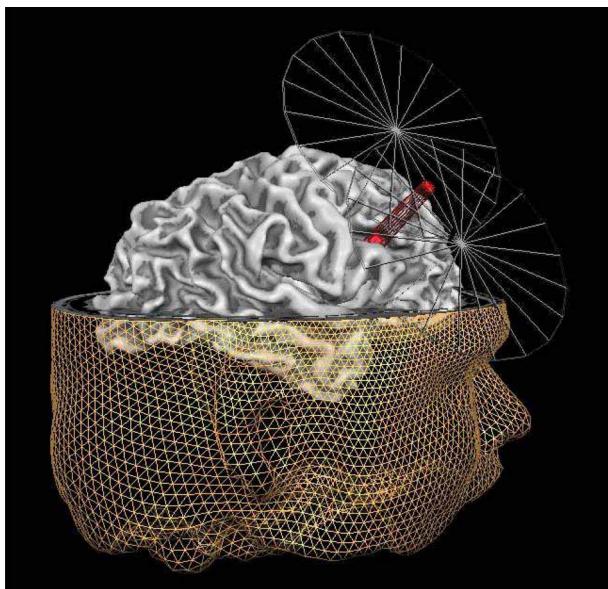


FIGURE 1 | Graphic representation of the MRI neuronavigated C-TBS at the right IFS. The inferior frontal sulcus (IFS) target point (red dot under the beam of the coil) for TMS, shown on the reconstruction of the right hemisphere of one exemplary participant. The target point is placed on the posterior part of the right IFS, in particular the area just anterior to the section of the precentral sulcus and the inferior frontal sulcus. In addition to the reconstruction of the right hemisphere of this participant, also the reconstruction of the head is displayed together with a simplified visualization of the coil. The tip of the red beam from the TMS figure-8 coil indicates the site of the maximal stimulation.

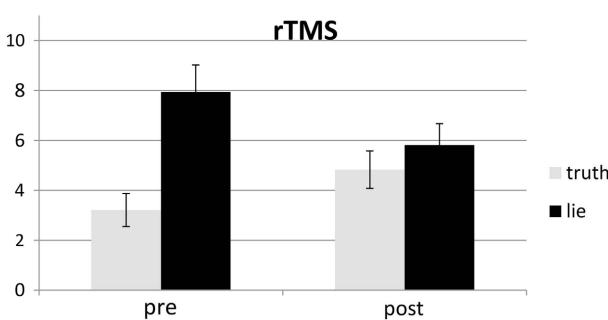


FIGURE 2 | Mean error (in %; \pm one SE) for lie and truth trials, pre and post real rTMS.

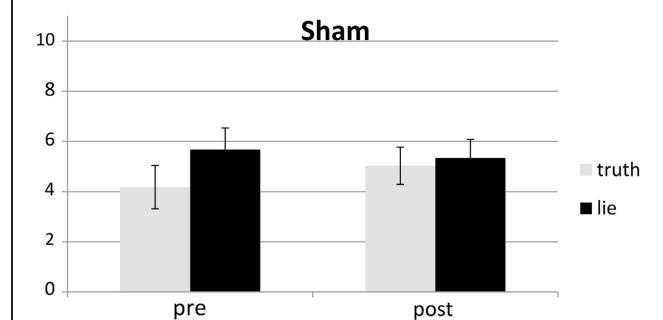


FIGURE 3 | Mean error (in %; \pm one SE) for lie and truth trials, pre and post sham.

found to be significantly longer-lasting compared to conventional rTMS (Huang et al., 2005) with shorter stimulation time and lower stimulation intensity needed. These factors could allow more comfortable stimulation conditions, especially when TBS is used as a therapeutical intervention over a long period of time (Cardenas-Morales et al., 2010). It has been suggested that cTBS decreases the effectiveness of synaptic connections that are recruited in circuits involved in both short interval intracortical inhibition (SICI) and intracortical facilitation (ICF) (Huang et al.). Some side effects were noted with this stimulation, most notably muscle twitches at the eye, cheek and mouth.

TMS localization

IFS corresponds to area 44 in Brodmann's cytoarchitectonic map (Brodmann, 1909). Based upon anatomical landmarks, we targeted the posterior part of the right IFS. Specifically, we targeted the area just anterior to the section of the precentral sulcus and the IFS. The stimulation site was localized using frameless stereotaxy (Brain Voyager TMS neuronavigation; Sack et al., 2006) for both real and sham stimulation. Using such a TMS neuronavigation system enabled us to account for inter-individual differences in anatomical brain structures while stimulating (see **Figure 1**).

TMS procedure

Individual AMTs were determined as the intensity at which the stimulation of the left motor cortex with single-pulse TMS resulted reliably in a visible movement of the first dorsal interosseous (FDI) muscle. The AMT of the participants ranged from 21 to 45% of maximum stimulator output [$M = 30.27\%$ ($47\text{ A}/\mu\text{S}$), $SD = 5.24$]. The mean stimulation intensity was set at 80% of the AMT and therefore resulted in 24.19% ($38\text{ A}/\mu\text{S}$) of maximum stimulator output (range 17–36%, $SD = 4.97$). Throughout the stimulation time, participants were wearing earplugs to protect their ears from the clicking sound and to minimize the interference of sounds during the task.

RESULTS

Separate $2 \times 2 \times 2$ ANOVAs with stimulation (rTMS vs. sham), session (pre vs. post), and deception (lie vs. truth) as the within-subjects factors were conducted on error percentage (%), and on mean (RTs) and variability (SD RTs) of correct response times.

ERRORS

Responses that did not match with the autobiographical questionnaire were considered behavioral errors. The only reliable effect was a main effect of deception, $F_{(1, 25)} = 10.22$, $p < 0.01$, with lying resulting in more errors than truth telling, see **Figures 2, 3**. Two other effects just failed short of reaching significance: Session \times Deception, $F_{(1, 25)} = 4.15$, $p = 0.05$, indicating that the lie vs. truth difference was somewhat greater at baseline than at test; and Stimulation \times Deception, $F_{(1, 25)} = 3.06$, $p = 0.09$, indicating that the lie vs. truth difference was somewhat greater in the rTMS session than in the sham session. Other F 's < 1 .

RTs

Behavioral errors were excluded from the RT analyses, as where RTs that deviated more than 2.5 SDs from the individual conditional mean (Ratcliff, 1993). There was only a main effect of deception, with participants being slower when lying than when telling the truth, $F_{(1, 25)} = 43.19$, $p < 0.001$, see **Figures 4, 5**. Other F 's < 1.5 .

SD RTs

SD RTs of the RTs included in the RT analyses were analyzed. There was only a main effect of deception, with participants being slower when lying than when telling the truth, $F_{(1, 25)} = 13.31$, $p < 0.01$, see **Figures 6, 7**. Other F 's < 2.2 .

DISCUSSION

Since deception by cognitive definition involves withholding the truth, response inhibition may be the cognitive function at the heart of deception. The behavioral data in the present study indeed showed that lying comes with a "cost," as lying was reliably associated with more errors and greater and more variable response times compared to truth telling, thereby replicating previous findings obtained with the Sheffield lie test (Spence et al., 2001, 2008a; Fullam et al., 2009; Verschuere et al., 2011) as well as with other deception paradigms (e.g., Sartori et al., 2008; Verschuere et al., 2009). A prominent cognitive neurobiological account of deception holds that this cost can be related to the active inhibition of the dominant truth response (Spence et al., 2008a), and that this response inhibition of the truth is regulated

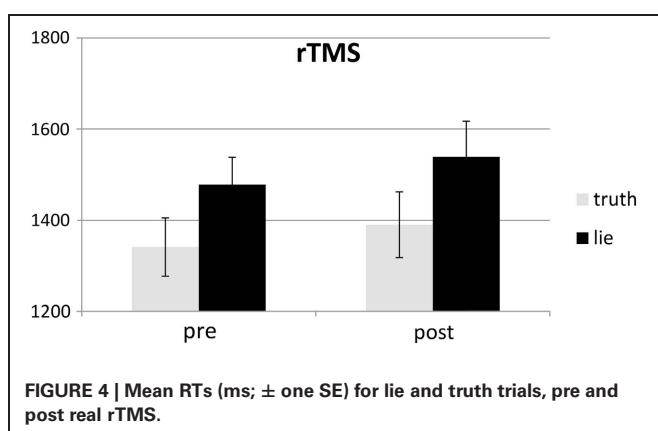


FIGURE 4 | Mean RTs (ms; \pm one SE) for lie and truth trials, pre and post real rTMS.

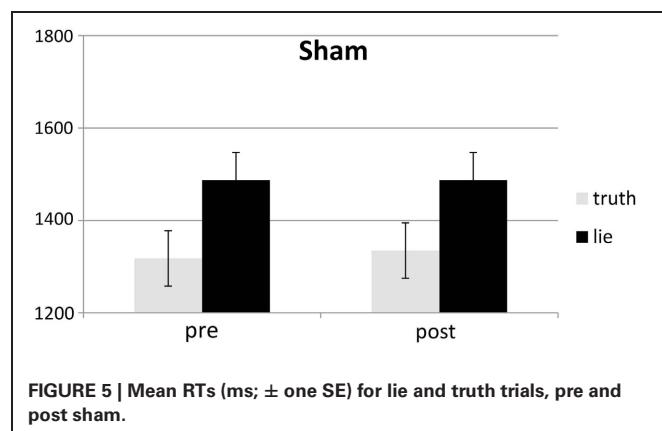


FIGURE 5 | Mean RTs (ms; \pm one SE) for lie and truth trials, pre and post sham.

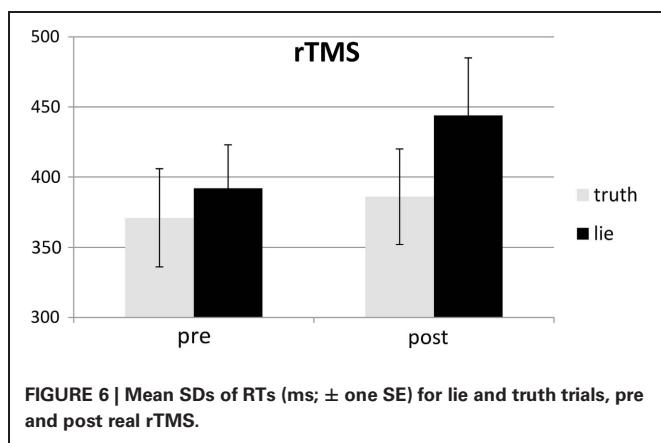


FIGURE 6 | Mean SDs of RTs (ms; \pm one SE) for lie and truth trials, pre and post real rTMS.

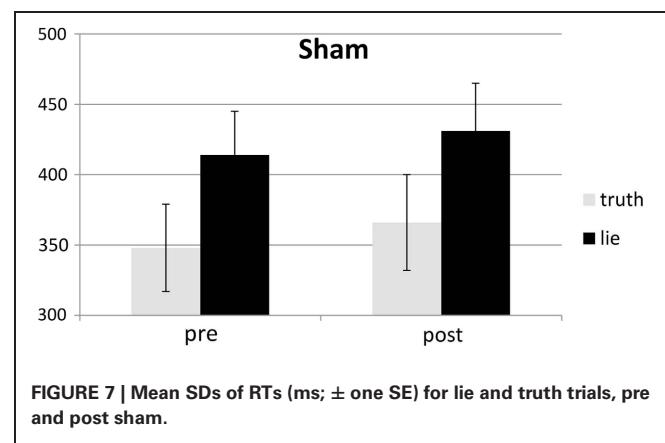


FIGURE 7 | Mean SDs of RTs (ms; \pm one SE) for lie and truth trials, pre and post sham.

mainly in right inferior frontal cortex (Spence et al., 2001, 2004; Kozel et al., 2005; Phan et al., 2005; Gamer et al., 2007; Christ et al., 2009). Being in essence correlation studies, imaging studies do not allow conclusions with regard to the functional necessity of brain regions. Here, we used rTMS to unravel the functional relevance of the right inferior frontal cortex for deception, expecting that a cTBS-induced disruption of right IFS would affect behavioral responding on the lying trials. However, real cTBS over right IFS had no effect on deception as compared to sham stimulation in the current study.

Our present findings failed to refute the null hypothesis, leaving us with the question whether the data can be meaningfully interpreted or not (De Graaf and Sack, 2011). To the extent that methodological aspects can explain our negative findings, interpretation is hazardous. Under certain methodological conditions, however, negative TMS findings provide a meaningful answer to the question that cannot be answered by imaging techniques: Is the specific brain region functionally relevant for the task or not? After all, TMS is an entirely different method than brain imaging, going beyond the correlation approach, and allowing to examine whether a region identified in imaging work is functionally relevant for the task or may be a non-functional by-product. Three important aspects need consideration to make meaningful interpretation of negative TMS findings (De Graaf and Sack, 2011): the localization argument (perhaps the coil was not positioned properly and the targeted brain region X was therefore not stimulated), the neural efficacy argument (did the expected neural effects occur?), and the power argument (maybe a non-significant TMS effect requires more participants). The *power argument* is not easily refuted, but is unlikely to explain our negative findings given the lack of statistical trends, and the use of a within-subjects design that seems sufficiently powered ($n = 26$) compared to previous research (Huang et al., 2005; Chambers et al., 2006; Verbruggen et al., 2010). With regard to the *localization argument*, the current study used individual MRI data to neuronavigate the coil to a specific individually-defined target point within IFS (see the “Materials and Methods” section). While we cannot rule out that individual fMRI data may have resulted in a slightly different TMS target site and potentially different results, we can conclude that stimulating the anatomical region within IFS shown here (Figure 1) does not affect

deception. With regard to the *neural efficiency argument*, the question can be raised whether the stimulation produced the intended change in cortical excitability. Unlike for the motor system, no direct and easily measurable assessment for the local cortical excitability level of right IFS is available, unless cTBS is directly combined with EEG or fMRI during stimulation. It has been shown that there is a considerable inter-individual variance in the cortical after effects of rTMS (Maeda et al., 2000) with some participants showing an increase in cortical excitability while others showing a respective decrease in cortical excitability, even when being stimulated with the same rTMS protocol. Moreover, it has been shown that the same rTMS protocol can induce opposite neural after effects (excitatory vs. inhibitory) when applied over different cortical target sites (Paus et al., 1997). Future research will benefit from direct concurrent neurophysiologic measurements to examine the direction of the change in cortical excitability induced by the rTMS/ transcranial direct current stimulation (tDCS) intervention. Furthermore, future studies should also include other control sites, and not only make use of sham stimulation as a control, since participants might be able to detect the difference between real and sham stimulation.

Whereas we cannot easily dismiss all methodological arguments relating to power, localization, and neural efficiency our negative findings may be meaningfully interpreted given that our study was based on a clear *a priori* hypothesis directly derived from the imaging literature, and conducted using state-of-the art TMS methodology—including (1) the employment of individual structural brain imaging data to select and target the right IFS in each individual participant, (2) a paradigm that reliably elicits stronger inferior frontal activation for lying compared to truth telling (Spence et al., 2001; Christ et al., 2009; Fullam et al., 2009), (3) a reasonably powered design (within-subjects; $n = 26$), and (4) a stimulation protocol (cTBS) that has been shown to produce immediate, profound and lasting effects on cognitive functioning generally and on inhibition specifically (Huang et al., 2005; Thut and Pascual-Leone, 2010; Verbruggen et al., 2010). As such our finding that our inhibitory protocol (cTBS) over the right IFS identified by individual MRI (see target site in Figure 1) did not have behavioral effects on deception as measured within the Sheffield lie test contains much more information than a

“pure” null result and is informative for the scientific community. The present study rejoins a handful of neuromodulation studies on deception. Unfortunately, the results of these studies are mixed and inconsistent. In the present study, we failed to find an effect of cTBS to the rIFC on deception. Previous studies have used related technique: tDCS or rTMS, both of which can be used to either increase or decrease neural excitability. Priori et al. (2008) unexpectedly found that anodal (excitatory) tDCS of the DLPFC hampered lying, with no effect of cathodal (inhibitory) stimulation. Karim et al. (2010), however, failed to find an effect of anodal tDCS to the anterior PFC. Rather, they found that cathodal tDCS to the same region facilitated lying. Rather than hampering lying as observed by Priori et al. (2008), Mameli et al. (2010) found that anodal tDCS of the DLPFC facilitated lying. Finally, Karton and Bachmann (2011) found that inhibiting the left DLPFC using low frequency rTMS makes people less truthful, whereas inhibiting the right DLPFC makes them more truthful. The small sample size ($n = 8$), and the lack of a baseline assessment are noteworthy shortcomings of this latter study. Taken together, these studies point to a functional role of the DLPFC in deception, yet also underscore that its exact role

remains unclear. Interestingly, rTMS studies of deception have received great media attention, headings “Magnets, the ultimate truth serum”, “Scientists can make you lie using magnets,” and “Magnetic pulses to the brain make it impossible to lie.” Our findings together with our review of previous rTMS studies of deception show these headline are misleading. Clearly, we are far from using this technology in applied setting, because we do not know exactly whether and how neuromodulation will affect lying ability. Still, neuromodulation is a powerful and promising technique that may help to reveal the neural underpinnings of deception. We hope that the present report provides an impetus to further investigate the functional necessity of brain regions associated with deception (Christ et al., 2009) using rTMS/tDCS.

ACKNOWLEDGMENTS

This research was supported by grants from the Scientific Research Foundation (FWO), and the Netherlands Organization for Scientific Research (NWO; grant number 452-06-003 and 400-04-215). We thank our medical supervisor Cees van Leeuwen, our independent physician Martin van Boxtel, and, for their aid in data collection, Mario Senden and Sonja Cornelsen.

REFERENCES

- Abe, N., Okuda, J., Suzuki, M., Sasaki, H., Matsuda, T., Mori, E., et al. (2008). Neural Correlates of true memory, false memory, and deception. *Cereb. Cortex* 18, 2811–2819.
- Abe, N., Suzuki, M., Tsukiura, T., Mori, E., Yamaguchi, K., Itoh, M., et al. (2006). Dissociable roles of prefrontal and anterior cingulate cortices in deception. *Cereb. Cortex* 16, 192–199.
- Aron, A. R., Fletcher, P. C., Bullmore, E. T., Sahakian, B. J., and Robbins, T. W. (2003). Stop-signal inhibition disrupted by damage to right inferior frontal gyrus in humans. *Nat. Neurosci.* 6, 115–116.
- Aron, A. R., Robbins, T. W., and Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends Cogn. Sci.* 8, 170–177.
- Brass, M., Derrfuss, J., Forstmann, B., and von Cramon, D. Y. (2005). The role of the inferior frontal junction area in cognitive control. *Trends Cogn. Sci.* 9, 314–316.
- Brodmann, K. (1909). *Vergleichende Lokalisationslehre der Großhirnrinde in Ihren Prinzipien Dargestellt auf Grund des Zellenbaues*. Leipzig: Johann Ambrosius Barth.
- Cardenas-Morales, L., Nowak, D. A., Kammer, T., Wolf, R. C., and Schonfeldt-Lecuona, C. (2010). Mechanisms and applications of theta-burst rTMS on the human motor cortex. *Brain Topogr.* 22, 294–306.
- Chambers, C. D., Bellgrove, M. A., Stokes, M. G., Henderson, T. R., Garavan, H., Robertson, I. H., et al. (2006). Executive “brake failure” following deactivation of human frontal lobe. *J. Cogn. Neurosci.* 18, 444–455.
- Christ, S. E., Essen, D. C., Watson, J. M., Brubaker, L. E., and McDermott, K. B. (2009). The contributions of prefrontal cortex and executive control to deception: evidence from activation likelihood estimate meta-analyses. *Cereb. Cortex* 19, 1557–1566.
- De Graaf, T. A., and Sack, A. T. (2011). Null results in TMS: from absence of evidence to evidence of absence. *Neurosci. Biobehav. Rev.* 35, 871–877.
- Di Lazzaro, V., Pilato, F., Dileone, M., Profice, P., Oliviero, A., Mazzone, P., et al. (2008). The physiological basis of the effects of intermittent theta burst stimulation of the human motor cortex. *J. Physiol.* 586, 3871–3879.
- Fullam, R. S., McKie, S., and Dolan, M. C. (2009). Psychopathic traits and deception: functional magnetic resonance imaging study. *Br. J. Psychiatry* 194, 229–235.
- Furedy, J. J., Davis, C., and Gurevich, M. (1988). Differentiation of deception as a psychological process – a psychophysiological approach. *Psychophysiology* 25, 683–688.
- Gamer, M., Bauermann, T., Stoeter, P., and Vossel, G. (2007). Covariations Among fMRI, Skin conductance and behavioral data during processing of concealed information. *Hum. Brain Mapp.* 28, 1287–1301.
- Ganis, G., Kosslyn, S. M., Stose, S., Thompson, W. L., and Yurgelun-Todd, D. A. (2003). Neural correlates of different types of deception: an fMRI investigation. *Cereb. Cortex* 13, 830–836.
- Garavan, H., Ross, T. J., and Stein, E. A. (1999). Right hemispheric dominance of inhibitory control: an event-related functional MRI study. *Proc. Natl. Acad. Sci. U.S.A.* 96, 8301–8306.
- Huang, Y. Z., Edwards, M. J., Rounis, E., Bhatia, K. P., and Rothwell, J. C. (2005). Theta burst stimulation of the human motor cortex. *Neuron* 45, 201–206.
- Inquisit. 3.0 ed. (2009). *Millisecond Software*. p. Computer software. Seattle, WA.
- Johnson, R., Barnhardt, J., and Zhu, J. (2005). Differential effects of practice on the executive processes used for truthful and deceptive responses: an event-related brain potential study. *Cogn. Brain Res.* 24, 386–404.
- Karim, A. A., Schneider, M., Lotze, M., Veit, R., Sauseng, P., Braun, C., et al. (2010). The truth about lying: inhibition of the anterior prefrontal cortex improves deceptive behavior. *Cereb. Cortex* 20, 205–213.
- Karton, I., and Bachmann, T. (2011). Effect of prefrontal transcranial magnetic stimulation on spontaneous truth-telling. *Behav. Brain Res.* 225, 209–214.
- Konishi, S., Nakajima, K., Uchida, I., Kikyo, H., Kameyama, M., and Miyashita, Y. (1999). Common inhibitory mechanism in human inferior prefrontal cortex revealed by event-related functional MRI. *Brain* 122, 981–991.
- Kozel, F. A., Johnson, K. A., Mu, Q. W., Grenesko, E. L., Laken, S. J., and George, M. S. (2005). Detecting deception using functional magnetic resonance imaging. *Biol. Psychiatry* 58, 605–613.
- Kriegeskorte, N., and Goebel, R. (2001). An efficient algorithm for topologically correct segmentation of the cortical sheet in anatomical mr volumes. *Neuroimage* 14, 329–346.
- Langleben, D. D., Schroeder, L., Maldjian, J. A., Gur, R. C., McDonald, S., Ragland, J. D., et al. (2002). Brain activity during simulated deception: an event-related functional magnetic resonance study. *Neuroimage* 15, 727–732.
- Luber, B., Fisher, C., Appelbaum, P. S., Ploessner, M., and Lisanby, S. H. (2009). Non-invasive brain stimulation in the detection of deception: scientific challenges and ethical consequences. *Behav. Sci. Law* 27, 191–208.
- Maeda, F., Keenan, J. P., Tormos, J. M., Topka, H., and Pascual-Leone, A. (2000). Interindividual variability of the modulatory effects of repetitive transcranial magnetic stimulation on cortical excitability. *Exp. Brain Res.* 133, 425–430.
- Mameli, F., Mrakic-Sposta, S., Vergari, M., Fumagalli, M., Macis, M., Ferrucci, R., et al. (2010).

- Dorsolateral prefrontal cortex specifically processes general - but not personal - knowledge deception: multiple brain networks for lying. *Behav. Brain Res.* 211, 164–168.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., and Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: a latent variable analysis. *Cogn. Psychol.* 41, 49–100.
- Monteleone, G. T., Phan, K. L., Fitzgerald, D. A., Irick, J. S., Hawley, L. C., and Cacioppo, J. T. (2006). Can lying be diagnosed using fMRI? *Psychophysiology* 43, S68.
- Paus, T., Jech, R., Thompson, C. J., Comeau, R., Peters, T., and Evans, A. C. (1997). Transcranial magnetic stimulation during positron emission tomography: a new method for studying connectivity of the human cerebral cortex. *J. Neurosci.* 17, 3178–3184.
- Phan, K. L., Magalhaes, A., Ziemlewicz, T. J., Fitzgerald, D. A., Green, C., and Smith, W. (2005). Neural correlates of telling lies: a functional magnetic resonance imaging study at 4 tesla. *Acad. Radiol.* 12, 164–172.
- Priori, A., Mameli, F., Cogiamanian, F., Marceglia, S., Tiriticco, M., Mrakic-Sposta, S., et al. (2008). Lie-specific involvement of dorsolateral prefrontal cortex in deception. *Cereb. Cortex* 18, 451–455.
- Ratcliff, R. (1993). Methods for dealing with reaction-time outliers. *Psychol. Bull.* 114, 510–532.
- Sack, A. T. (2006). Transcranial magnetic stimulation, causal structure-function mapping and networks of functional relevance. *Curr. Opin. Neurobiol.* 16, 593–599.
- Sack, A. T., Kohler, A., Linden, D. E. J., Goebel, R., and Muckli, L. (2006). The temporal characteristics of motion processing in hMT/V5+: combining fMRI and neuronavigated TMS. *Neuroimage* 29, 1326–1335.
- Sartori, G., Agosta, S., Zogmaister, C., Ferrara, S. D., and Castiello, U. (2008). How to accurately assess autobiographical events. *Psychol. Sci.* 19, 772–780.
- Spence, S. A., Farrow, T. F. D., Herford, A. E., Wilkinson, I. D., Zheng, Y., and Woodruff, P. W. R. (2001). Behavioural and functional anatomical correlates of deception in humans. *Neuroreport* 12, 2849–2853.
- Spence, S. A., Hunter, M. D., Farrow, T. F. D., Green, R. D., Leung, D. H., Hughes, C. J., et al. (2004). A cognitive neurobiological account of deception: evidence from functional neuroimaging. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 359, 1755–1762.
- Spence, S. A., Kaylor-Hughes, C., Farrow, T. F. D., and Wilkinson, I. D. (2008a). Speaking of secrets and lies: the contribution of ventrolateral prefrontal cortex to vocal deception. *Neuroimage* 40, 1411–1418.
- Spence, S. A., Kaylor-Hughes, C. J., Brook, M. L., Lankappa, S. T., and Wilkinson, I. D. (2008b). ‘Munchausen’s syndrome by proxy’ or a ‘miscarriage of justice’: an initial application of functional neuroimaging to the question of guilt versus innocence. *Eur. Psychiatry* 23, 309–314.
- Thut, G., and Pascual-Leone, A. (2010). A review of combined TMS-EEG studies to characterize lasting effects of repetitive TMS and assess their usefulness in cognitive and clinical neuroscience. *Brain Topogr.* 22, 219–232.
- Tian, F. H., Sharma, V., Kozel, F. A., and Liu, H. L. (2009). Functional near-infrared spectroscopy to investigate hemodynamic responses to deception in the prefrontal cortex. *Brain Res.* 1303, 120–130.
- Verbruggen, F., Aron, A. R., Stevens, M. A., and Chambers, C. D. (2010). Theta burst stimulation dissociates attention and action updating in human inferior frontal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 107, 13966–13971.
- Verschuere, B., Prati, V., and De Houwer, J. (2009). Cheating the lie detector: faking in the autobiographical IAT. *Psychol. Sci.* 20, 410–413.
- Verschuere, B., Spruyt, A., Meijer, E. H., and Otgaar, H. (2011). The ease of lying. *Conscious. Cogn.* 20, 908–911.
- Vrij, A. (2008). *Detection of Lies and Deceit: Pitfalls and Opportunities*. 2nd Edn. Chichester: John Wiley and Sons Ltd.
- Vrij, A., Fisher, R., Mann, S., and Leal, S. (2006). Detecting deception by manipulating cognitive load. *Trends Cogn. Sci.* 10, 141–142.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 12 June 2012; accepted: 28 September 2012; published online: 18 October 2012.

*Citation: Verschuere B, Schuhmann T and Sack AT (2012) Does the inferior frontal sulcus play a functional role in deception? A neuronavigated theta-burst transcranial magnetic stimulation study. *Front. Hum. Neurosci.* 6:284. doi: 10.3389/fnhum.2012.00284*

Copyright © 2012 Verschuere, Schuhmann and Sack. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.



Markers of deception in Italian speech

Katelyn Spence, Gina Villar and Joanne Arciuli *

Faculty of Health Sciences, University of Sydney, Sydney, NSW, Australia

Edited by:

Wolfgang Ambach, Institute for Frontier Areas of Psychology and Mental Health, Germany

Reviewed by:

Kenny Smith, University of Edinburgh, UK

Jaume Masip, University of Salamanca, Spain

Ray Bull, University of Leicester, UK

***Correspondence:**

Joanne Arciuli, Faculty of Health Sciences, University of Sydney, PO Box 170, Lidcombe, Sydney, 1825 NSW, Australia.

e-mail: joanne.arciuli@sydney.edu.au

Lying is a universal activity and the detection of lying a universal concern. Presently, there is great interest in determining objective measures of deception. The examination of speech, in particular, holds promise in this regard; yet, most of what we know about the relationship between speech and lying is based on the assessment of English speaking participants. Few studies have examined indicators of deception in languages other than English. The world's languages differ in significant ways, and cross-linguistic studies of deceptive communications are a research imperative. Here we review some of these differences amongst the world's languages, and provide an overview of a number of recent studies demonstrating that cross-linguistic research is a worthwhile endeavor. In addition, we report the results of an empirical investigation of pitch, response latency, and speech rate as cues to deception in Italian speech. True and false opinions were elicited in an audio-taped interview. A within-subjects analysis revealed no significant difference between the average pitch of the two conditions; however, speech rate was significantly slower, while response latency was longer, during deception compared with truth-telling. We explore the implications of these findings and propose directions for future research, with the aim of expanding the cross-linguistic branch of research on markers of deception.

Keywords: deception, lying, linguistic markers of deception, cross-linguistic, Italian

INTRODUCTION

Deception can take many forms. Whether it be exaggeration, equivocation, concealment, or an outright lie, deception is a deliberate act that originates with the intent to mislead others (DePaulo et al., 2003). It has been suggested that people lie on average once a day during routine social interactions (DePaulo et al., 1996). Given that we come into contact with lies every day, it is perhaps surprising to discover that many people find it difficult to detect deception. A meta-analysis of 206 studies revealed that humans perform near chance (54%) when making veracity judgments (Bond and DePaulo, 2006). However, most studies involve the elicitation of lies through low-stakes, laboratory-based paradigms and it should be acknowledged that some professional lie-catchers are capable of accuracy rates that are significantly higher than this (Frank and Svetevia, 2012), particularly when they are asked to make veracity judgments in real-life, high-stakes circumstances (Mann et al., 2004). One explanation for poor deception detection performance is that, generally, people hold inaccurate beliefs about what constitutes a reliable indicator of deception (Vrij, 2000). Examination of participants from 75 countries and 43 languages demonstrated that inaccurate beliefs about lie detection are common (Global Deception Research Team, 2006). For example, many people believe that gaze aversion indicates deception (Vrij et al., 2006), a conviction that can compromise lie detection accuracy (Forrest et al., 2004). More recently, it has been suggested that difficulties in lie detection stem from weak associations between cues and deception, rather than people's reliance on inaccurate beliefs about reliable indicators of deception (Hartwig and Bond, 2011).

Regardless of the underlying cause, the mediocre deception detection rates of the average human observer have impelled the search for objective indicators of lying. Traditionally, objective

analyses of lying behavior have been grouped into psychophysiological measures (e.g., heart rate and skin conductivity), "verbal" cues (e.g., the presence of emotive words), and other cues. The latter have sometimes included visual behaviors (e.g., gestures, facial expressions) and what have been referred to as "vocal" or "paraverbal" indices (e.g., pitch and speech rate, see Sporer and Schwandt, 2006, for a review). Here, we have chosen to adopt the term "linguistic" cues, which includes any behavior that is directly associated with oral or written communication. From this perspective, linguistic indicators of lying include the content of both spoken utterances and written communications (e.g., lexical content such as parts-of-speech), along with measures that reflect the way that communication is being delivered (e.g., the analysis of pitch in the case of spoken utterances). A now sizeable body of research has investigated the utility of linguistic cues to deception; however, this research has focused primarily on speakers of English. Lying is a universal activity; hence, it is important to examine linguistic markers of deception beyond English. In the current study, we provide an overview of cross-linguistic research on markers of deception and present empirical data on three potential markers of deception in Italian speech: pitch, response latency, and speech rate.

THEORIES OF DECEPTION

A number of theories have been proposed to explain behavioral differences between deception and truth-telling, including the Four-Factor Model, Interpersonal Deception Theory, the Motivation Impairment Effect, and the Self-Presentational Perspective (for a review, see DePaulo et al., 2003). One of the most influential of these theories is Zuckerman et al.'s (1981a) Four-Factor Model. This model attempts to explain cues to deception in terms of four

psychological processes that may occur during lying compared to truth-telling, specifically: *generalized arousal*, in response to *increased emotion* (fear, guilt, or excitement at deceiving), *cognitive load* (presumably it requires concerted cognitive effort to fabricate a coherent, plausible, consistent account, and maintain a deception), and *attempted control* (deliberate self-regulatory strategies to suppress any leakage of cues). Too much control could result in telling behaviors such as a reduction in emotional expressiveness or reduced hand movement. Alternatively, it may be difficult for deceivers to control all communication channels simultaneously. For example, a deceiver may focus primarily on controlling their facial expression but exert less control over other behaviors.

There is some evidence for Zuckerman et al.'s (1981a) Four-Factor Model to suggest that people do experience one or more of these psychological processes more frequently during deceptive than truthful behavior (e.g., Dionisio et al., 2001; Walczyk et al., 2003; Caso et al., 2005; Gombos, 2006). However, which of these processes will dominate under what circumstances, and which cues to deception are indicative of each of these processes is still being debated in the literature (DePaulo et al., 2003; Caso et al., 2005; Gombos, 2006). While there is debate over the extent to which such processes are under the control of the deceiver, there is general agreement that some cues to deception are non-strategic and frequently outside the deceiver's awareness (DePaulo et al., 2003). It is feasible that some acoustic behaviors, such as pitch and speech rate, might be less vulnerable to behavioral control than other linguistic markers of lying (Villar et al., in press). Vocal pitch, for example, may be more difficult to manipulate when it represents an autonomic response to strong emotion, such as the anxiety an individual may experience while lying (Zuckerman et al., 1981a).

MARKERS OF DECEPTION

The ongoing challenge in lie detection is that there is no single behavior that occurs in all people in every situation and is exclusively related to deceptive behavior (DePaulo et al., 2003). However, some behaviors appear to be more reliable than others. In their meta-analysis, DePaulo and colleagues reviewed 116 studies and coded 158 different cues to deception. These included facial expressions, physical behaviors, and language-related measures (including acoustic measurements). Significant relationships were found between deception and behavioral cues in each of these categories. The results led to the conclusion that liars are "less forthcoming, less compelling, more negative, more tense, and suspiciously bereft of ordinary imperfections and unusual details" (p. 104). Sporer and Schwandt (2006) conducted a meta-analysis of 41 studies that focused on nine cues: speech rate, response latency, message duration, number of words, filled and unfilled pauses, repetitions, speech errors, and pitch. Results indicated that of these cues only pitch ($d = 0.268$) and response latency ($d = 0.177$) were reliably associated with deception, with both showing increases during lying compared to truth-telling.

CROSS-LINGUISTIC RESEARCH

The world's languages differ in many ways, and it follows that there might be differences in the extent to which the cues which have been previously identified as viable markers of lying in English can be applied across languages. Take the case of grammatical

category. A decrease in personal pronoun use has been observed in lying compared to truthful speech for English speaking participants. However, personal pronoun use is overt in English most of the time, so this begs the question: does the deception detection utility of pronoun use extend to null personal pronoun languages, such as Italian and Spanish, where pronoun use is overt only 20–30% of the time (Serratrice, 2005), or to languages such as Japanese which uses considerably fewer pronouns in general than Indo-European languages (Shibatani, 1990)? Likewise, an increase in adjective and adverb use has been observed in lying compared to truthful speech for English speaking participants (Zhou et al., 2004). Yet, not all languages use the same grammatical categories; for instance, Russian has no phrasal verbs (Mudraya et al., 2008), and Polish has no articles (Wierzbicka, 1985). Silent pause duration is another linguistic variable thought to be an indicator of deception in English (Mann et al., 2002); yet, pause duration differs among languages. For example, native speakers of Russian use longer pauses during informal monologs than do native speakers of English (Riazantseva, 2009), while the latter demonstrate shorter silent pauses in read speech than do native speakers of Italian (Campione and Véronis, 2002). The extent to which these differing characteristics are culturally derived is open to debate. Nonetheless, such differences underscore the importance of investigating cues to deception in a range of speakers including but not restricted to English speaking participants.

Previous research on linguistic indicators of deception includes a substantial body of work devoted to two language assessment tools, namely, Criteria-Based Content Analysis (CBCA; Steller and Köehnken, 1989) and Reality Monitoring (RM; Johnson and Raye, 1981). These tools have been successfully used with adult speakers of German, Swedish, Dutch, French, Spanish, and English for the identification of true versus fabricated narratives (Ruby and Brigham, 1997; Vrij et al., 2004; Blandón-Gitlin et al., 2009). Another credibility assessment technique, which is partly derived from CBCA and RM, is Assessment Criteria Indicative of Deception (ACID; Colwell et al., 2007). This tool has been applied to the credibility assessment of Arabic speakers; although, the analysis was performed on the English translation of their oral statements, as opposed to assessing the Arabic utterances directly (Colwell et al., manuscript in progress, cited in Suckle-Nelson et al., 2010). When implemented by trained assessors, each of these techniques can discriminate deceptive from truthful narratives at rates that are higher than chance; however they are labour-intensive and dependent upon contextual clues to veracity (Masip et al., 2005; Vrij, 2005). Evaluating the utility of other markers of lying, that can be measured independent of the judgment of a trained observer, is a worthwhile endeavor. To this end, computerized text analysis programs, such as Linguistic Inquiry Word Count (LIWC; Pennebaker et al., 2007) have been applied to the identification of deceptive text and transcribed verbal utterances in languages other than English, including Spanish, Dutch, Italian, and German (e.g., Schelleman-Offermans and Merckelbach, 2010; Fornaciari and Poesio, 2011; Almela et al., 2012; Hauch et al., 2012; Masip et al., 2012; Sporer, 2012).

Some deception studies do not specify the language in which the lies are elicited, and we are left to deduce the target language from the location of the laboratory in which the research

was conducted. Of those that do specify a language other than English, there appear to be few studies which have examined linguistic markers of deception without the input of a trained assessor (e.g., Anolli and Ciceri, 1997; Anolli et al., 2003; Zhou and Sung, 2008; Schelleman-Offermans and Merckelbach, 2010). Some of the variables that were revealed to be viable markers of deception in English have shown mixed results in studies of other languages. For example, Zhou and Sung (2008) examined the computer-mediated communications of Chinese players engaged in a so-called Mafia Game. Results revealed that, consistent with some studies of English speakers, the use of third person pronouns increased during deception. Inconsistent with findings from some studies of English speakers, there were no significant differences between the proportional use of first person pronouns in the deceivers' versus truth-tellers' messages; however, one limitation of the study reported by Zhou and Sung (2008) was the use of a between-participants design. In a within-subjects design, Schelleman-Offermans and Merckelbach (2010) examined the presence of self-references in the true compared to the fabricated written stories of Dutch speakers. Among other findings, the results showed no significant differences between the presence of self-references in participants' true versus deceptive narratives. While there are methodological differences that may account for the dissimilarities between these findings and those of studies with English speakers, it is possible that some cues which have shown promise in English are not as useful in other languages.

Notably, most studies in languages other than English have examined lying in computer-mediated communication (e.g., Zhou and Sung, 2008), through the written modality (e.g., Schelleman-Offermans and Merckelbach, 2010) or via language analysis of transcribed speech (e.g., CBCA, RM, and ACID). Only a handful of studies (e.g., Anolli and Ciceri, 1997; Anolli et al., 2003) have examined the cross-linguistic utility of acoustically quantifiable markers of deceptive speech. Pitch, response latency, and speech rate are three such variables which have received some attention in studies of English and non-English speaking participants.

Pitch

Pitch refers to our perceptions of how "low" or "high" a voice sounds. The acoustic correlate of pitch is fundamental frequency (F_0), which is a measure of the frequency of vibrations of the vocal tract during speech production. Automated acoustic analysis programs, such as Praat (Boersma and Weenink, 2011), can be used to measure F_0 . Adult males produce an average F_0 between 100 and 150 Hz, while adult females' F_0 tends to be higher with an average between 175 and 250 Hz (Baken and Orlikoff, 2000). The effects of pitch have been noted in situations that vary in terms of emotional involvement. For example, pitch has been shown to increase in situations that evoke strong emotions such as viewing pictures of burn victims (Ekman et al., 1991), and discussing personal beliefs and future plans (Streeter et al., 1977).

While some studies have reported no pitch differences between liars and truth-tellers (Buller and Aune, 1987; Bond et al., 1990; Vrij and Winkel, 1991; Fiedler and Walka, 1993), the findings of two seminal meta-analyses provide support for an overall increase in average pitch across multi-word deceptive compared to truthful utterances (DePaulo et al., 2003; Sporer and Schwandt, 2006).

In addition to studies of average pitch, the deception literature contains examinations of pitch variability (measured as standard deviation of F_0). Various studies have found that there is a significantly greater variation in pitch during deceptive speech compared to truthful speech.

An increase in average pitch, and pitch variability during lying, might be due to an increase in arousal during lying that leads to physiological responses in the body that are difficult to control (Zuckerman et al., 1981a; Sporer and Schwandt, 2006). Heightened emotion, such as the anxiety that is commonly experienced during deception, is thought to intensify tension in the vocal tract, which is responsible for the increase in pitch that accompanies lying. Of relevance to the current study, increases in average pitch and pitch variability have been observed during lying compared to truth-telling in the speech of 31 male Italian undergraduate students (Anolli and Ciceri, 1997). An examination of pitch in deceptive Italian speech, using a sample that includes female participants and older participants (as opposed to a sample comprised entirely of college students), is required.

Response latency

Response latency is the amount of time taken to respond to a question or statement. Several studies have used this definition to measure response latency in relation to deception (e.g., Rockwell et al., 1997b; Feeley and deTurck, 1998; Vrij et al., 2000). Some have reported no difference (Buller et al., 1989) or a decrease (O'Hair et al., 1981; Dulaney, 1982) in response latency in deceptive compared with truthful speech. It has been suggested that decreases in response latency during lying might be a result of the speakers' beliefs that faster responses are associated with a more credible impression (Dulaney, 1982; Buller et al., 1989).

As revealed by the results of Sporer and Schwandt's (2006) meta-analysis, other studies have found that response latency increases during deception compared to truth-telling (e.g., Harrison et al., 1978; deTurck and Miller, 1985; Feeley and deTurck, 1998; Vrij et al., 2000). An increase in response latency has been attributed to the increased cognitive load experienced by a deceiver (Vrij et al., 2000; Sporer and Schwandt, 2006). At the time of writing, we are unaware of any studies which have examined response latency in the speech of Italian speakers during lying compared to truth-telling.

Speech rate

Speech rate refers to the speed with which someone speaks, and can be measured in a variety of ways. Measures of the number of words and syllables, divided by the acoustic length of the utterance (in seconds) are the most common in the deception literature (DePaulo et al., 1982; Riggio and Friedman, 1983; Buller and Aune, 1992; Rockwell et al., 1997a; Feeley and deTurck, 1998; Vrij et al., 2000). Significant variations in speech rate between speakers within the same language have been reported (Ramus, 2002); therefore, it is difficult to refer to an average speech rate for adult speakers. However, the average articulation rate of spontaneous Italian speech has been estimated at 4.9 syllables and 3.4 words per second (Caldognetto et al., 1997). Cross-linguistic investigations have found that speech rate can also vary between languages.

For example, German speakers articulate significantly faster than Italian speakers (Russo and Barry, 2008).

The relationship between speech rate and deception is equivocal in the deception literature. In several studies, significant decreases in speech rate during deceptive versus truthful utterances have been observed (Fiedler and Walka, 1993; Ebisu and Miller, 1994; Rockwell et al., 1997b; Vrij et al., 2000; Vrij and Mann, 2001; Vrij et al., 2008), while non-significant decreases have been observed in some (Mehrabian, 1971; Hocking and Leathers, 1980; Feely and deTurck, 1998), including one study of 31 male Italian speakers (Anolli and Ciceri, 1997). Decreases in speech rate during lying have been attributed to the increase in cognitive load that is thought to accompany lying (Vrij et al., 2008). Significant increases in speech rate during deception have been observed in other studies (Mehrabian, 1971; Klaver et al., 2007). It is possible that methodological differences, particularly in the extent to which participants are cognitively challenged by the experimental task, might account for the different outcomes that have been observed across studies. For example, when given little time for planning, liars speak more slowly than truth-tellers; however, the opposite has been observed when liars are given opportunities to prepare their lie (Sporer and Schwandt, 2006). Participants in the current study were given no preparation time prior to the elicitation of their deceptive response, in order to increase the cognitive challenges of the task.

THE CURRENT STUDY

In summary, deceivers are prone to experiencing (consciously or otherwise) heightened emotion, increased cognitive effort, and attempts at behaviour control (DePaulo et al., 2003; Vrij, 2008). Deceivers may experience the same psychological processes regardless of their background; however, these processes may have different behavioral manifestations depending upon linguistic and/or cultural context. Previous research has investigated the utility of a number of cues to deception. Of these potential deception markers, pitch, response latency, and speech rate were selected for the current study.

In line with previous research conducted with English speakers, and one study of male Italian speakers (Anolli and Ciceri, 1997), it was hypothesized that pitch would be higher in the deceptive speech compared to the truthful speech of Italian speakers. Additionally, it was hypothesized that response latency would be longer in deceptive speech. Due to inconsistencies in the findings of previous studies, the direction and significance of differences in speech rate during deception versus truth-telling was an open empirical question. In light of individual variability amongst participants in terms of their personal speaking style, including differences in pitch, response latency, and speech rate, we employed a within-participants design.

MATERIALS AND METHODS

PARTICIPANTS

Nineteen native speakers of Italian (12 females and 7 males) with a mean age of 56.1 years ($SE = 3.36$) participated in this study. They were recruited in Sydney, Australia, through a variety of methods including word of mouth, advertisements in a local

Italian newspaper, and flyers distributed at Italian community organizations. All participants were born and educated in Italy.

PROCEDURE

Recruitment materials described the study as an investigation of communication skills relating to social issues, in order to avoid attracting participants who considered themselves to be particularly good liars, or those who considered themselves to be poor liars and were hoping to improve their abilities. The same researcher, who was a native speaker of Italian, conducted all of the individual testing sessions in Italian, which took approximately 30 minutes each. All materials and consent forms were provided in Italian.

We employed the well-established false opinion paradigm based on the procedure described by Frank and Ekman (2004) which has been used in a variety of laboratory-based studies of deception (Newman et al., 2003; Arciuli et al., 2010; Villar et al., in press). Participants completed a questionnaire to determine their opinions on various social issues. These social issues are listed in **Table 1**.

Participants were asked to rate the extent to which they agreed or disagreed with each social issue ("1" = completely disagree, "7" = completely agree) as well as the strength of their feelings about the issue ("1" = No feelings, "7" = Very strong feelings). Two issues were then selected for each participant, one about which they would lie, and one about which they would tell the truth. Topics where participants reported strong opinions and strong feelings were chosen. The mean absolute difference of *opinion* ratings from the midpoint of 4 (i.e., mean strength of agreement or disagreement measured as the distance of the value from zero: 1 and 7 become 3, 2 and 6 become 2, and 3 and 5 become 1) were 2.84 ($SE = 0.12$) for the truthful target topics and 2.74 ($SE = 0.15$) for the untruthful target topics. One-sample *t*-tests revealed significant differences between zero and the mean absolute difference of *opinion* ratings for the strength of agreement with the truthful topics [$t(18) = 24.705, p < 0.0001$] and the untruthful topics [$t(18) = 18.258, p < 0.0001$]. A paired samples *t*-test revealed no significant difference between these mean of 2.84 and 2.74 [$t(18) = 0.622, p = 0.542$]. The mean absolute difference

Table 1 | Topics addressed in social issues questionnaire (translated into English here).

Topic	Description
Smoking in public	Should smoking be banned in all enclosed public spaces?
Capital punishment	Should the death penalty be reintroduced in Italy/Australia?
Legalization of marijuana	Should marijuana be legalized for public use?
Legalization of abortion	Should abortion be legal in Australia/Italy?
Same-sex marriage	Should same-sex couples be allowed to marry?
Sex offender registry	Should the identity and location of sex-offenders be made public on the internet?
Church versus state	Should the church be allowed to intervene in political decisions?
Blood alcohol limit	Should the legal blood alcohol limit for driving be lowered from 0.05 to 0.02?

of *feelings* ratings from the midpoint were 2.63 ($SE = 0.18$) for the truthful target topics and 2.26 ($SE = 0.25$) for the untruthful target topics. One-sample *t*-tests revealed significant differences between zero and the mean absolute difference of ratings of the strength of participants' *feelings* toward the truthful topics [$t(18) = 15.076$, $p < 0.0001$] and the untruthful topics [$t(18) = 8.988$, $p < 0.0001$]. A paired samples *t*-test revealed no significant difference between these mean of 2.63 and 2.26 [$t(18) = 1.235$, $p = 0.233$]. Hence, participants' opinions and feelings were (i) sufficiently strong and (ii) equivalent across true and false topics.

Participants were randomly assigned to lie about one of the designated issues and tell the truth about the other. The order of topics was counterbalanced such that half the participants started the interview with a lie and half with the truth. To determine the effect of topic on each of the target variables, one-way ANOVA were conducted. Results revealed that there was no significant effect of topic on pitch [$F(7, 11) = 1.947$, $p = 0.155$], response latency [$F(7,11) = 0.857$, $p = 0.566$], or speech rate [$F(7,11) = 2.362$, $p = 0.098$].

Participants were instructed to provide an honest account of their true opinion of the topic designated for the truthful condition, along with a false representation of their true opinion for the topic designated for the deceptive condition. Participants were told that the interviewer would not know whether they were lying or telling the truth and that they should aim to convince him of their credibility in each of the interviews. Participants were not given any planning time during which to prepare their false or true opinion. The topic was read aloud to the participant who was then asked to state whether they agreed or disagreed and explain why. This was then followed up with a question enquiring whether they were telling the truth. At the conclusion of the interview participants were debriefed and thanked for their cooperation. Interviews were recorded using a Sony Digital Voice Recorder, which has a frequency response of between 80 and 20,000 Hz. All audio files were stored in uncompressed linear PCM (.wav) format for later analysis.

DATA PREPARATION AND ANALYSIS

A native speaker of Italian performed a verbatim Italian transcription of all the interviews. Praat software (Boersma and Weenink, 2011) was used to measure pitch, response latency, and length of utterance (used to calculate speech rate) in each of the audio recordings. In line with Praat software instructions, the speech samples were analyzed using a pitch range of 75–500 Hz for females, and 75–300 Hz for males. Response latency was determined by measuring the time lapse from the end of the first question asked by the interviewer and the start of the participants' response in milliseconds. Duration of response latency was measured via visual examination of the wave form. The portion of the wave form that represented the response latency was magnified, permitting accurate selection and measurement of the latency duration in milliseconds (ms). Recent research suggests that interjections such as "erm" and "um" constitute lexical terms (Arciuli et al., 2010; Villar et al., 2012), and so these were included in the total word count in each transcription. Speech rate was calculated by dividing the total number of words in the utterance, by the acoustic length (measured in seconds).

RESULTS

WORD COUNT

The average number of words produced in the deceptive speech condition was 189.63 ($SE = 17.21$), while the average number of words in the truthful speech condition was 218.84 ($SE = 20.27$). A paired samples *t*-test showed no significant difference between these means [$t(18) = 1.162$, $p = 0.260$, two-tailed].

ACOUSTIC DURATION

The average acoustic duration of the responses in the deceptive speech condition was 100.20 s ($SE = 9.46$). The average duration of the responses in the truthful speech condition was 104.03 s ($SE = 9.70$). A paired samples *t*-test revealed no significant differences between these means [$t(18) = 0.322$, $p = 0.751$, two-tailed].

In order to assess the reliability of the measure of duration of utterance, a second rater measured this variable for just over 50% of the 38 observations ($n = 20$). The inter-rater reliability coefficient was significant ($r = 0.927$, $p < 0.001$), indicating a high consistency between the measurements of acoustic duration that were recorded by the two raters.

PITCH

The average pitch of participants in the deceptive condition was 160.88 Hz ($SE = 7.73$). The average pitch in the truthful condition was very similar at 160.67 Hz ($SE = 7.19$). A paired samples *t*-test revealed no significant difference between the average pitch across conditions [$t(18) = 0.093$, $p = 0.927$, two-tailed] and the effect size was small ($d = 0.006$). In view of the differences in pitch between male and female speakers, additional analyses were performed. The average pitch of female speakers was 178.14 Hz ($SE = 6.24$) during their truthful utterances and 180.21 Hz ($SE = 6.50$) during their deceptive utterances. The average male pitch was 130.74 Hz ($SE = 7.87$) during their truthful utterances and 127.76 Hz ($SE = 8.04$) during their deceptive utterances. As expected, an analysis of gender effects on pitch production, a 2 (veracity: lying versus truth-telling) \times 2 (gender: male versus female) mixed ANOVA revealed a significant main effect of gender [$F(1,17) = 24.632$, $p < 0.0001$, partial $\eta^2 = 0.592$]. However, there was no significant main effect of veracity [$F(1,17) = 0.037$, $p = 0.850$, partial $\eta^2 = 0.002$] and no interaction between gender and veracity [$F(1,50) = 1.147$, $p = 0.299$, partial $\eta^2 = 0.063$].

Further analyses were performed to determine variability in pitch (measured as the standard deviation of F_0 , in Hz) in each condition. A paired samples *t*-test revealed no significant difference between pitch variability in the true ($M = 56.12$, $SE = 5.67$) compared to the lying ($M = 57.60$, $SE = 5.91$) condition [$t(18) = 0.344$, $p = 0.735$, two-tailed, $d = 0.06$]. Using a median split analysis, the variable of average F_0 was dichotomized into groups (variability: low/high) for each condition (truthful/deceptive). An independent samples *t*-test revealed no significant difference between the average F_0 of the truthful ($M = 163.30$, $SE = 7.30$) compared to the deceptive ($M = 173.70$, $SE = 6.30$) conditions for the low variability group [$t(17) = 1.066$, $p = 0.301$, two-tailed], even in view of a moderate effect size ($d = 0.49$). Similarly, for the high variability group, there were

no significant differences between the average F_0 of the truthful ($M = 157.75$, $SE = 12.30$) and the deceptive ($M = 149.35$, $SE = 149.35$) conditions [$t(17) = 0.454$, $p = 0.655$, two-tailed]. The effect size was small ($d = 0.20$).

RESPONSE LATENCY

A Kolmogorov-Smirnov test of the entire data-set revealed that the distribution of scores for response latency in the truthful speech condition, $D(19) = 0.350$, $p < 0.001$, and the deceptive speech condition, $D(19) = 0.378$, $p < 0.001$, were both significantly non-normal. Consequently, the data were analyzed using a non-parametric alternative to a paired samples t -test: the Wilcoxon signed-rank test. Results showed that response latency (in ms) was longer in the deceptive speech condition ($Mdn = 1200.77$) than in the truthful speech condition ($Mdn = 775.26$). This difference was significant, $T = 9.50$, $p = 0.02$, and the effect size was large ($r = -0.51$).

A second rater measured response latency in just over 50% of the 38 observations ($n = 20$). The inter-rater reliability coefficient was significant ($r = 0.998$, $p < 0.001$), indicating a high consistency between the measurements of response latency that were recorded by the two raters.

SPEECH RATE

The average speech rate (in words per second) was slower in the deceptive speech condition ($M = 1.95$, $SE = 0.08$) compared to the truthful speech condition ($M = 2.10$, $SE = 0.07$). A paired samples t -test revealed a significant difference between the means [$t(18) = 2.454$, $p = 0.025$, two-tailed]. The effect size was medium ($d = 0.447$).

DISCUSSION

Here we examined whether pitch, response latency, and speech rate are helpful in distinguishing between deceptive and truthful communications in Italian. Our hypothesis that pitch would be higher in the deceptive speech condition was not supported. As hypothesized, we found that response latency was significantly longer in the deceptive speech condition compared to the truthful speech condition. It was an open empirical question as to whether participants' speech rate would differ during lying compared to truth-telling. The data revealed a significant difference between the average speech rate for the two conditions: speech rate was significantly slower in the deceptive versus the truthful speech condition. The lies in the present study were, on average, of a relatively short duration (around 100 s); yet, they were of a sufficient length to enable the detection of significant changes in response latency and speech rate during lying compared to truth-telling.

PITCH

It has been documented that increased pitch is one of the cues that people associate with deceptive speech (Zuckerman et al., 1981b; Vrij and Semin, 1996; Anderson et al., 1999; Lakhani and Taylor, 2003; Colwell et al., 2006). Speakers sometimes employ counter-measures to appear more believable when they lie (Sip et al., 2008). Consequently, it is possible that some of the subjects in our study strategically managed their vocal pitch in an attempt to appear more credible. However, a recent study found

that those individuals who believed that pitch increases during deception, demonstrated a significantly higher pitch during their own deceptive utterances (Villar et al., in press). Thus, it is unlikely that attempts at behavioral control can explain the findings in the current study.

The pitch values we observed are in line with previous reports of pitch in adult females and males (Villar et al., in press). Additional analyses were conducted in order to determine whether differences in pitch across females and males may have influenced the mean pitch results. There was no main effect for veracity, nor a significant interaction between gender and veracity. Therefore, it is unlikely that gender had a systematic impact on our mean pitch results.

In addition to measuring mean pitch, variability in pitch is another frequently used measure in voice research (Neil et al., 2003). Deception research, also, has looked at the effects of lying on pitch variation and found greater pitch variation in deceptive speech compared to truthful speech (Anolli and Ciceri, 1997; Rockwell et al., 1997a). Our analyses indicated no significant difference in pitch variability across the truthful versus deceptive conditions, nor was there a significant difference in the average F_0 between the truthful and deceptive conditions for either the high variability group or the low variability group. Thus, regardless of whether a speaker's pitch variability was high or low there was no difference in the average vocal pitch of their truthful compared to their deceptive speech.

Sporer and Schwandt's (2006) meta-analysis found that pitch was significantly higher during lying when participants lied about "facts and feelings," as opposed to "facts only." The explanation offered for this finding is that, in the absence of increased emotional arousal, pitch remains the same during lying compared to truth-telling. It is possible that we did not observe the expected increases in pitch during lying because of the paradigm we used to elicit the lies. Perhaps, despite our attempts to elicit topics about which participants felt strongly (see Materials and Methods), the topics were not sufficiently arousing to be accompanied by pitch changes. However, this explanation seems problematic given that our paradigm was successful in eliciting differences in response latency and response rate in lying versus truthful Italian speech.

While all the participants in our study were native speakers of Italian who were born and educated in Italy, they were also speakers of English who were visiting or residing in Australia. It is possible that the bilingual status of the participants influenced their speech. However, we think it unlikely that bilingual status would have a systematic impact upon truth-telling and lies such that exposure to English as a second language would result in a lack of pitch differences in native Italian speech. The speech of bilinguals has been shown to incorporate phonological and prosodic features from both languages (Jusczyk, 1997). Thus, having English as a second language might increase the likelihood that Italian speakers would speak in a higher pitch during lying compared to truth-telling, in the same way that English speakers appear to. However, this was not the case in the current study. Anolli and Ciceri (1997) reported significant differences in the mean, range, and variability in pitch of the deceptive versus truthful utterances of their male Italian participants. Their participants were younger than ours ($M = 24.4$ years versus $M = 56.1$ years). Perhaps there

are age-related differences in pitch during lying compared to truth-telling that could account for the differences between our findings and those of Anolli and Ciceri.

Lastly, it is feasible that there are differences between languages in the efficacy of pitch as an indicator of deception, which are culturally, as opposed to linguistically determined. For example, Van Bezooijen (1995) showed that Japanese women produce a higher pitch on average than Dutch women, and suggested that these differences reflect the characteristics that are perceived to be desirable in women in each culture (i.e., a preference for high pitch in Japanese women, and low to medium pitch in women from the Netherlands). Future studies might consider the socio-cultural factors that could influence the viability of pitch as a marker of deception in languages other than English.

Further research is required in order to explore these possibilities.

RESPONSE LATENCY

The response latency results are in line with the findings of Sporer and Schwandt's (2006) meta-analysis. Of note, we observed a large effect size concerning response latency. As explained by the four-factor theory (Zuckerman et al., 1981a), a longer response latency during deception may be due to the increased cognitive load associated with lying which can lead to "leakage" of certain behaviors (Vrij et al., 2000; Sporer and Schwandt, 2006). Sporer and Schwandt (2006) proposed that the increased cognitive load experienced during deception is due to increased demands on working memory. In other words, when a pre-existing schema or script is not available, which is often the case during lying, the formulation of novel ideas is required. This increases the load on working memory, leaving less capacity available for speech production, which can lead to increased latencies. Short planning times for lie formulation have been associated with longer latencies (Sporer and Schwandt, 2007), and it is possible that the low levels of preparation time in the current study contributed to the efficacy of this variable.

SPEECH RATE

Previous studies of speech rate during lying compared to truth-telling have produced conflicting results. Our findings are consistent with those studies which have found that speech rate is significantly slower during lying compared to truth-telling (Fiedler and Walka, 1993; Ebisu and Miller, 1994; Rockwell et al., 1997b; Vrij et al., 2000; Vrij and Mann, 2001; Vrij et al., 2008). Once again, the increases in cognitive load that are thought to accompany lying might reduce the cognitive capacity available for other activities, such as speech production (Sporer and Schwandt, 2006). One consequence of this might be the slower speech that we have observed here during lying. Notably, our findings are in the same direction

REFERENCES

- Almela, Á., Valencia-García, R., and Cantos, P. (2012). "Seeing through deception: a computational approach to deceit detection in written communication," in *Proceedings of the EACL 2012 Workshop on Computational Approaches to Deception Detection* (Avignon: The Association for Computer Linguistics), 15–22.
- Anderson, D. E., DePaulo, B. M., Ansfield, M. E., Tickle, J. J., and Green, E. (1999). Beliefs about cues to deception: mindless stereotypes or untapped wisdom? *J. Nonverbal Behav.* 23, 67–89.
- Anolli, L., and Ciceri, R. (1997). The voice of deception: vocal strategies of naive and able liars. *J. Nonverbal Behav.* 21, 259–284.
- Anolli, L. M., Balconi, M., and Ciceri, R. (2003). Linguistic styles in deceptive communication: dubitative ambiguity and elliptic eluding in packaged lies. *Soc. Behav. Pers.* 31, 687–711.
- Arciuli, J., Mallard, D., and Villar, G. (2010). "Um, I can tell you're lying":

as those of Anolli and Ciceri (1997) who found a decrease (albeit a non-significant one) in speech rate during lying compared to truth-telling for their 31 male Italian speakers. It is worth noting the different methodologies that were utilized: Anolli and Ciceri's participants described a black and white picture, while participants in the current study described opinions of social topics. It could be argued that the latter involves a more emotive and cognitively demanding task (but that explanation becomes a little problematic when interpreting discrepant results between the two studies concerning pitch).

Future studies might consider the cross-linguistic utility of speech rate as a marker of lying in languages other than English and Italian. It may be that measures of words per second are not appropriate for all languages. For instance, in languages such as Japanese and Filipino, where the morphology is highly agglutinative, a more appropriate measure of speech rate might be the number of morphemes per second.

CONCLUSION

This research investigated the effects of veracity on pitch, speech rate, and response latency in the speech of native speakers of Italian. Each of these variables has been linked to deception in the speech of native speakers of English. Our findings revealed that response latency and speech rate are associated with deception in the speech of native speakers of Italian in the same way that they are for English speakers. No relationship was found between pitch and lying in the present study. Additional studies are required to determine whether pitch is a reliable marker of lying in languages other than English. Further investigations of the extent to which differences in deceptive communications across languages are linguistically, as opposed to culturally derived, are also required. In our view, a systematic analysis of the utility of a range of linguistic variables in cross-linguistic and cross-cultural contexts would be invaluable for deception research. Another very interesting avenue for research is the comparison of linguistic cues to deception in monolingual and bilingual speakers (including comparison of simultaneous bilinguals versus those that acquired a second language after acquiring their first). It would also be valuable to assess lying versus truth across multiple languages in a within-subjects design to see if cues to deception are used by the same multilingual speakers regardless of the language they are speaking. We hope that the current study encourages expansion of this line of deception research. It remains to be seen whether a unique pattern of cues to deception will emerge for each language or whether we will discover that there are some markers of lying that are common across languages.

ACKNOWLEDGMENTS

Data collection was undertaken with the assistance of Alessio Barsaglini from the University Degli Studi di Padova, Padova, Italy.

- linguistic markers of deception versus truth-telling in speech. *Appl. Psycholinguist.* 31, 397–411.
- Baken, R. J., and Orlikoff, R. F. (2000). *Clinical Measurement of Speech and Voice*. San Diego: Singular Pub Group.
- Blandón-Gitlin, I., Pezdek, K., Lindsay, D. S., and Hagen, L. (2009). Criteria-based content analysis of true and suggested accounts of events. *Appl. Cogn. Psychol.* 23, 901–917.
- Boersma, P., and Weenink, D. (2011). *Praat 5.2.0.1. [Software]*. Amsterdam: Institute of Phonetic Sciences.
- Bond, C., Omar, A., Mahmoud, A., and Bonser, R. N. (1990). Lie detection across cultures. *J. Nonverbal Behav.* 14, 189–204.
- Bond, C. E., and DePaulo, B. M. (2006). Accuracy of deception judgments. *Pers. Soc. Psychol. Rev.* 10, 214–234.
- Buller, D. B., and Aune, R. K. (1987). Nonverbal cues to deception among intimates, friends, and strangers. *J. Nonverbal Behav.* 11, 269–290.
- Buller, D. B., and Aune, R. K. (1992). The effects of speech rate similarity on compliance: application of communication accommodation theory. *West. J. Speech Commun.* 56, 37–53.
- Buller, D. B., Comstock, J., Aune, R. K., and Strzyzewski, K. D. (1989). The effect of probing on deceivers and truthtellers. *J. Nonverbal Behav.* 13, 155–170.
- Caldognetto, E. M., Zmarich, C., and Ferrero, F. (1997). “A comparative acoustic study of spontaneous and read Italian speech,” in *Proceedings of Eurospeech*, eds G. Kokkinakis, N. Fakotakis and E. Dermatas (Rhodes: ISCA), 770–782.
- Campione, E., and Véronis, J. (2002). “A large-scale multilingual study of silent pause duration,” *Speech Prosody 2002 International Conference*, (Aix-En-Provence: Laboratoire Parole et Langage).
- Caso, L., Gnisci, A., Vrij, A., and Mann, S. (2005). Processes underlying deception: an empirical analysis of truth and lies when manipulating the stakes. *J. Invest. Psychol. Offender Profil.* 2, 195–202.
- Colwell, K., Hiscock-Anisman, C. K., Memon, A., Taylor, L., and Prewett, J. (2007). Assessment criteria indicative of deception (ACID): an integrated system of investigative interviewing and detecting deception. *J. Invest. Psychol. Offender Profil.* 4, 167–180.
- Colwell, L. H., Miller, H. A., Miller, R. S., and Lyons, P. M. (2006). US police officers’ knowledge regarding behaviours indicative of deception: implications for eradicating erroneous beliefs through training. *Psychol. Crime Law* 12, 489–503.
- DePaulo, B., Kashy, D., Kirkendol, S., Wyer, M., and Epstein, J. (1996). Lying in everyday life. *J. Pers. Soc. Psychol.* 70, 979–995.
- DePaulo, B., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., and Cooper, H. (2003). Cues to deception. *Psychol. Bull.* 129, 74–118.
- DePaulo, B. M., Rosenthal, R., Rosenkrantz, J., and Green, C. R. (1982). Actual and perceived cues to deception: a closer look at speech. *Basic Appl. Soc. Psych.* 3, 291–312.
- deTurck, M. A., and Miller, G. R. (1985). Deception and arousal: isolating the behavioral correlates of deception. *Hum. Commun. Res.* 12, 181–201.
- Dionisio, D. P., Granholm, E., Hillix, W. A., and Perrine, W. F. (2001). Differentiation of deception using pupillary responses as an index of cognitive processing. *Psychophysiology* 38, 205–211.
- Dulaney, E. F. (1982). Changes in language behavior as a function of veracity. *Hum. Commun. Res.* 9, 75–82.
- Ebesu, A. S., and Miller, M. D. (1994). Verbal and nonverbal behaviors as a function of deception type. *J. Lang. Soc. Psychol.* 13, 418–442.
- Ekman, P., O’Sullivan, M., Friesen, W., and Scherer, K. (1991). Invited article: face, voice, and body in detecting deceit. *J. Nonverbal Behav.* 15, 125–135.
- Feeley, T. H., and deTurck, M. A. (1998). The behavioral correlates of sanctioned and unsanctioned deceptive communication. *J. Nonverbal Behav.* 22, 189–204.
- Fiedler, K., and Walka, I. (1993). Training lie detectors to use nonverbal cues instead of global heuristics. *Hum. Commun. Res.* 20, 199–223.
- Fornaciari, T., and Poesio, M. (2011). “Sincere and deceptive statements in Italian criminal proceedings,” in *Proceedings of the International Association of Forensic Linguists Tenth Biennial Conference* (Wales: Cardiff), 126–138.
- Forrest, J., Feldman, R., and Tyler, J. (2004). When accurate beliefs lead to better lie detection. *J. Appl. Soc. Psychol.* 34, 764–780.
- Frank, M., and Ekman, P. (2004). Appearing truthful generalizes across different deception situations. *J. Pers. Soc. Psychol.* 86, 486–495.
- Frank, M. G., and Svetieva, E. (2012). Lies worth catching involve both emotion and cognition: commentary on Vrij and Granham (2012). *J. Appl. Res. Mem. Cogn.* 1, 110–117.
- Global Deception Research Team. (2006). A world of lies. *J. Cross Cult. Psychol.* 37, 60–74.
- Gombos, V. A. (2006). The cognition of deception: the role of executive processes in producing lies. *Genet. Soc. Gen. Psychol. Monogr.* 132, 197–214.
- Harrison, A. A., Hwalek, M., Raney, D. F., and Fritz, J. G. (1978). Cues to deception in an interview situation. *Soc. Psychol.* 41, 156–161.
- Hartwig, M., and Bond, C. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychol. Bull.* 137, 643.
- Hauch, V., Blandón-Gitlin, I., Masip, J., and Sporer, S. L. (2012). “Linguistic cues to deception assessed by computer programs: a meta-analysis,” in *Proceedings of the EACL 2012 Workshop on Computational Approaches to Deception Detection* (Avignon: The Association for Computer Linguistics), 1–4.
- Hocking, J. E., and Leathers, D. G. (1980). Nonverbal indicators of deception: a new theoretical perspective. *Commun. Monogr.* 47, 119–131.
- Johnson, M. K., and Raye, C. L. (1981). Reality monitoring. *Psychol. Rev.* 88, 67.
- Jusczyk, P. W. (1997). *The Discovery of Spoken Language*. Massachusetts: The MIT press.
- Klaver, J. R., Lee, Z., and Hart, S. D. (2007). Psychopathy and nonverbal indicators of deception in offenders. *Law Hum. Behav.* 31, 337–351.
- Lakhani, M., and Taylor, R. (2003). Beliefs about the cues to deception in high- and low-stake situations. *Psychol. Crime Law* 9, 357–368.
- Mann, S., Vrij, A., and Bull, R. (2002). Suspects, lies, and videotape: an analysis of authentic high-stake liars. *Law Hum. Behav.* 26, 365–376.
- Mann, S., Vrij, A., and Bull, R. (2004). Detecting true lies: police officers’ ability to detect suspects’ lies. *J. Appl. Psychol.* 89, 137–149.
- Masip, J., Bethencourt, M., Lucas, G., Sanchez-San Segundo, M., and Herrero, C. (2012). Deception detection from written accounts. *Scand. J. Psychol.* 53, 103–111.
- Masip, J., Sporer, S. L., Garrido, E., and Herrero, C. (2005). The detection of deception with the reality monitoring approach: a review of the empirical evidence. *Psychol. Crime Law* 11, 99–122.
- Mehrabian, A. (1971). Nonverbal betrayal of feeling. *J. Exp. Res. Pers.* 5, 64–73.
- Mudraya, O., Piao, S. S. L., Rayson, P., Sharoff, S., Babych, B., and Lofberg, L. (2008). “Automatic extraction of translation equivalents of phrasal and light verbs in English and Russian,” in *Phraseology: An Interdisciplinary Perspective*, ed. S. Granger and F. Meunier (Amsterdam: John Benjamins Publishing Co.), 293–309.
- Neil, E., Worral, L., Day, A., and Hickson, L. (2003). Voice and speech characteristics and vocal hygiene in novice and professional broadcast journalists. *Adv. Speech Lang. Pathol.* 5, 1–14.
- Newman, M. L., Pennebaker, J. W., Berry, D. S., and Richards, J. M. (2003). Lying words: predicting deception from linguistic styles. *Pers. Soc. Psychol. Bull.* 29, 665–675.
- O’Hair, H. D., Cody, M. J., and McLaughlin, M. L. (1981). Prepared lies, spontaneous lies, Machiavellianism, and nonverbal communication. *Hum. Commun. Res.* 7, 325–339.
- Pennebaker, J., Booth, R., and Francis, M. (2007). *Linguistic Inquiry and Word Count (LIWC2007). Operator’s manual*. Austin, TX: LIWC Inc.
- Ramus, F. (2002). “Acoustic correlates of linguistic rhythm: perspectives,” in *Conference on Speech Prosody*, eds B. Bel and I. Marlien, (Aix-en-Provence: Laboratoire Parole et Langage), 115–120.
- Riazantseva, A. (2009). Second language proficiency and pausing: a study of Russian speakers of English. *Stud. Second Lang. Acq.* 23, 497–526.
- Riggio, R. E., and Friedman, H. S. (1983). Individual differences and cues to deception. *J. Pers. Soc. Psychol.* 45, 899–915.
- Rockwell, P., Buller, D., and Burgoon, J. K. (1997a). Measurement of deceptive voices: comparing acoustic and perceptual data. *Appl. Psycholinguist.* 18, 471–484.
- Rockwell, P., Buller, D. B., and Burgoon, J. K. (1997b). The voice of deceit: refining and expanding vocal cues to deception. *Commun. Res. Rep.* 14, 451–459.
- Ruby, C. L., and Brigham, J. C. (1997). The usefulness of the criteria-based content analysis technique in distinguishing between truthful and fabricated allegations: a critical review. *Psychol. Public Policy Law* 3, 705.
- Russo, M., and Barry, W. J. (2008). “Isochrony reconsidered: objectifying relations between rhythm measures and speech tempo,” in *Conference on Speech Prosody*, eds P. Barbosa, S. Madureira, and C. Reis (Campinas: ISCA), 419–422.

- Schelleman-Offermans, K., and Merckelbach, H. (2010). Fantasy proneness as a confounder of verbal lie detection tools. *J. Invest. Psychol. Offender Profil.* 7, 247–260.
- Serratrice, L. (2005). The role of discourse pragmatics in the acquisition of subjects in Italian. *Appl. Psycholinguist.* 26, 437–462.
- Shibatani, M. (1990). *The Languages of Japan*. Cambridge: Cambridge University Press.
- Sip, K. E., Roepstorff, A., McGregor, W., and Frith, C. D. (2008). Detecting deception: the scope and limits. *Trends Cogn. Sci.* 12, 48–53.
- Sporer, S. L. (2012). “Making the subjective objective? Computer-assisted quantification of qualitative content cues to deception,” in *Proceedings of the EACL 2012 Workshop on Computational Approaches to Deception Detection* (Avignon: The Association for Computer Linguistics), 78–85.
- Sporer, S. L., and Schwandt, B. (2006). Paraverbal indicators of deception: a meta analytic synthesis. *Appl. Cogn. Psychol.* 20, 421–446.
- Sporer, S. L., and Schwandt, B. (2007). Moderators of nonverbal indicators of deception: a meta-analytic synthesis. *Psychol. Public Policy Law* 13, 1–34.
- Steller, M., and Köhnken, G. (1989). “Criteria-based statement analysis,” in *Psychological Methods in Criminal Investigation and Evidence*, ed. D. C. Raskin (New York, NY: Springer Publishing Co.), 217–245.
- Streeter, L. A., Krauss, R. M., Geller, V., Olson, C., and Apple, W. (1977). Pitch changes during attempted deception. *J. Pers. Soc. Psychol.* 35, 345–350.
- Suckle-Nelson, J. A., Colwell, K., Hiscock-Anisman, C., Florence, S., Youschak, K. E., and Duarte, A. (2010). Assessment criteria indicative of deception (ACID): replication and gender differences. *Open Criminol. J.* 3, 23–30.
- Van Bezooijen, R. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. *Lang. Speech* 38, 253–265.
- Villar, G., Arciuli, J., and Mallard, D. (2012). Use of “um” in the deceptive speech of a convicted murderer. *Appl. Psycholinguist.* 33, 83–95.
- Villar, G., Arciuli, J., and Paterson, H. (in press). Vocal pitch production during lying: beliefs about deception matter. *Psychiatry Psychol. Law*.
- Vrij, A. (2000). *Telling Lies and Detecting Deceit: The Psychology of Lying and the Implications for Professional Practice*. Chichester: Wiley.
- Vrij, A. (2005). Criteria-based content analysis: a qualitative review of the first 37 studies. *Psychol. Public Policy Law* 11, 3.
- Vrij, A. (2008). *Detecting Lies and Deceit: Pitfalls and Opportunities*. West Sussex: Wiley-Interscience.
- Vrij, A., Akehurst, L., and Knight, S. (2006). Police officers’, social workers’, teachers’ and the general public’s beliefs about deception in children, adolescents and adults. *Legal Criminol. Psychol.* 11, 297–312.
- Vrij, A., Akehurst, L., Soukara, S., and Bull, R. (2004). Let me inform you how to tell a convincing story: CBCA and reality monitoring scores as a function of age, coaching, and deception. *Can. J. Behav. Sci.* 36, 113–126.
- Vrij, A., Edward, K., Roberts, K. P., and Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *J. Nonverbal Behav.* 24, 239–263.
- Vrij, A., and Mann, S. (2001). Telling and detecting lies in a high-stake situation: the case of a convicted murderer. *Appl. Cogn. Psychol.* 15, 187–203.
- Vrij, A., Mann, S., Fisher, R., Leal, S., Milne, R., and Bull, R. (2008). Increasing cognitive load to facilitate lie detection: the benefit of recalling an event in reverse order. *Law Hum. Behav.* 32, 253–265.
- Vrij, A., and Semin, G. R. (1996). Lie experts’ beliefs about nonverbal indicators of deception. *J. Nonverbal Behav.* 20, 65–80.
- Vrij, A., and Winkel, F. (1991). “The frequency and scope of differences in nonverbal behavioral patterns: an observation study of Dutchmen and Surinamers,” in *Contemporary Issues in Crosscultural Psychology*, eds N. Bleichrodt and P. Drenth (Amsterdam: Swets en Zeitlinger), 120–137.
- Walczuk, J. J., Roper, K. S., Seemann, E., and Humphrey, A. M. (2003). Cognitive mechanisms underlying lying to questions: response time as a cue to deception. *Appl. Cogn. Psychol.* 17, 744–755.
- Wierzbicka, A. (1985). Different cultures, different languages, different speech acts: Polish vs. English. *J. Pragmat.* 9, 145–178.
- Zhou, L., Burgoon, J. K., Nunamaker, J. F., and Twitchell, D. (2004). Automating linguistics-based cues for detecting deception in text-based asynchronous computer-mediated communication. *Group Decis. Negot.* 13, 81–106.
- Zhou, L., and Sung, Y. W. (2008). “Cues to deception in online Chinese groups,” *Annual Hawaii International Conference on System Sciences*, (Big Island: IEEE Computer Society).
- Zuckerman, M., DePaulo, B. M., and Rosenthal, R. (1981a). Verbal and nonverbal communication of deception. *Adv. Exp. Soc. Psychol.* 14, 1–59.
- Zuckerman, M., Koestner, R., and Driver, R. (1981b). Beliefs about cues associated with deception. *J. Nonverbal Behav.* 6, 105–114.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 25 July 2012; accepted: 09 October 2012; published online: 30 October 2012.

Citation: Spence K, Villar G and Arciuli J (2012) Markers of deception in Italian speech. *Front. Psychology* 3:453. doi: 10.3389/fpsyg.2012.00453

This article was submitted to Frontiers in Cognitive Science, a specialty of Frontiers in Psychology.

Copyright © 2012 Spence, Villar and Arciuli. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.



Exploring the movement dynamics of deception

Nicholas D. Duran^{1*}, Rick Dale¹, Christopher T. Kello¹, Chris N. H. Street² and Daniel C. Richardson²

¹ Cognitive and Information Sciences, University of California Merced, Merced, CA, USA

² Cognitive, Perceptual and Brain Sciences, University College London, London, UK

Edited by:

Wolfgang Ambach, Institute for Frontier Areas of Psychology and Mental Health (IGPP), Germany

Reviewed by:

Giorgio Ganis, Plymouth University, UK

Birthe Aßmann, Niedersächsisches Institut für Frühkindliche Bildung und Entwicklung, Germany

Leanne Ten Brinke, University of California Berkeley, USA

***Correspondence:**

Nicholas D. Duran, Cognitive and Information Sciences, University of California Merced, 5200 Lake Road, Merced, CA 95343, USA.
e-mail: nduran2@ucmerced.edu

Both the science and the everyday practice of detecting a lie rest on the same assumption: hidden cognitive states that the liar would like to remain hidden nevertheless influence observable behavior. This assumption has good evidence. The insights of professional interrogators, anecdotal evidence, and body language textbooks have all built up a sizeable catalog of non-verbal cues that have been claimed to distinguish deceptive and truthful behavior. Typically, these cues are discrete, individual behaviors—a hand touching a mouth, the rise of a brow—that distinguish lies from truths solely in terms of their frequency or duration. Research to date has failed to establish any of these non-verbal cues as a reliable marker of deception. Here we argue that perhaps this is because simple tallies of behavior can miss out on the rich but subtle organization of behavior as it unfolds over time. Research in cognitive science from a dynamical systems perspective has shown that behavior is structured across multiple timescales, with more or less regularity and structure. Using tools that are sensitive to these dynamics, we analyzed body motion data from an experiment that put participants in a realistic situation of choosing, or not, to lie to an experimenter. Our analyses indicate that when being deceptive, continuous fluctuations of movement in the upper face, and somewhat in the arms, are characterized by dynamical properties of less stability, but greater complexity. For the upper face, these distinctions are present despite no apparent differences in the overall amount of movement between deception and truth. We suggest that these unique dynamical signatures of motion are indicative of both the cognitive demands inherent to deception and the need to respond adaptively in a social context.

Keywords: **deception, non-linear measures, Dynamical Systems Theory, embodiment, recurrence quantification analysis, multiscale entropy analysis, body and facial movements, time series analysis**

INTRODUCTION

The keystone of “dynamical cognition” is the intimate relationship between mental and motor processes. Rather than the mind being limited to abstract computation, encapsulated from the body and its interactions with the environment, the connections between cognition, action, and perception are tightly intertwined (Port and Van Gelder, 1995; Riley et al., 2012). Consider the interlocked rhythms of speech and gesture, where hand and arm movements are timed to coincide with the articulation of words and phrases during communication. The exact timings suggest that information carried in gesture subserves the transmission of meaning, with both arising from the same underlying cognitive processes (McNeill, 1996). Such a relationship counters notions that the path between cognition and movement is one of discrete, sequential steps, where instructions to act are handed down from a central executive. Instead, cognition and action formed a coupled system that co-varies in systematic ways.

The connection between thought and action also suggests that hidden cognitive processes can be revealed in the dynamics of movement, such as those that occur during deception. Indeed, deception likely elicits unique cognitive demands that vary markedly from truthful communication (Vrij et al., 2010). By definition, deception requires mental partitioning of what is and what is not the case, and an intentional effort to convince

listeners of the latter. In addition, it often occurs face-to-face, where a large array of motor cues are available, from movements of the hands and eyes, to facial movements and changes in articulatory patterns. Given this mind–body relationship, the possible consequences on deceptive behavior have not gone unstudied. However, overwhelming focus has been placed on discrete individual behaviors that can be noted and counted by human observers (e.g., see Vrij et al., 1996; Hill and Craig, 2002). In doing so, the dynamics of how movements are patterned across time have not been examined, and may in part explain why detection reliability in existing studies remains quite low (Bond and DePaulo, 2006).

Here, we take a different tack by examining the moment-by-moment temporal dependencies that reside in patterns of motion. At this more granular level, we are able to provide a *dynamical systems* account of deceivers’ continuous movements in naturalistic contexts. By examining how fluctuations of movement are structured in time, new insights can be had about the manner in which mental dynamics are expressed in bodily dynamics. These insights are particularly relevant for evaluating existing studies based on an implicit assumption that deception negatively interferes with normal processes of communication. Such an assumption leads to explanations that are typically couched in terms of greater processing load, whereby attentional resources

are presumably diverted away from, or overly committed to, the control of action (Ekman and Friesen, 1972; DePaulo, 1992; DePaulo and Friedman, 1998; Vrij et al., 2008). A consequence is that normal behavior is believed to be impaired in some way, often evidenced by decreases in movement frequency and duration (DePaulo et al., 2003; Porter and ten Brinke, 2010; Vrij et al., 2010).

From a dynamical systems perspective, this conclusion is based on a relatively coarse relationship between mind and body. As will be discussed further in the following section (“Complexity in Movement Variability”), increases or decreases in movement can serve only as gross indicators of how the cognitive and motor systems are indeed impaired. Rather, what is most telling are the structural properties of *stability* and *complexity* that are derived from the fine-grained changes in movement variability. It is here that the influences of deception might be more directly revealed. We hypothesize that the outcome may not be one of impairment, but instead a reorganization of behavior over time that is better able to flexibly respond to the changing demands in deceptive contexts. Although we provide additional justification for this claim (see section “Adaptive Responding During Deception”), it is important to note that our arguments can only be, at present, speculative. Nonetheless, combining existing cognitive accounts of deception and deception detection with further exploration of dynamics may be a fruitful avenue of investigation. We will argue that dynamics may hold great promise in distinguishing deception from truth, as well as in understanding the underlying cognitive processes during deception.

We examine such possibilities by reanalyzing the bodily dynamics of participants in a deception experiment performed by Eapen et al. (2010). They designed two scenarios to elicit deception in participants who believed they were taking part in a study of mathematical ability and balance. Throughout the experiment, 29 points on the body, head, and on the face were rapidly sampled in three-dimensional space every 5 ms¹.

In the first scenario, participants performed two math tests, and were offered a £5 reward if they performed better on the second test. Crucially, only they knew how well they actually performed on the second test, but since the difficulty was calibrated carefully, we could be confident that they performed worse.

As part of the second scenario, participants witnessed a laptop being accidentally dropped by a junior investigator. In fact, the accident was staged and purposefully occurred while the senior researcher was out of the room. Later, the senior researcher returned, found his laptop not working, and asked the participant if anything had happened to it. Part of the participants’ motivation to lie was the demeanor of the experimenters. The senior researcher was brusque and unpleasant throughout, but the junior researcher was very friendly toward the participant and expressed anxiety that she would be found out.

In both scenarios the participant was given the means, motive, and the opportunity to spontaneously lie to the experimenter. About 60% did so in each case. Eapen et al. found that while lying, compared to telling the truth, participants tended to move less.

This conclusion was based on overall movement displacement across all motion points on the body. It echoes previous findings in the literature, albeit with a more refined, automated analysis. Here, we aim to extend these findings in two critical ways. First, by introducing two non-linear measures used in the biological and physical sciences that provide a novel analysis of the motor dynamics of deception. Second, by considering the theoretical implications that such characterizations of behavior have on the responsiveness of the cognitive system during deception. To better serve these goals, we turn next to an area of dynamical systems research that strongly motivates the current approach.

UNRAVELING THE DYNAMICS OF MOVEMENT

COMPLEXITY IN MOVEMENT VARIABILITY

Even with the most basic types of control, the motor system faces the problem of how to constrain multiple and redundant bodily degrees of freedom in producing coherent, functional behaviors (Bernstein, 1967; Dickinson et al., 2000; Turvey, 2007). Given the countless physiological, contextual, and environmental interactions that are undoubtedly at play, assemblies of behavior cannot be captured by simple linear measures of more or less movement (Newell, 1998; Harbourne and Stergiou, 2009; Riley et al., 2012). Rather, the interactions are expressed as a process of self-organization, whereby the coordination of the musculoskeletal and nervous systems, coupled with ever-changing environmental demands, lead behavioral repertoires into stable response modes. To be maximally adaptive, movements should not stay fixed in any one mode, but must be able to rapidly transition to new stable modes of organization (Kelso, 1995; Port and Van Gelder, 1995; Riley and Turvey, 2002; Van Orden et al., 2003; Halley and Winkler, 2008). These transitions are the hallmark of complexity, expressed as short- and long-term dependencies in movement stability and instability.

The complexity exhibited in motor control also sheds new light on the influences of cognitive demand during processing tasks, an issue that is pertinent to deception. Despite the paucity of examples that can be drawn from the deception literature, this is offset by the extensive research involving the self-organization of postural control under dual-task conditions. The dual-task context is similar in form to deception, where one is trying to balance both what is true and what is a lie. In these postural dual-task designs, intentions and cognitive demands act to shape behavior in meaningful, albeit subtle ways. In a typical set-up, participants attempt to maintain an upright stance while performing cognitive tasks presented visually or auditorily, and that can vary in attentional and processing demands. The resulting outcomes suggest that there is no one-to-one correspondence between the cognitive constraints and how movements are expressed, such as saying that increased task difficulty leads to degraded movements (Riley et al., 2005; Fraizer and Mitra, 2008). Even when attentional resources are heavily drawn upon, the behavioral system does not necessarily break down, as would be the case if cognitive and motor processes were separate components competing for a limited pool of resources (e.g., as proposed in *limited capacity* theories, see Woollacott and Shumway-Cook, 2002; Schmidt, 2003; Schmidt and Lee, 2005, for review). Rather, because these cognitive and motor processes are tightly coupled, new solutions

¹This study was originally published as a proceeding article for the Cognitive Science Society. Face data results were not included in the original report.

as to how to optimally redistribute resources are more quickly realized and expressed. Put simply, the cognitive system is not just breaking down or being overwhelmed, but is *reorganizing dynamically* in response to a new situation. How this might be relevant for deception is considered next.

ADAPTIVE RESPONDING DURING DECEPTION

Deception makes heavy demands on cognitive resources (see Vrij et al., 2011 for discussion). The truth also seems to be spontaneously activated with a lie, requiring additional effort to overcome (Osman et al., 2009; Duran et al., 2010). It is thought that performing concurrent tasks with deception, such as controlling one's body movements, will leave fewer resources available for successful deceptive performances (Leal et al., 2008). With less to work with, the movements of deceivers will become impaired in some way, whether it is an overall decrease in animation or overly controlled movements that appear rigid and unnatural (Zuckerman et al., 1981; Vrij et al., 1996; DePaulo and Friedman, 1998). However, from a dynamical systems perspective, this impairment interpretation does not necessarily reflect how the cognitive and motor systems are actually operating. Instead, the contextually and socially rich environment in which deception occurs provides a myriad of constraints that allow for the adaptive and functional reorganization of movement.

This view is inspired by Interpersonal Deception Theory (IDT), in which emphasis is placed on deceivers' ability to adapt within real-time interaction (Buller and Burgoon, 1996; Burgoon, 2005; Burgoon and Qin, 2006). Here, intentional and motivational factors allow deceivers to better regulate their behavior, doing so in a way that is highly responsive to their communication partner. According to this account, and the account considered here, deceptive displays of movement may not be driven by limited cognitive resources *per se* (i.e., impairment), but by the larger context. There is an important caveat however, in that IDT claims that resulting movements are largely under strategic control. We remain agnostic on this conclusion. Rather, our focus is on the reorganization of underlying "micro-behaviors" that are not intentionally controlled, and that may suggest a more subtle level of adaptivity. These movements are a non-conscious consequence of being on the ready in a situation that requires quick thinking and responsiveness in averting suspicion or detection. Finding greater complexity in the deceptive movements would support such a claim. Of course, if deceptive behavior has less complexity than honest behavior, doubt would be cast on our hypothesis and support would be lent to the impairment position. By adopting a dynamical systems approach, we can test these predictions.

We employed two measures used in the motor control literature, as well as the cognitive sciences more broadly. These two measures, recurrence quantification analysis (RQA) and multi-scale entropy analysis (MSE), provide complementary insights into the structure (as opposed to the amount) of variability exhibited in motor behavior. They do so by quantifying patterns of stability and complexity of body movement, expressed as time series of marker positions in a motion capture system. In the sections that follow, we first turn to a more detailed, albeit introductory,

tutorial of the conceptual and technical underpinnings of RQA and MSE (section "Quantifying the Structure in Time"). In the section "Extending an Analysis of Spontaneous Deception," we outline the methodology from Eapen et al. (2010), and detail our analytical approach for reinterpreting the collected data, targeting the undifferentiated movements of the arms, head, and upper face. To draw distinctions between deceptive and truthful behavior, we then contrast a displacement measure of movement (a traditional summary approach) with the RQA and MSE results (section "Results and Interpretation"). Finally, we return to the theoretical and diagnostic potential of the current research in the discussion (section "Discussion").

QUANTIFYING THE STRUCTURE IN TIME

Human cognition is driven by many factors, all of which must work together in a coherent, integrated fashion. This multiscale characteristic is a hallmark of a complex, dynamical system. In such systems, subtle fluctuations of behavior may reveal transitions between stable behaviors, strategies, or states. If a system transitions frequently, this may reflect the buildup and breakdown of constraints over system elements as new potentials for movement are formed. Sticking to a single strategy will work against an individual when vigilance is required. These frequent transitions between strategies or states, then, maximize the potential for adaptive responding. To capture this underlying stability and complexity, a number of non-linear measures have been developed to quantify these properties (Seely and Macklem, 2004; Dale et al., 2011).

The first of the two measures employed here, RQA, makes use of a method called "phase-space reconstruction" to capture geometric properties of how a system evolves in time (Eckmann et al., 1987; Webber and Zbilut, 1994; Marwan et al., 2007). As will be explained below, a measure of stability can be derived based on how often a system revisits various regions within its phase space. In essence, more visits to the same region of phase space represents greater stability. The second measure, MSE, provides an assessment of system complexity as variation in sequences of observations in a time series, measured across different temporal window sizes (Costa et al., 2005; Gao et al., 2007). Rather than phase-space reconstruction, this measure is based on *sample entropy*, which is computed over coarse-grained versions of the original series. The result offers insights into meaningful complexity, where less complexity is a system with too few or excessive transitions across stable states, and is either locked into a limited number of behavioral repertoires, or devolves into stochastic noise. An example of a system with less complexity can be seen in the movements of young children who are first learning to walk (Newell, 1998). Their movements are often rigidly fixed or seemingly random, both conditions that suggest a lack of motor control in adapting to changing situational demands. Taken together, RQA and MSE may serve as powerful new tools for assessing non-linear changes in movement. In the next section, we flesh out the details of these methods in simple, qualitative terms².

²For a more technical treatment of each approach, we recommend Riley and Van Orden (2005), Dale et al. (2011), and Marwan et al. (2007) for RQA, and Costa et al. (2005) for MSE.

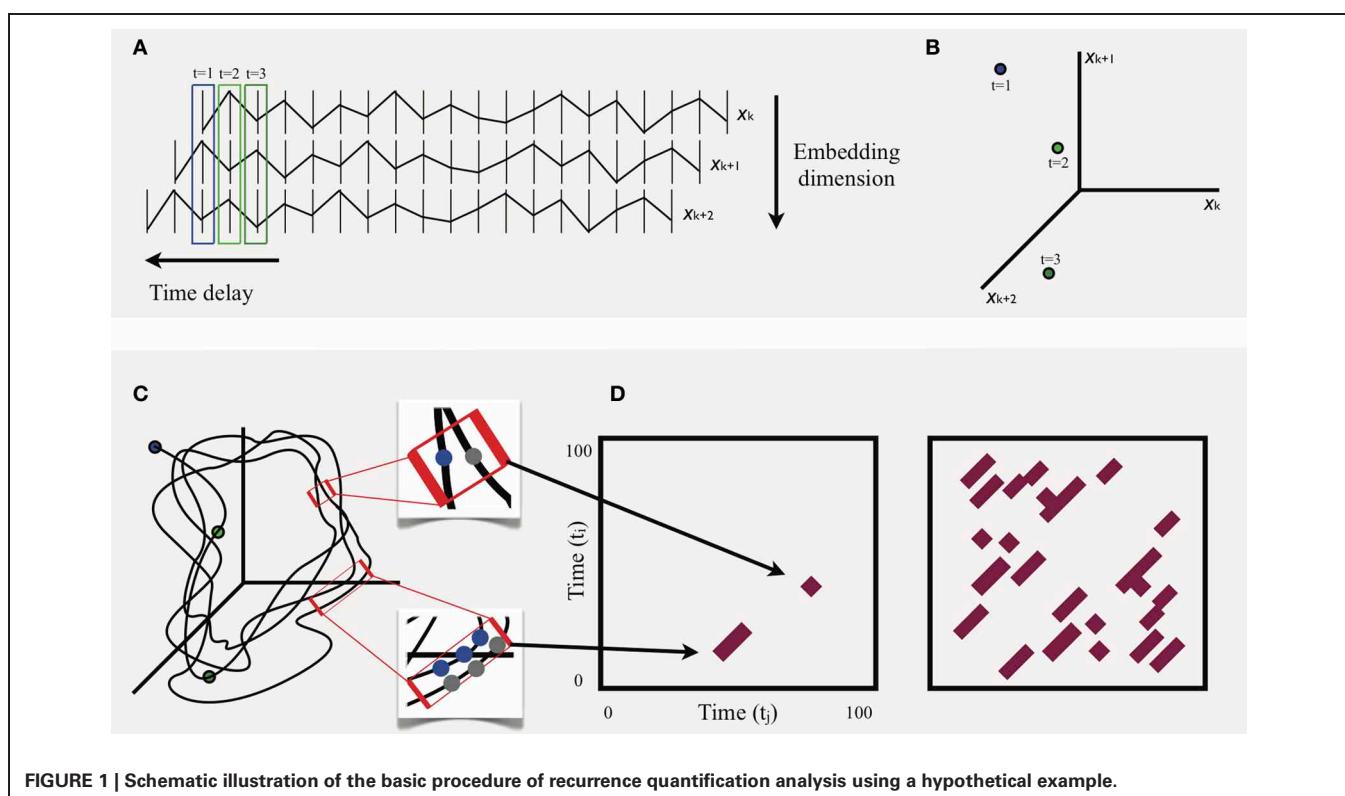
RECURRENCE QUANTIFICATION ANALYSIS

As already touched upon, the idea of phase space is critical to RQA. It is worth carefully explaining the concept of a “phase space,” and how it is reconstructed from a time series. A phase space is defined by the variables (i.e., dimensions) that govern a dynamical system. For example, velocity and angle of the arms are necessary variables in explaining movement coordination, just as temperature and pressure are necessary variables for defining a thermodynamic system. Because these variables are time varying and directional, temporal succession over them produces a “behavioral trajectory” in a system’s phase space. By examining the shape of the trajectory, it is possible to identify dynamic stabilities and instabilities as they emerge. One problem with this approach is that many state variables are unknown or cannot be measured. Another problem is the need to perform complex mathematics over a set of differential equations (e.g., integrating velocity vectors associated with state variables). To compensate, a solution is to reconstruct a phase space from time-lagged copies of a single time series of behavioral change. As originally observed by Takens (1981), a single state variable will be tightly coupled with all other state variables and thus is able to “stand in” for those that are unknown (Marwan, 2003; Stephen et al., 2009). Once plotted in high dimensional space, these surrogate variables are able to estimate the topography of system organization. Put simply, by analyzing just one behavioral time series, we can “reconstruct” the phase space.

Figure 1 provides an illustrative example of phase space reconstruction, as well as how RQA makes use of this space to derive

measures that describe a system’s behavior. To begin, in (A), a univariate time series of movement fluctuation, x_k , is shifted by any number of time steps (horizontal bars) to produce new *time-delayed* copies, x_{k+1} and x_{k+2} , of the original series. The number of copies (i.e., *embedding dimensions*) is inferred to be the number of dimensions in which the system is really operating. These are limited to three for current purposes. The resulting vectors are then plotted in temporal order, with the first three time points, enclosed in colored boxes, plotted in (B), and with all hypothetical points plotted in (C). The result is a phase space trajectory that, from visual inspection, tends to pass through regions previously visited at earlier points in time. It is the proximity of these recurrent points that is crucial to RQA. Recurrent points, particularly sequences of recurrent points, indicate that the system is in a preferred region of its state space, i.e., an attractor. In the top inset of (C), the Euclidean distance between two points, say at $t_i = 45$ and $t_j = 85$, fall within a predetermined *threshold radius* that defines a narrow region of space. When this occurs, it is simply plotted in what is known as a *recurrence plot*, shown in (D; left panel). Using the same logic, sequences of points that fall within the threshold radius are also captured: bottom inset of (C). Thus, the corresponding diagonal in (D; left panel) can be interpreted as follows: the system at time points; $t_j = 49, t_j = 50, t_j = 51$, is also where the system was at points; $t_i = 22, t_i = 23, t_i = 24$; a stable region.

A complete (albeit hypothetical) recurrence plot is shown in (D; right panel). Properties of this plot provide the basis for all RQA measures. Here, we focus on just two: *percent recurrence* and *determinism*. The first is simply the percentage of filled points



given the number of possible points, calculated according to the equation³,

$$RR = \frac{1}{N^2} \sum_{i,j=1}^N R_{i,j},$$

that counts all points between the two time series, (i, j) , that fall within a predetermined radius. The latter, determinism, is the percentage of points that fall on diagonal lines, where diagonal lines indicate continuous sequences of repeating movements at different time points⁴. This is computed as a ratio between diagonal sequences and overall recurrence,

$$DET = \frac{\sum_{l=l_{\min}}^N IP(l)}{\sum_{i,j}^N R_{i,j}},$$

where $P(l) = \{l_i; i = 1, \dots, N_l\}$ is the frequency distribution of all lengths of diagonal lines. Determinism is thus derived from basic recurrence, and is especially relevant for the current study. Specifically, it provides an intuitive measure of overall movement stability. However, as discussed earlier, determinism does not necessarily have a straightforward correspondence with system complexity. Movements that are highly predictable, occurring at regular, unchanging intervals, will exhibit high determinism, but are not complex. Likewise, movements characterized by random noise will show low determinism, but again are void of meaningful complexity. To identify what is meaningful, a suite of entropy-based measures has been developed that are based

³See the excellent resource <http://www.recurrence-plot.tk/rqa.php> by Norbert Marwan for these and other quantifications.

⁴RQA also produces 11 additional measures that capture further dynamical properties of the recurrence plots, such as averaged diagonal length and length of the longest diagonal line. These measures may provide new directions for analysis, but for current purposes of examining general stability, we focus on a parsimonious set of variables.

on the degree of repetitiveness in a time series. One measure in particular, MSE, provides a powerful technique for assessing complexity over multiple spatiotemporal scales in a single series, a method we turn to next⁵.

MULTISCALE ENTROPY

MSE is a two-step process, with the first step being the computation of sample entropy over a univariate time series. As previously stated, sample entropy is a measure of regularity, and captures, as Richman and Moorman (2000) observe, “the rate generation of new information.” This new information is related to the degree to which sequences of some length (m) in a time series remain similar after the sequence length is extended by an additional time point ($m + 1$). **Figure 2**, adapted from Costa et al. (2005), is presented to help conceptually ground what is meant by the given definition. A relevant pattern constitutes a short sequence of consecutive points, represented here as sequences of two points. This pattern is tallied as it repeats in the time series. For example, the consecutive values at $t = 2$ and $t = 3$ are a candidate pattern of interest (enclosed by box), and can be seen to repeat starting at $t = 10$ and at $t = 27$, as they occur within a similar range (or *threshold radius*; designated by horizontal dashed lines). This brings the total tally count to three. What needs to be determined is whether these two-point sequences can be extended by a similar, consecutive point. Returning to the original pattern in **Figure 2**, this value corresponds to $t = 5$ (marked by red arrow), and is only extendable at the $t = 28$ location (marked by

⁵It should be noted that RQA also produces an entropy measure based on recurrence plots. This measure is derived from the number of diagonal lines of different lengths, with a greater number indicating greater entropy. However, results can sometimes be difficult to interpret if long diagonal lines are present with many smaller lines. Such a system would be considered highly entropic, yet the presence of long diagonals indicates high stability. The MSE measure allows for a more straightforward interpretation of entropy and complexity. Furthermore, by turning to a measure outside of RQA, we can ensure that the observed patterns are not limited to the RQA-based analysis.

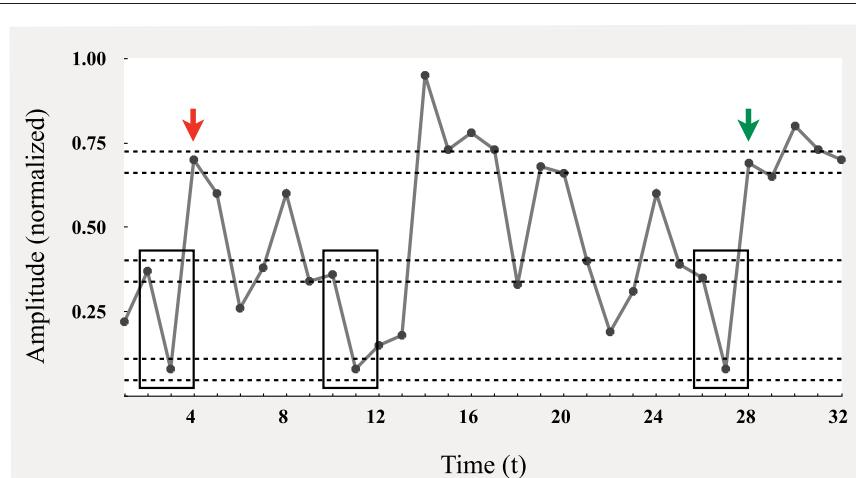


FIGURE 2 | Schematic illustration of the procedure for computing sample entropy [adapted from Costa et al. (2005)].

green arrow), resulting in a tally of two three-point sequences. After repeating this process over all possible patterns, the natural log of the ratio between the final two-point and three-point tallies is computed. The result is sample entropy (a conditional probability), where greater values indicate that there are more two-point sequence patterns that cannot be extended by a similar third point; thus, there are a greater number of unique patterns, i.e., more information, greater complexity, and less regularity.

Although not immediately obvious, this measure has a fundamental problem in that higher entropy values also scale with increasing amounts of random noise (Costa et al., 2005). In other words, if there is less repetitiveness in a signal, it may not necessarily be due to complexity. One way to solve this problem is to evaluate how sample entropy changes over various spatiotemporal scales of the time series. Motor behavior is composed of a number of interacting elements that must come together to perform a task. Although these elements are closely bound and depend on each other for expression, each has its own intrinsic frequency that, when combined, produce organized structure across multiple spatiotemporal scales. The reader may ask: "What elements, what scales?" The relevant ones could be the various structures (head, torso, arms, etc.), cognitive processes (e.g., memory, language, etc.), and even finer-grained scales of neural organization. It is obvious that any organized cognitive performance, such as deception, is grounded in such an array of elements and processes. Yet, even without making any commitments about the physical or cognitive constraints on the system, this coherent self-organization is a fundamental characteristic of a dynamical process (Bar-Yam, 2004). Thus, a complex system reveals new information (complexity) across scales of decreasing frequency, whereas a random signal (void of underlying element interactions) will show less and less new information.

To produce a range of scales, the second step of MSE, the original time series is divided into non-overlapping windows of

increasing sizes (i.e., coarse-graining). The values in each window are then averaged and replotted as a new point in a reduced series, producing a new time series, calculated by the following equation

$$y_j^{(\tau)} = 1/\tau \sum_{i=(j-1)\tau+1}^{j\tau} x_i, \quad 1 \leq j \leq N/\tau.$$

Here, the original time series, X_1, \dots, X_N , is divided into non-overlapping windows of length τ , with the datapoints in each window averaged to produce $y_j^{(\tau)}$. An example of this process is shown in **Figure 3** with an original time series of x_1, \dots, x_{12} that is reduced by a scale of 2 ($\tau = 2$), to y_1, \dots, y_6 , and then by a scale of 3 ($\tau = 3$), to z_1, \dots, z_4 . In actual time series, which are comprised of thousands of points, reduction continues to a scale of 9 ($\tau = 9$). These resulting scales correspond to signals of lower and lower frequencies. Finally, sample entropy is computed for each new reduced series and plotted with scale increasing along the x-axis (**Figure 3B**). The resulting curves are then used to compare relative differences between groups, an issue we return to when comparing deceptive and truthful movements in the following section.

EXTENDING AN ANALYSIS OF SPONTANEOUS DECEPTION OVERVIEW OF EAPEN ET AL. (2010)

To apply these dynamical techniques to deception, data captured during an interaction between a participant and two experimenters are explored here⁶. To ensure recordings were of natural spontaneous behavior, participants were told their behaviors would be captured while they took part in a study supposedly examining the relationship between mathematical ability and

⁶This experiment was conducted under the permission of the UCL Research Ethics Committee.

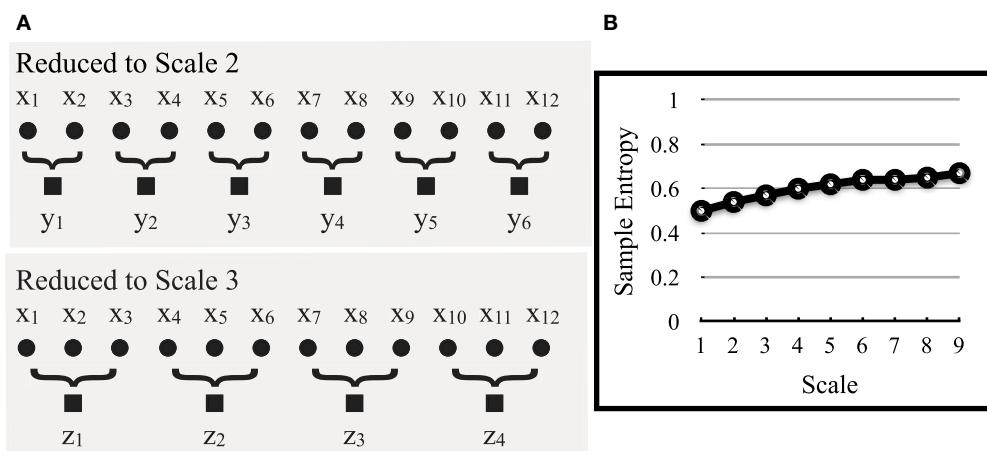


FIGURE 3 | In (A), the original time series, x_{1-12} (scale 1), is reduced by a lower-order scale to produce new time series, y_{1-6} (scale 2) and z_{1-4} (scale 3). Although not shown, this continues to scale 9.

In (B), sample entropy is computed for these new lower frequency time series and plotted as a function of scale, from 1 to 9 [adapted from Costa et al. (2005)].

body sway. In reality, two critical recording periods were captured when the experiment was apparently at an end: one regarding their performance on a math test and the other regarding an accident they witnessed.

An amiable female experimenter welcomed participants. Soon after, a male experimenter entered and acted in a cold and unpleasant manner⁷. The male experimenter placed a laptop on the edge of a table and told the female experimenter, “I’ve got that report of yours on my laptop. Remind me about it at the end.” Participants donned a body motion tracking shirt and hat and were calibrated before being seated at a computer to take part in a math test. The test consisted of two stages of 30 multiplication questions with three multiple choices. Pilot testing indicated people scored ~75% correct.

After the first stage, the male experimenter excused himself while the female experimenter explained what the second stage would entail. She told them what we had found and hoped to continue to find was that standing improves math ability, purposely violating good experimental practice to give the impression that it was normative to perform well on the second stage. In addition, participants were offered £5 if they performed better. They were also told that since they were standing they would be unable to reach the keyboard, so it was also their task to mentally keep track of approximately how many they calculated correctly, but not to voice this. That is, they were encouraged to claim they performed better on the second stage and they were aware there was no way to verify their claim. At this point the female experimenter accidentally knocked the laptop to the floor. She quickly expressed relief saying, “Thank God the cameras were off,” implying that only she and the participant were witnesses to the accident.

The second block was initiated as the male experimenter re-entered the room. The block was designed to become increasingly difficult over time, such that the absolute difference between the three multiple choices was smaller on all trials in comparison to the first stage and that the time to respond was gradually reduced with each successive trial. All participants in a norming test performed worse on the second stage.

After completing the math test, participants were asked a baseline question (“Did you feel the second stage took more or less time to complete?”) and a critical question (“Did you feel you performed better on the first or the second test?”). The responses to these two questions, from the onset of their reply, constitute the neutral and critical recording periods for the math test. Participants who claimed to have performed better were paid the additional £5. Participants were then thanked for taking part and asked to remain in the kit while the male experimenter

⁷A reviewer raised the interesting point that had we used different gender roles, our results would have been quite different, citing Wraga et al. (2006) as support. Although this is an intriguing possibility, our aim was to set up a social situation that draws upon social norms about lying and honesty, and correct behavior between participants and experimenter. The goal was to rely upon these schemas of social interaction to elicit a higher rate of spontaneous deception. Had we used other gender roles in doing so, we might expect the rates of deception to decrease. Nevertheless, we believe that the roles used here adhere to reasonable expectations about social interaction and are optimized for the current research question.

took a backup of the data onto his laptop. During this time, the neutral (“Did the math experiment run ok?”) and critical laptop-accident questions (“My computer doesn’t seem to be working. Did you see anything happen?”) were posed to the participant and recorded.

CAPTURING MOVEMENT

A Vicon Nexus body motion tracker captured three-dimensional movement at 200 Hz by recording near-infrared reflections from 20 plastic markers attached to a tight-fitting shirt and cap. An additional nine markers were attached around the face, on the back of each hand and on the tips of each index finger. Marker positions were captured with an accuracy of 0.1 mm in terms of position in space (Figure 4).

MOVEMENT DISPLACEMENT

We focus here on undifferentiated movements of the arms, head, and upper face. These regions have been targeted in deception research as being especially relevant for detection purposes (Ekman and Friesen, 1969, 1972; Vrij et al., 1996, 1997; Hill and Craig, 2002; DePaulo et al., 2003; Jensen et al., 2010; Hurley and Frank, 2011). In the majority of these previous studies, participants are asked to rate the frequency, duration, or functional purpose of the movements, such as whether the movement has communicative intent (e.g., gestures used to emphasize verbal statements) or is unintentional (e.g., a “leakage” cue flashed across the face). In the current work, we avoid the assumptions needed to make these distinctions, evaluating only the rhythmic sequences of movement over time.

As mentioned, the output of the motion tracker system is in three-dimensional coordinate positions across multiple body markers; and as such, we need to convert position to a single-dimensional measure of movement displacement. To begin, we first averaged the three-dimensional coordinate positions of body markers within each region of interest. For the arms, this includes six points distributed across right/left forearms, hands, and wrists; for the head, five points distributed across the top, right/left, and back/front; and for the face, five points distributed across the eyes and nose, thus minimizing influences from speech articulation.

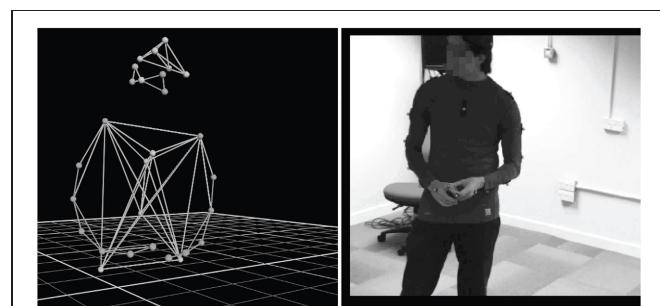


FIGURE 4 | Marker placement for body, head, and face, reconstructed with an accuracy of 0.1 mm using Vicon Nexus motion tracking software.

Averaging produces a single vector of coordinate positions for each region. Change in movement displacement was computed over windows of 250 ms, equivalent to 20 time steps (based on a sampling rate of 200 Hz). For arms and head, this was done by averaging the Euclidean distances between contiguous (x , y , z) coordinate positions in the moving window. A sample time series is shown in **Figure 5**. For the face, a slight modification was made based on the observation that movements of the face will co-vary with movements of the head. To remove this influence, Euclidean distances were computed between each face point and a composite head position, and then averaged in the moving window of 20 time steps.

PARAMETER SELECTION

The generated displacement time series were normalized (mean zero and standard deviation of one) and used for the RQA and MSE analyses. It should be noted that although the movements here differ from those typically used in the motor control literature, they are still amenable to non-linear analyses and interpretation. Various types of movements have been assessed using a similar approach; for example, changes in the angular velocity of hand movements (Stephen et al., 2009), and movement displacement in the video recordings of facial/head movements (D'Mello, 2011). The main requirement for these analyses is a movement signal that is thought to be generated by a complex

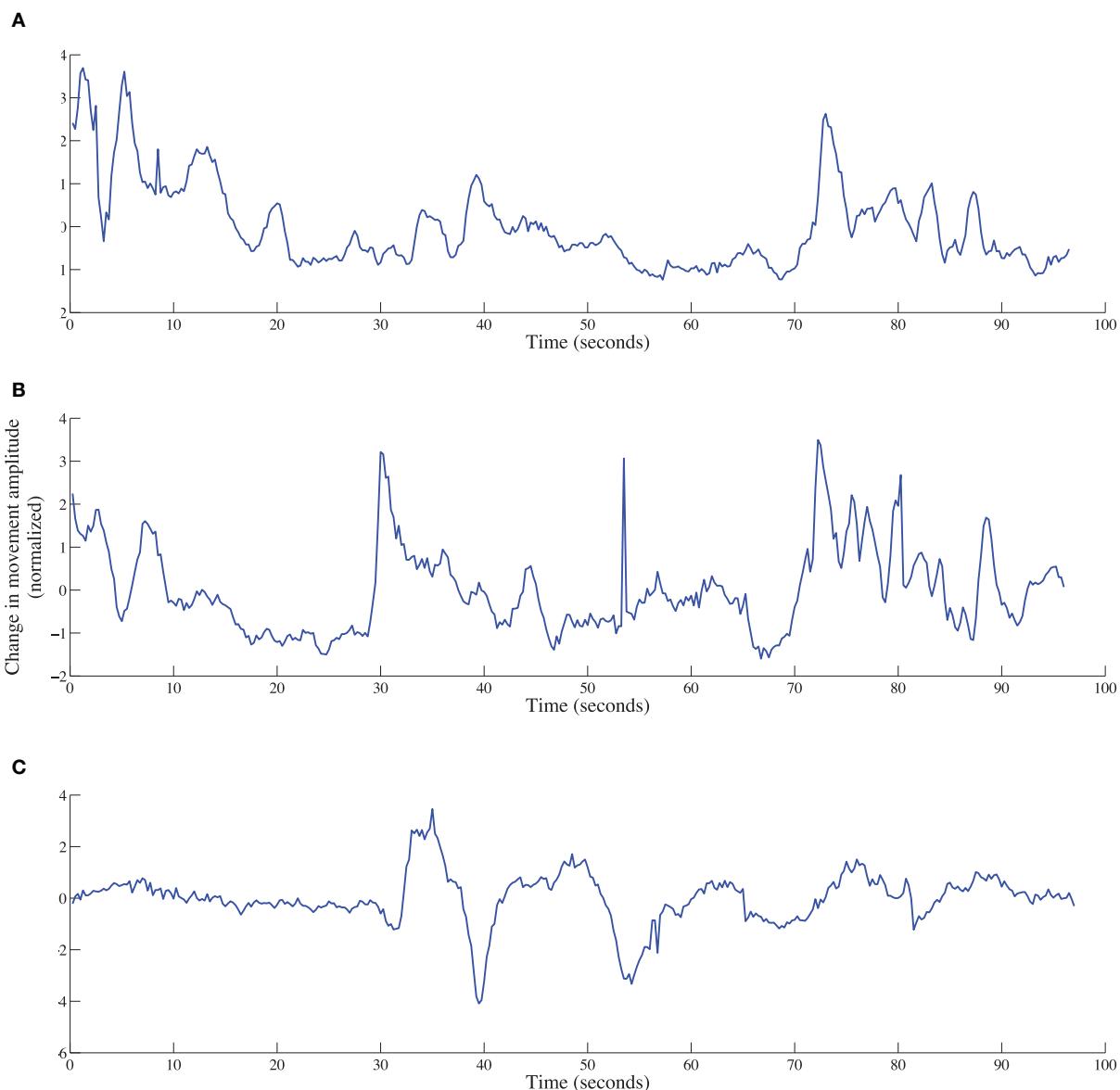


FIGURE 5 | Time series of movement displacement (based on Euclidean distance) for arms (A), head (B), and upper face (C) for a deceptive responder in the math-test condition.

system. However, the parameters for RQA and MSE still need to be uniquely specified for signal source in order to avoid spurious or unaccounted structure.

For RQA, the critical parameters correspond to time delay, embedding dimension, and radius for determining whether two points in phase space are sufficiently close (with radius expressed as a percentage of the standard deviation of a normalized time series). Following Shockley (2005) and Shockley et al. (2003), we selected parameter values by first conducting RQA on four randomly selected time series across multiple embedding dimensions, along a range of delay and radius parameter values. Using a surface plot, we plotted the recurrence rate (y-axis) from each analysis, for each embedding dimension, as a function of delay (x-axis) and radius (z-axis). This produces multiple three-dimensional landscapes of valleys and peaks corresponding to recurrence rates that rise or fall depending on parameter value combinations. The optimal parameters are those that are in the flat regions of each series landscape, thus ensuring that the values are stable and not reflecting idiosyncratic change (i.e., small increases or decreases in the selected embedding dimension, time delay, and radius would have little effect on recurrence rates). It is also typical to select values that produce an overall recurrence percentage around 5% and that avoid ceiling effects in determinism. As such, we settled on an embedding dimension of three, a delay of eight, and radius of 15% for all analyses⁸.

For MSE, parameter selection is more straightforward. Here, we followed the precedent of Costa et al. (2005) in setting the parameters corresponding to sample entropy and coarse-graining. As described in the previous section, we began with two-point sequences that were extended by a third point. We also used a threshold radius of 15%, which like RQA, sets the boundary of whether time points are considered similar, and is expressed as a percentage of time series standard deviation. Coarse-grained versions of the original series, in which sample entropy was computed, were reduced by a factor of 2–9 (retaining the original series with a factor of one). This is depicted in **Figure 3**⁹.

PARTICIPANTS

Data from 28 participants were analyzed in this study (18 females and 10 males, mean age 22.5 years old). Most participants were consistent in how they responded between the math-test and laptop-accident conditions, either lying in both or telling the truth in both. However, six participants split their responses between conditions, telling a lie in one and the truth in another. Also, due to some data loss with the Vicon motion tracking system, movements for six participants were unavailable in the accident condition and unavailable for one participant in the math-test condition. In the end, for all analyses, there were 26 deceptive time series (combined across the math-test and laptop-accident conditions; 16 participants; 3 males and 13 females),

⁸The “max norm” method was also used to compute distance between vectors in the reconstructed phase space (Marwan, 2003). Shockley (2005) offers an excellent summary of these issues, and is available as an open access chapter online here: www.nsf.gov/sbe/bcs/pac/nmbs/chap4.pdf

⁹In general, the setting of these specific parameters does not adversely affect the general pattern of results, which hold across a range of these values.

and 21 truthful time series (combined across the math-test and laptop-accident conditions, 17 participants; 5 males and 12 females).

DATA PREPARATION

Responses in the math-test and laptop-accident conditions were combined for all analyses. This combination was done partly for purposes of generalizability, as the structure of movements associated with deception should be somewhat consistent across similar contexts, thus bolstering claims of detectability. The other reason is more pragmatic, as limitations in statistical power for the RQA and MSE analyses warranted combination. This is often a consequence of using previously collected datasets, particularly sets that involve naturalistic, and somewhat noisy, expressions of behavior. As such, our claims are somewhat limited (an issue we address in the Discussion), but nevertheless, the goals of introducing non-linear measures to the deception literature and relating these measures to the underlying cognitive processes involved in deception are still intact. It should be noted, however, that the pattern of results presented here in fact holds in each case of deception separately.

STATISTICAL APPROACH

For the displacement and RQA determinism results, differences between deception and truth, across neutral and critical questions, were analyzed using linear mixed effects models. Given that participants sometimes contributed to both or only one of the deceptive responses across conditions, participant and condition variables were entered as random factors in the model to control for associated random variance. Also, because the error term in this model class is not amenable to traditional *F*-test methods for computing a *p*-statistic, an MCMC method was instead used for estimating statistical significance (see Pinheiro and Bates, 2000; Baayen et al., 2008). Next, for MSE curves, differences between relevant groups were analyzed by generating intercept and slope coefficients for each participant’s time series data, using a curve-fitting model with linear fit. The resulting coefficient terms were then compared across deceptive and true responses using a two-sample *t*-test.

RESULTS AND INTERPRETATION

In this section, we begin with the results of movement displacement, an aggregate measure of magnitude change that has traditionally been used in analytic approaches that average over time series. We then turn to our two non-linear measures, RQA and MSE, that may be useful in capturing additional information about movement dynamics.

DISPLACEMENT RESULTS

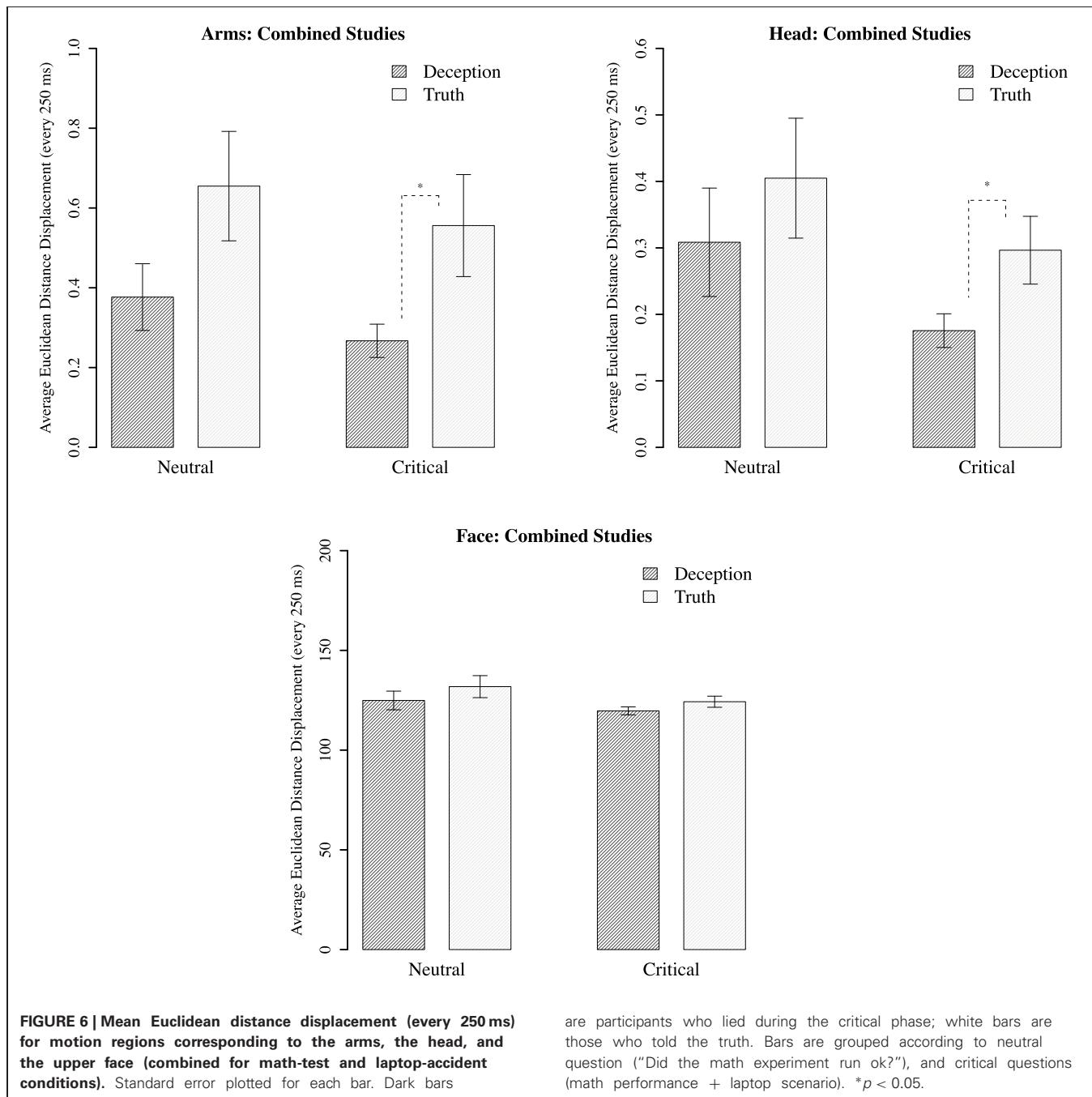
Separate analyses were conducted on the arms, head, and upper face regions¹⁰. In comparing deception with truth, the neutral

¹⁰For these and subsequent analyses, the total N for each comparison varied slightly between body regions due to dropped recordings with the Vicon motion tracking system. For arms, there were 26 deceptive and 20 truth time series; for head, there were 23 deceptive and 21 truth time series; and for face, there were 25 deceptive and 20 truth time series.

questions showed no statistically significant differences across all three motion regions. However, for critical questions, the movements of the arms and head reveal significantly less displacement in deception than the truth; for arms, $B = 0.264$, $p = 0.022$; for head, $B = 0.121$, $p = 0.038$. There are no statistically significant differences in displacement for face movements. And for all regions, there were no significant differences between neutral and critical questions for deception or truth (see **Figure 6**).

For critical questions, we replicated the basic effect found by Eapen et al. (2010), who found less movement for deception

across all motion points. Here, using a slightly different operationalization of displacement, decreases were isolated to the arms and head. This finding may suggest that participants are seeking to minimize incriminating behaviors by clamping down on their movements. Conversely, the null finding for the face suggests that the generated movements are much more subtle and spontaneous, and the same control exhibited over the arms and head is not possible. But this may be because the wrong level of movement has been examined, leaving open the possibility that non-linear measures offer a more sensitive means of identifying differences between conditions.



Another issue that is evident from **Figure 6** is the lack of significant differences between the neutral and critical questions. Yet the direction of mean values for neutral questions is very similar to that of the critical. Given that the neutral questions always preceded the critical in the experimental setup, participants who cheated on the math test or who were witnesses to the experimenter dropping a computer, may anticipate that a follow-up question will be asked that requires deception (such as being asked about their performance or why the computer was broken). Thus, their response behavior during the neutral question may indicate a preparation to lie that is ultimately expressed when a deceptive response is required. Whether the behavioral system was poised to react in this way is difficult to interpret from movement magnitude alone. Again, non-linear measures may prove useful in clarifying this issue.

RECURRENCE QUANTIFICATION ANALYSIS RESULTS

For each motion region of interest, measures of percentage recurrence, and determinism were generated based on recurrence plots for deceptive and true responses (**Figure 7**). The recurrence rate for all analyses were within 4–8%, and did not differ between comparisons of deception vs. truth, or neutral vs. critical questions. However, determinism rate did show statistically significant differences between groups, most notably in upper face movements, with less determinism in deception than in the truth, $B = 0.126$, $p < 0.05$ (**Figure 8**). There was also marginally less

determinism in deception with arm movements, $B = 0.135$, $p = 0.09$; but for head movements, no statistically significant differences were found. There were also no significant differences within neutral questions, and in comparison with the critical questions.

The trend for all regions is for less determinism for the critical questions during deception. This is most safely concluded for the upper face, with some cautious support for arm movements. Even so, this is suggestive that stability, as assessed by determinism, decreases in deception. Although it may be tempting to draw the conclusion that less movement causes a drop in determinism, the results of the upper face indicate otherwise, as no differences were found with displacement (based on the previous analysis). In other words, movement displacement appears to be independent of the influences driving determinism. That is, the non-linear dynamics of the motion reveals new detail about the act of deception that is unavailable to the oft-used frequency counts of more or less movement in prior research.

As with displacement, the pattern of determinism between deceptive and truthful responses was also similar for neutral and critical questions. That is, there were lowered levels of determinism when participants both anticipated and expressed a lie. However, although there is decreased determinism/stability, it is not necessarily characterized by meaningful complexity. Before considering what a decrease in stability might mean in a deceptive context, we interpret the results alongside the MSE analysis.

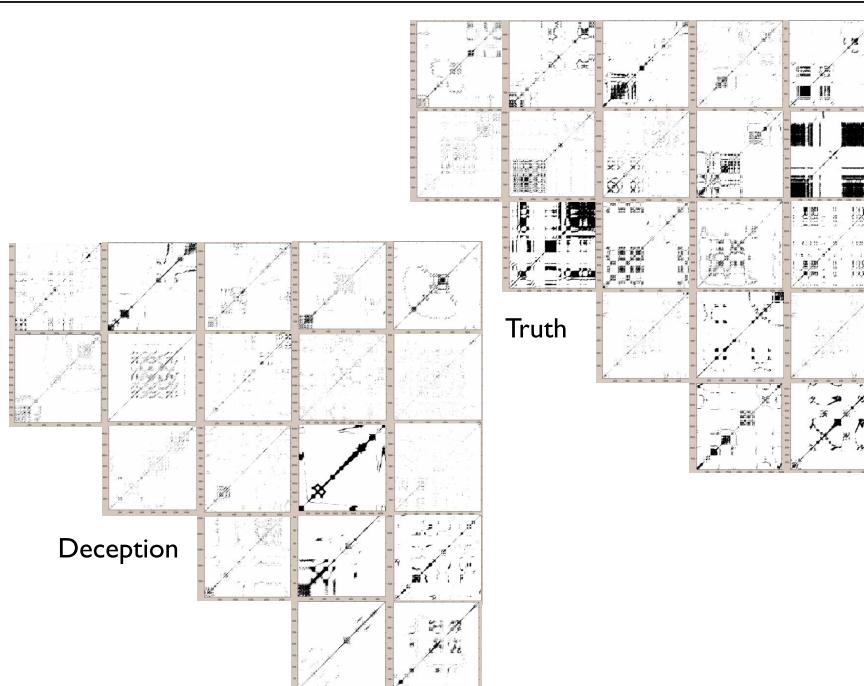
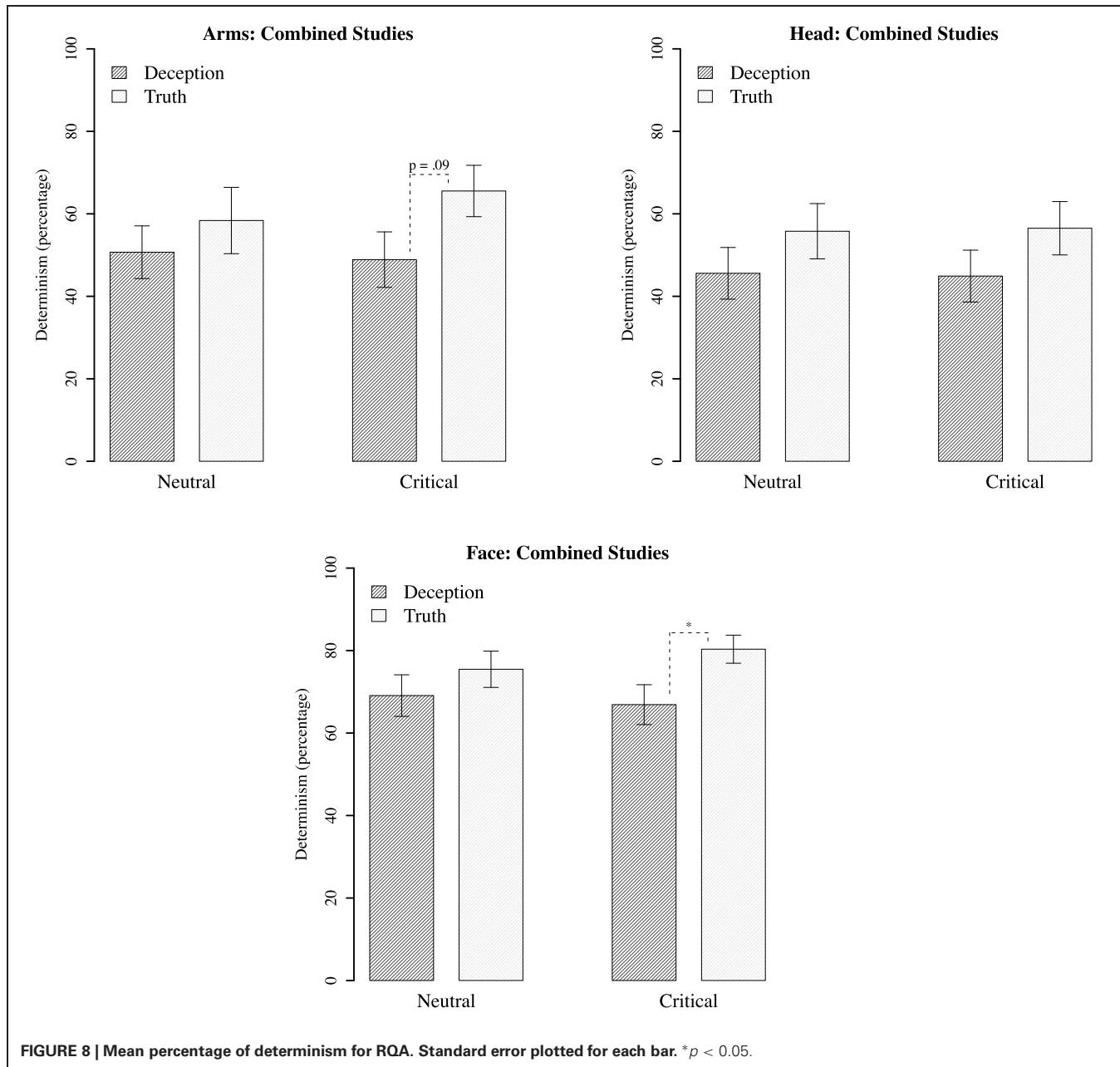


FIGURE 7 | For upper face movements, mosaic of recurrence plots for randomly selected subset of deceptive and truthful responses for critical questions. Deception is shown in the lower panel and truth in the upper panel. For truth, there is overall higher determinism than deception, as indicated by the greater percentage of recurrent diagonal lines. Each plot shown in this array is a reflection of the “recurrences”

of face movements over time; the more points there are, the more the time series of movements exhibits similar fluctuations. Glancing at the plots does reveal that Truth plots seem to have more dense appearance of recurrence structures (for details on method, see **Figure 1**). This is quantified using the Determinism percentage shown in **Figure 8**.

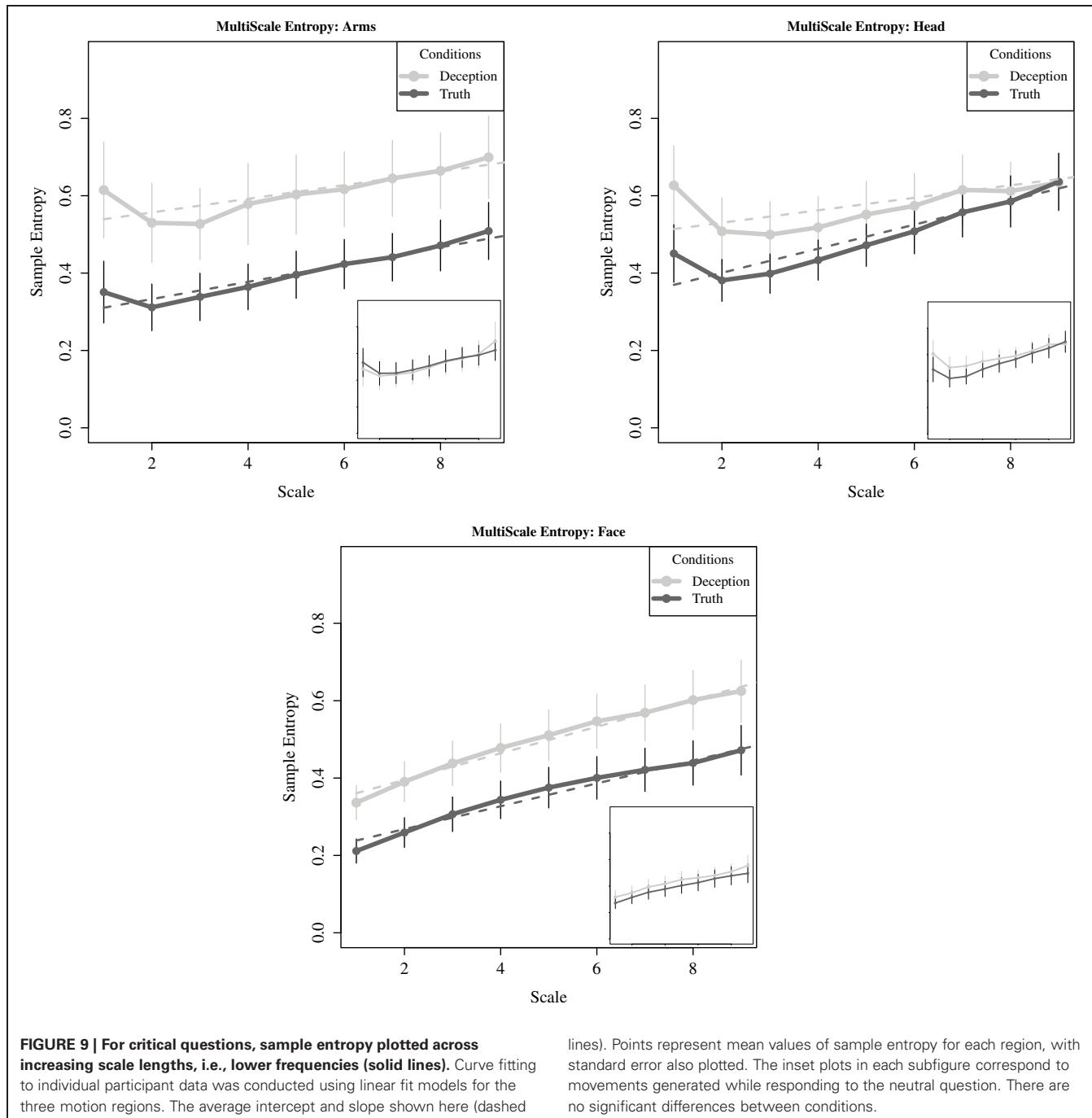


MULTISCALE ENTROPY ANALYSIS

As a reminder, MSE relies on sample entropy, a measure that evaluates the repetition of consecutive sequences in a time series (as opposed to variance). Sample entropy is then plotted over multiple time scales increasing in length, with time scales derived from the original movement time series. For each deceptive and truthful response, within each motion region, an MSE curve is generated and fitted with a linear model. To compare the relative complexity between groups, the resulting intercept coefficients for deceptive and truthful responses are evaluated using two-sample t -tests. In this way, differences across all scales can be evaluated in one statistic. The slope terms are also examined to compare differences in the rate by which

complexity increases over scales. Composite slopes are shown in **Figure 9**.

For the intercept coefficients, we found statistically significant differences with the movements of the upper face, $t_{(41)} = 1.976$, $p < 0.05$; and once again marginal statistical significance for the arms, $t_{(44)} = 1.654$, $p = 0.09$. There are no statistically significant differences for the head. Thus, the pattern for the upper face and the arms is for greater relative complexity with deception compared to the truth. Next, turning to the rate in which complexity increases for both deception and truth, there is equivalent gain for all regions except the head, where the complexity in the truth rises at a faster rate than deception, $t_{(42)} = 2.27$, $p < 0.05$. Here, truth and deception converge at the larger timescales, and may account



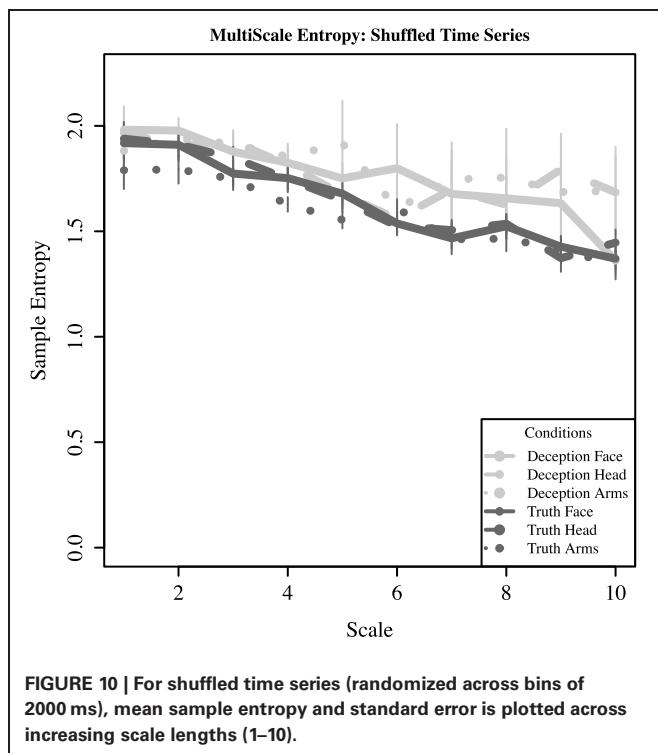
for the failure in finding significant differences between deception and truth. Finally, for neutral questions, complexity was present in the neutral responses, but as has been evident in the previous analyses, there were no differences with critical questions.

The findings of greater complexity in deception for the upper face (and somewhat for the arms), is further qualified when one examines what happens when the time series for each response is randomly shuffled while preserving local temporal interdependencies. Binned sequences of 2000 ms sequences were randomly shuffled, effectively removing the time-dependent complexity

hypothesized to be present in each series. Based on **Figure 10**, the monotonic downward slope indicates that the number of new structures drops as the length of the window for coarse-graining increases; thus, there is no new information to be found.

DISCUSSION

Despite a long tradition in seeking out bodily cues of deception, temporal dependencies in how movement is organized across time have largely been overlooked. In the current paper, we captured these dependencies as emergent properties of a complex



system, characterized by structural properties of stability and complexity. Using two non-linear measures, RQA and MSE, we found that the movements about the upper face, and somewhat in the arms, tend to have lower determinism/stability (based on RQA) and higher complexity (based on MSE). These patterns suggest greater flexibility in movement responsiveness that would have remained hidden with a measure of movement displacement alone, as deceptive and truthful facial movements were shown to have similar summary statistics (mean and standard error). Though suggestive, it is important to note that these results are indeed statistically subtle, based on a convenience sample, and also show that the neutral and critical contexts are about the same in most measures within each subject. However, if we take these results for granted, here we consider some potential theoretical implications of these dynamical methods.

These results challenge the notion that the demands introduced by deception exclusively deplete attentional resources and negatively affect the control of movement. That is, rather than only a breakdown in processing, the dynamic signatures of movement are structured in such a way to permit rapid adjustments to emerging demands unique to deceptive, social contexts. To support this claim, we have drawn from a dynamical systems framework for understanding how non-linear systems come to exhibit structured behavior. Human motor behavior is often held up as a primary example, in that patterns of movement are rapidly formed, maintained, and transformed by the release or restriction of system-wide degrees of freedom (Turvey, 1990, 2007; Newell, 1998). What results is increased complexity that speaks to the ability of the motor system to flexibly adjust and adapt to ever-changing situational demands, much like the behaviors of a skilled athlete or a child mastering the ability

walk. Such behavior may be necessary in handling the challenges inherent to deception.

Greater flexibility also appears to be present during the neutral questions prior to the actual deception. This finding may point to participants who anticipate that they will need to lie. Although they did not know that they would be put on the spot about their own guilty behaviors (assuming they cheated on the math test), or the guilty actions of another (witnessing a confederate drop a laptop), the possibility of investigative questioning by the experimenter, as well as the experimenter's possible suspicion, was always present. Such a situation would support an increased need for heightened responsiveness (i.e., adaptiveness, see Eapen et al., 2010). One reviewer remarked that this may instead be a sign of a sluggish system that is incapable of rapidly adapting to a more local context. Holding up the results from another perspective, this is a viable interpretation. But one timescale's sluggishness may be another timescale's adaptiveness. The way in which the dynamic signatures seem to be present (i.e., in both neutral and critical questions) suggests adaptiveness at a longer timescale; while this adaptiveness may force more local moments to be under the control of these longer timescales. In other words, the system could be adapting for a future potential event; and before it happens the situation at hand is subject to this structure.

It is also revealing that responsiveness was most apparent in the subtle movements of the upper face. The face has largely been implicated as a "dynamic canvas" for expressive behavior, where intentional and unintentional information about mental states are optimally conveyed (DePaulo, 1992; Rozin and Cohen, 2003). Given that accurate assessments of these states are easily and rapidly seized upon by outside observers (Ambady et al., 2000), it is sensible to hypothesize that these movements need to be particularly flexible in deceptive contexts. Also, unlike the movements of the body and head, the control of the musculature around the eyes may also produce a signal that is most appropriate for the non-linear analyses employed here. Both factors may explain why the reported results were statistically significant for the face alone.

The rapid and small-scale movements in the face are also thought to be susceptible to the inadvertent "leakage" of hidden emotional states (Hill and Craig, 2002; Ekman and Friesen, 2003). Such leakage forms the basis for the *inhibition hypothesis*, whereby attempts to conceal true emotions are revealed in "micro-expressions" of the face that last only tenths of a second (Ekman, 1992; Ekman and Friesen, 2003). Of the few empirical studies that directly examine this claim, evidence suggests that masked negative emotions may elicit the greatest leakage; and that transitory patterns of emotional states, particularly from negative to positive emotions, may also be a predictor of deception (Porter and ten Brinke, 2008; ten Brinke et al., 2011). For the current study, this raises the interesting possibility that the transitional nature of momentary emotional states can account for the current results. However, such transitions are much too coarse-grained to drive the moment-by-moment millisecond fluctuations that were analyzed. Also, given the short duration of participants' interactions with the experimenter, a wide array of changing emotional states is unlikely. Nevertheless, the role of emotions in the current

study cannot be discounted. The need to adapt emotional displays to changing circumstances may very well contribute to the increased movement complexity found during deception. Such questions pave a way for future work.

We were limited by certain characteristics of the data, such as participants that unevenly self-selected into deceptive and truthful response groups, and who sometimes lied in both or only one of the math-test and laptop-accident conditions. Statistical power concerns were also limiting, and required us to combine the math-test and laptop-accident conditions. There is also the inescapable fact that statistical effects were somewhat weak. Nevertheless, the upside of the current dataset is that we could draw conclusions from behavior that possesses defining characteristics of deception; that is, participants who deliberately attempted to mislead unsuspecting recipients (a rarity in laboratory-based studies). The dataset also allowed us to examine continuously sampled movements as fluctuations over time. Such data are quite rare in the deception literature, with the exception of a promising line of research that extracts continuous body movements from video recordings (Meservy et al., 2005; Jensen et al., 2010). Although this research uses participants who were instructed to lie and analyses were based on movement displacement alone, a number of these variables have proved to be highly effective in detecting deception. When entered into machine learning models, the classification algorithms produced surprisingly high accuracy rates. Given that we show dynamical measures provide information above and beyond movement displacement, these additional variables could further improve the accuracy of classification.

REFERENCES

- Ambady, N., Bernieri, F. J., and Richeson, J. A. (2000). Toward a histology of social behavior: judgmental accuracy from thin slices of the behavioral stream. *Adv. Exp. Soc. Psychol.* 32, 201–271.
- Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412.
- Bar-Yam, Y. (2004). Multiscale variety in complex systems. *Complexity* 9, 37–45.
- Bernstein, N. A. (1967). *The Coordination and Regulation of Movements*. New York, NY: Pergamon.
- Bond, C. F., and DePaulo, B. M. (2006). Accuracy of deception judgments. *Pers. Soc. Psychol. Rev.* 10, 214–234.
- Buller, D. B., and Burgoon, J. K. (1996). Interpersonal deception theory. *Commun. Theory* 6, 203–242.
- Burgoon, J. K. (2005). The future of motivated deception and its detection. *Commun. Yearb.* 29, 49–92.
- Burgoon, J. K., and Qin, T. (2006). The dynamic nature of deceptive verbal communication. *J. Lang. Soc. Psychol.* 25, 76–96.
- Costa, M., Goldberger, A. L., and Peng, C. K. (2005). Multiscale entropy analysis of biological signals. *Phys. Rev. E* 71, 1–18.
- D'Mello, S. (2011). "Dynamical emotions: bodily dynamics of affect during problem solving," in *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (Boston, MA).
- Dale, R., Warlaumont, A. S., and Richardson, D. C. (2011). Nominal cross recurrence as a generalized lag sequential analysis for behavioral streams. *Int. J. Bifurcat. Chaos* 21, 1153–1161.
- DePaulo, B. M. (1992). Nonverbal behavior and self-presentation. *Psychol. Bull.* 111, 203.
- DePaulo, B. M., and Friedman, H. S. (1998). "Nonverbal communication," in *The Handbook of Social Psychology*, eds D. T. Gilbert, S. T. Fiske, and G. Lindzey (New York, NY: Oxford University Press), 3–40.
- DePaulo, B. M., Kirkendol, S. E., Tang, J., and O'Brien, T. P. (1988). The motivational impairment effect in the communication of deception: replications and extensions. *J. Nonverbal Behav.* 12, 177–202.
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., and Cooper, H. (2003). Cues to deception. *Psychol. Bull.* 129, 74–118.
- Dickinson, M. H., Farley, C. T., Koehl, M. A. R., Kram, R., and Lehman, S. (2000). How animals move: an integrative view. *Science* 288, 100–106.
- Duran, N. D., Dale, R., and McNamara, D. S. (2010). The action dynamics of overcoming the truth. *Psychon. Bull. Rev.* 17, 486–491.
- Eapen, N. M., Baron, S., Street, C. N. H., and Richardson, D. C. (2010). "The bodily movements of liars," in *33rd Annual Conference of the Cognitive Science Society*, eds L. Carlson, C. Hölscher, and T. Shipley (Austin, TX: Cognitive Science Society).
- Eckmann, J. P., Kamphorst, S. O., and Ruelle, D. (1987). Recurrence plots of dynamical systems. *Europhys. Lett.* 4, 973–977.
- Ekman, P. (1992). *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage*. New York, NY: W. W. Norton.
- Ekman, P., and Friesen, W. V. (1969). The repertoire of nonverbal behavior: categories, origins, usage, and coding. *Semiotica* 1, 49–98.
- Ekman, P., and Friesen, W. V. (1972). Hand movements. *J. Commun.* 22, 353–374.
- Ekman, P., and Friesen, W. V. (2003). *Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues*. Los Altos, CA: ISHK.
- Fraizer, E. V., and Mitra, S. (2008). Methodological and interpretive issues in posture-cognition dual-tasking in upright stance. *Gait Posture* 27, 271–279.
- Gao, J., Cao, Y., Tung, W., and Hu, J. (2007). *Multiscale Analysis of Complex Time Series: Integration of Chaos and Random Fractal Theory, and Beyond*. New York, NY: John Wiley and Sons.
- Halley, J. D., and Winkler, D. A. (2008). Classification of emergence and its relation to self-organization. *Complexity* 13, 10–15.
- Harbourne, R. T., and Stergiou, N. (2009). Movement variability and the use of nonlinear tools: principles

Lastly, the current approach addresses an important debate in the deception literature concerning the tendency for deceivers to move less. It is unclear whether fewer movements are caused by excessive strategic management to the point that deceivers ironically overcompensate (DePaulo et al., 1988; see also Wegner, 2009) or a strategic move to prevent leakage cues (Burgoon, 2005). This is an important distinction for the lie detector. After all, if the behavior is strategic then its diagnosticity cannot be relied upon. An important facet of accurate lie detection, then, is not only discovering those behaviors that give liars away, but also determining if those behaviors are strategic in an attempt to minimize irrepressible "tells." Accordingly, dynamical measures of stability and complexity might have a great deal of relevance here. Although people may strategically minimize the overall magnitude of their movements, the dynamical structure of these movements are certainly outside of conscious control. And where a minimization of movement might be considered unintentional, it does not necessarily have to reflect impairment on part of the cognitive system. According to a main hypothesis, when the dynamical properties of movements are examined, what may be expressed are complex patterns of adaptation that emerge in task-specific ways. There are new and exciting ways to spot a liar.

ACKNOWLEDGMENTS

The research reported here was supported by a NSF Minority Postdoctoral Research Fellowship awarded to the first author and Grant NSF BCS-0826825 awarded to the second author. The opinions expressed are those of the authors and do not represent the views of the National Science Foundation.

- to guide physical therapist practice. *Phys. Ther.* 89, 267–282.
- Hill, M. L., and Craig, K. D. (2002). Detecting deception in pain expressions: the structure of genuine and deceptive facial displays. *Pain* 98, 135–144.
- Hurley, C. M., and Frank, M. G. (2011). Executing facial control during deception situations. *J. Nonverbal Behav.* 35, 119–131.
- Jensen, M. L., Meservy, T. O., Burgoon, J. K., and Nunamaker, J. F. (2010). Automatic, multimodal evaluation of human interaction. *Group Decis. Negot.* 19, 367–389.
- Kelso, J. A. S. (1995). *Dynamic Patterns: The Self-organization of Brain and Behavior*. Boston, MA: The MIT Press.
- Leal, S., Vrij, A., Fisher, R. P., and van Hooff, H. (2008). The time of the crime: cognitively induced tonic arousal suppression when lying in a free recall context. *Acta Psychol.* 129, 1–7.
- Marwan, N. (2003). *Encounters with Neighbours: Current Developments of Concepts Based on Recurrence Plots and Their Applications*. University of Potsdam, Doctoral dissertation.
- Marwan, N., Carmen Romano, M., Thiel, M., and Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Phys. Rep.* 438, 237–329.
- McNeill, D. (1996). *Hand and Mind: What Gestures Reveal About Thought*. Chicago: University of Chicago Press.
- Meservy, T. O., Jensen, M. L., Kruse, J., Burgoon, J. K., Nunamaker, J. F. Jr., Twitchell, D. P., et al. (2005). Deception detection through automatic, unobtrusive analysis of non-verbal behavior. *IEEE Intell. Syst.* 20, 36–43.
- Newell, K. M. (1998). “Degrees of freedom and the development of postural center of pressure profiles. Applications of nonlinear dynamics to developmental process modeling,” in *Applications of Nonlinear Dynamics to Developmental Process Modeling*, eds K. M. Newell and P. C. M. Molenaar (Mahwah, NJ: Lawrence Erlbaum Associates Publishers), 63–84.
- Osman, M., Channon, S., and Fitzpatrick, S. (2009). Does the truth interfere with our ability to deceive? *Psychon. Bull. Rev.* 16, 901–906.
- Pinheiro, J. C., and Bates, D. M. (2000). *Mixed-Effects Models in S and S-PLUS*. New York, NY: Springer Verlag.
- Port, R. F., and Van Gelder, T. (1995). *Mind as Motion: Explorations in the Dynamics of Cognition*. Boston, MA: MIT Press.
- Porter, S., and ten Brinke, L. (2010). The truth about lies: What works in detecting high-stakes deception? *Leg. Criminol. Psychol.* 15, 57–75.
- Porter, S., and ten Brinke, L. (2008). Reading between the lies. *Psychol. Sci.* 19, 508.
- Richman, J. S., and Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol.* 278, H2039–H2049.
- Riley, M., and Turvey, M. T. (2002). Variability and determinism in motor behavior. *J. Mot. Behav.* 34, 99–125.
- Riley, M. A., Baker, A. A., Schmit, J. M., and Weaver, E. (2005). Effects of visual and auditory short-term memory tasks on the spatiotemporal dynamics and variability of postural sway. *J. Mot. Behav.* 37, 311–324.
- Riley, M. A., Shockley, K., and Van Orden, G. (2012). Learning from the body about the mind. *Top. Cogn. Sci.* 4, 21–34.
- Riley, M. A., and Van Orden, G. C. (2005). *Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences*. Arlington, VA: National Science Foundation.
- Rozin, P., and Cohen, A. B. (2003). High frequency of facial expressions corresponding to confusion, concentration, and worry in an analysis of naturally occurring facial expressions of Americans. *Emotion* 3, 68.
- Schmidt, R. A. (2003). Motor schema theory after 27 years: reflections and implications for a new theory. *Res. Q. Exerc. Sport* 74, 366–375.
- Schmidt, R. A., and Lee, T. D. (2005). *Motor Control and Learning: A Behavioral Emphasis*. Champaign, IL: Human Kinetics Publishers.
- Seely, A. J., and Macklem, P. T. (2004). Complex systems and the technology of variability analysis. *Crit. Care* 8, R367–R384.
- Shockley, K. (2005). “Cross recurrence quantification of interpersonal postural activity,” in *Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences*, eds M. A. Riley and G. C. Van Orden (Arlington, VA: National Science Foundation), 142–177.
- Shockley, K., Santana, M. V., and Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *J. Exp. Psychol. Hum. Percept. Perform.* 29, 326.
- Stephen, D. G., Boncoddo, R. A., Magnuson, J. S., and Dixon, J. A. (2009). The dynamics of insight: mathematical discovery as a phase transition. *Mem. Cogn.* 37, 1132–1149.
- Takens, F. (1981). Detecting strange attractors in turbulence. *Lect. Notes Math.* 898, 366–381.
- ten Brinke, L., MacDonald, S., Porter, S., and O’Connor, B. (2011). Crocodile tears: facial, verbal, and body language behaviours associated with genuine and fabricated remorse. *Law. Hum. Behav.* 36, 51–59.
- Turvey, M. T. (1990). Coordination. *Am. Psychol.* 45, 938.
- Turvey, M. T. (2007). Action and perception at the level of synergies. *Hum. Mov. Sci.* 26, 657–697.
- Van Orden, G. C., Holden, J. G., and Turvey, M. T. (2003). Self-organization of cognitive performance. *J. Exp. Psychol. Gen.* 132, 331–350.
- Vrij, A., Akehurst, L., and Morris, P. (1997). Individual differences in hand movements during deception. *J. Nonverbal Behav.* 21, 87–102.
- Vrij, A., Granhag, P. A., Mann, S., and Leal, S. (2011). Outsmarting the liars: toward a cognitive lie detection approach. *Curr. Dir. Psychol. Sci.* 20, 28–32.
- Vrij, A., Granhag, P. A., and Porter, S. (2010). Pitfalls and opportunities in nonverbal and verbal lie detection. *Psychol. Sci. Public Interest* 11, 89–121.
- Vrij, A., Mann, S. A., Fisher, R. P., Leal, S., Milne, R., and Bull, R. (2008). Increasing cognitive load to facilitate lie detection: the benefit of recalling an event in reverse order. *Law Hum. Behav.* 32, 253–265.
- Vrij, A., Semin, G., and Bull, R. (1996). Insight into behavior displayed during deception. *Hum. Commun. Res.* 22, 544–562.
- Webber, C. L. Jr., and Zbilut, J. P. (1994). Dynamical assessment of physiological systems and states using recurrence plot strategies. *J. Appl. Physiol.* 76, 965–973.
- Wegner, D. M. (2009). How to think, say, or do precisely the worst thing for any occasion. *Science* 325, 48–50.
- Woollacott, M., and Shumway-Cook, A. (2002). Attention and the control of posture and gait: a review of an emerging area of research. *Gait Posture* 16, 1–14.
- Wraga, M., Duncan, L., Jacobs, E. C., Helt, M., and Church, J. (2006). Stereotype susceptibility narrows the gender gap in imagined self-rotation performance. *Psychon. Bull. Rev.* 13, 813–819.
- Zuckerman, M., DePaulo, B. M., and Rosenthal, R. (1981). “Verbal and nonverbal communication of deception,” in *Advances in Experimental Social Psychology*, ed L. Berkowitz (New York, NY: Academic Press), 1–59.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 20 September 2012; **accepted:** 04 March 2013; **published online:** 27 March 2013.

Citation: Duran ND, Dale R, Kello CT, Street CNH and Richardson DC (2013) Exploring the movement dynamics of deception. *Front. Psychol.* 4:140. doi: 10.3389/fpsyg.2013.00140

This article was submitted to Frontiers in Cognitive Science, a specialty of Frontiers in Psychology.

Copyright © 2013 Duran, Dale, Kello, Street and Richardson. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.



How do incentives lead to deception in advisor-client interactions? Explicit and implicit strategies of self-interested deception

Barbara Mackinger * and Eva Jonas

Department of Psychology, Paris Lodron University Salzburg, Salzburg, Austria

Edited by:

Matthias Gamer, University Medical Center Hamburg-Eppendorf, Germany

Reviewed by:

Richard J. Tunney, University of Nottingham, UK

Shaul Shalvi, University of Amsterdam, Netherlands

***Correspondence:**

Barbara Mackinger, Department of Psychology, Paris Lodron University Salzburg, Hellbrunnerstraße 34, 5020 Salzburg, Austria.

e-mail: barbara.mackinger@sbg.ac.at

When confronted with important questions we like to rely on the advice of experts. However, uncertainty can occur regarding advisors' motivation to pursue self-interest and deceive the client. This can especially occur when the advisor has the possibility to receive an incentive by recommending a certain alternative. We investigated how the possibility to pursue self-interest led to explicit strategic behavior (bias in recommendation and transfer of information) and to implicit strategic behavior (bias in information processing: evaluation and memory). In Study 1 explicit strategic behavior could be identified: self-interested advisors recommended more often the self-serving alternative and transferred more self-interested biased information to their client compared to the advisor without specific interest. Also deception through implicit strategic behavior was identified: self-interested advisors biased the evaluation of information less in favor of the client compared to the control group. Self-interested advisors also remembered conflicting information regarding their self-interest worse compared to advisors without self-interest. In Study 2 beside self-interest we assessed accountability which interacted with self-interest and increased the bias: when accountability was high advisor's self-interest led to higher explicit strategic behavior (less transfer of conflicting information), and to higher implicit strategic behavior (devaluated and remembered less conflicting information). Both studies identified implicit strategic behavior as mediator which can explain the relation between self-interest and explicit strategic behavior. Results of both studies suggest that self-interested advisors use explicit and implicit strategic behavior to receive an incentive. Thus, advisors do not only consciously inform their clients "self-interested," but they are influenced unconsciously by biased information processing – a tendency which even increased with high accountability.

Keywords: strategic behavior, deception, self-interest, incentive, advice-giving, motivated information processing, principal-agent theory

"If you want people to perform better, you reward them, right?
Bonuses, commissions, their own reality show. Incentivize them."
(Daniel Pink¹)

This quote of a well known American career analyst explains one common strategy of motivating employes in the business world. It follows a simple analogy: sell or produce X for the company and you will get Y as a reward. Especially, at a time where companies are in trouble to survive on the market, they are challenged to perform well. Taking the competitive nature of the market into account, it is comprehensible that companies use incentives as an instrument to motivate employes. Incentives are assumed to encourage employes in accomplishing a stated goal, or even in following a goal that they would normally have no other reason for pursuing. Incentives should help by guiding self-interested behavior and adjusting the employes' interests to the company's interests.

Indeed, there are many examples showing that using incentives often work out well for companies. One company, which has been working successfully with incentives, is Tupperware. Their marketing principle of using Tupperware-parties to sell their products is very effective. These parties take place in the private atmosphere of someone's home. The host provides the location, snacks, and refreshments, and invites the guests. Well, how can that be that a private person starts a marketing event for a company he or she does not work for? Back in the 1950s, bringing women together and discussing housekeeping and kitchen secrets was nothing new, but receiving additional incentives for hosting the event and for providing people who buy a lot, was an innovation. Today, the promise of incentives especially in form of turnover-dependent commission is still assumed to have a direct impact on peoples' behavior. The host invites a lot of people, creates a nice surrounding atmosphere with sandwiches and drinks because he or she assumes that this behavior will pay off in the end.

However, the question of interest is whether incentives really just motivate employes or hosts of Tupperware-parties to behave

¹Author of "Drive: The Surprising Truth About What Motivates Us."

in vein with the company's strategic interest, or does it lead to more wide reaching consequences? For the company it is somehow desirable that the host behaves strategically in order to pursue self-interest to earn a lot of money. This type of strategic behavior is in line with the intention to sell a lot. Therefore, the company directs the hosts' employees' behavior by incentives which are adapted to its own interests. Thus, this type of strategic behavior might also lead to deceive the other party consciously: for example the host is recommending especially expensive products to the customers because he or she receives a turnover-dependent commission. For this reason, he or she may communicate only the advantages of the product and withhold the disadvantages (e.g., Buller and Burgoon, 1994, 1996; Steinle and De Dreu, 2004). This behavior is henceforth called *explicit strategic behavior*. Explained more in detail, people showing strategic behavior can have the explicit goal to deceive and therefore alter the customer's behavior or opinion in order to receive an incentive.

However, we speculate that the promise of incentives could also entail unconscious risks. In order to gain reward, people might behave more implicitly strategic which supports the deception of the other party. Past research already showed that deception is also accompanied by more automatic and less conscious actions, such as smiling longer and nodding when deceiving our counterpart in face-to-face interaction (e.g., Buller and Burgoon, 1994; Burgoon et al., 1996). However, this is not an explicit strategy to deceive the other party. People simply seem to implicitly engage in this deception process. In our opinion, deception can also already start before the interaction with a counterpart, for example through biased information processing. This occurs when the host evaluates and remembers expensive products biased in favor of one's possibility to receive more reward. We call the biased information processing henceforth *implicit strategic behavior*, which in contrast to explicit strategic behavior is not clearly linked to deception, because it is not used to alter the client's behavior or thinking in order to increase their own incentives. However, we think that it is triggered by the wish to receive a high incentive and we therefore investigate it as an implicit strategic behavior of deceiving the other party.

As is generally known, the promise of incentives is not only common practice in occurrences like Tupperware-parties, where peer-advice is given. Incentives are also widespread in professional consultancy where advisors have specialized knowledge, which is demanded by clients to improve their decision (e.g., financial advisors, physicians, and personnel advisors). Clients are in a clear disadvantage because of their lack of knowledge and advisors can use these scopes to deceive the client in self-interested manner. In the current paper we investigate if the promise of incentives motivates advisors to make consciously use of their advantage and behave explicitly strategic by deceiving the other party. However, we assume advisors are also influenced by the incentive more implicitly, which should display in biased information processing. This is especially risky and therefore highly relevant for advisors who prepare information for their clients. So far, past research did not investigate this type of strategic behavior which might support the explicit strategic behavior to deceive a counterpart. We test our assumptions in two studies. Whereas our second study additionally examines to what extent incentivized advisors

behave strategically when they feel highly accountable for their clients.

EXPLICIT STRATEGIC BEHAVIOR

Strategic behavior is primarily used to create a false belief in order to deceive the other party. One theoretical background of deception is the Interpersonal Deception Theory (IDT, Buller and Burgoon, 1994, 1996), which focuses primarily on the face-to-face interaction and the dynamic of deceptive interaction between sender and receiver (verbal and non-verbal). Buller and Burgoon (1994) describe observable strategic behavior during the deception (e.g., ambiguous, vague, and intentional messages), which is used by the senders to alter the communication in order to achieve their goal.

An alternative description of deception and strategic behavior is proposed by the economic principle-agent theory (PAT; Ross, 1973), which focuses less on communication and more on strategic actions within working relationships. According to this theory, strategic behavior is likely to occur when one party (principal) delegates work to a more knowledgeable party (agent) and the two parties have different interests. In other words, the client is under uncertainty when asking a knowledgeable advisor for support to find the best solution for a problem, if the advisor really acts in client's best interest or in their own interest. Advisors, in the role of an expert, have wide-ranging possibilities to use scopes to behave strategically and to pursue their self-interest. The client lacks the knowledge to fully evaluate the quality of the advisor's recommendation. Therefore, the advisor tends to provide recommendations which support their own self-interest rather than the client's interest.

Deception and the different levels of knowledge between two parties were already found as crucial in bargaining experiments especially when stakes were high (Boles et al., 2000): In this situation participants in the role of the proposer behaved strategically by offering less and providing a worse bargain to their unknown counterpart (who did not know the size of the pie) compared to a more knowledgeable counterpart. This means proposers pursued self-interest, when there was a low possibility to be detected by the counterpart and only when the stakes were high (Boles et al., 2000).

We would now like to turn to advice-giving situations, where likewise many opportunities exist to deceive the client. This should be illustrated with the help of an example of a personnel advisor: Companies hire a personnel advisor when needing support for a specific job placement and typically pay him or her with an agency fee after successful staffing. This procedure, as typically used by companies, leads to an incentive for the advisor to find a suitable candidate as soon as possible in order to fulfill the contract and to receive the incentive (agency fee). In other words, a fast fulfillment of the contract enhances the chances of the advisor to soon be available for a new contract and a new possibility to earn money. Thus, for the company and the possible job applicant the risk exists that the advisor is rather interested in a fast than in the best job placement (goal conflict). Because of advisor's information advance (e.g., job market, job duties, salary, and education), he or she can make use of existing scopes and deceive by behaving explicitly strategic. This information asymmetry is existing

in many advisor-client interactions and similar behavior by the advisors can be assumed, such as physicians or financial advisors.

According to the PAT (Ross, 1973; for overview Eisenhardt, 1989), in such situations strategic behavior occurs in different forms, such as "Moral hazard," withholding actions or information ("hidden action," "hidden information"), and also in terms of "Hold up" ("hidden intention"), such as concealing the own goals and intentions. Similarly, IDT (Buller and Burgoon, 1994, 1996) list (explicit) strategic behaviors regarding the regulation of the content of the spoken information ("information management"). Deceivers, for instance, were found to behave strategically and use higher rates of irrelevant and vague information (Buller and Burgoon, 1994).

Indeed, further research shows that participants behave strategically by managing and controlling information. For example, Steinel and De Dreu (2004) found that participants of an information provision game behaved strategically. Participants were in the position to guide their counterpart which had opposing interests (losing points when the counterpart gain points) through passing accurate or inaccurate information. The aim of the participants was to reach as many points as possible, since these determined the amount of lottery tickets they would receive in the end. Results demonstrated that participants behaved explicitly strategic by withholding more accurate information and passing more inaccurate one. Additionally the authors could identify that strategic behavior was mainly driven by greed within this interdependent relationship with opposing interests (Steinel and De Dreu, 2004).

Looking back at the case of the personnel advisors this means that the possibility to receive an incentive for recommending a specific alternative may frequently lead to explicit strategic behavior: the advisor may offer only jobs to applicants which allows a quick recruitment process with low effort (PAT: "hidden action"). To achieve this, the advisor may strategically transfer information to the client to accelerate the process of convincing the client to accept the easily available job instead of prolonging the search for the best job alternative. This can occur by solely presenting those aspects of the easily available job which are in line with the applicant's demands (good labor-market situation, career opportunities) or by withdrawing the conflicting information (salary, job characteristics) (PAT: "hidden information" and Steinel and De Dreu, 2004). Based on this assumption we assume the following hypothesis regarding the explicit strategic behavior:

Hypothesis 1: We assume that the self-interested advisors recommend the easily available job more often compared to the advisor without specific interest.

Hypothesis 2: We suppose that self-interested advisors enhance supporting information and devalue conflicting information regarding their self-interest whereas advisors without specific interest do not make this difference.

Well, if these predictions become true they are a large problem for advice-taking situations, where clients often have to rely on the knowledge of an advisor as expert. However, past deception research has only tested the conscious modification of information (PAT: Ross, 1973; IDT: Buller and Burgoon, 1996). Heretofore,

it has failed to incorporate the more implicit biasing of the information processing within the deception. But we assume that the motivation to gain reward might influence the advisor even more implicit.

MOTIVATED DIRECTIONAL GOAL AND IMPLICIT STRATEGIC BEHAVIOR

First, before we fully understand implicit strategic behavior, we need to describe the process of advice-giving and to differentiate the phase of explicit advice-giving, and the preceding phase of preparing the recommendation (Jungermann, 1999). Within the latter one the more implicit information processing has to be taken into account. In this phase the advisors search, evaluate, and process information. At the latest since Kunda (1990) we know that there are motives, such as receiving an incentive, that influence our reasoning. This means people's goals, wishes, fears, and desires lead people to engage in biased information processing and direct our thinking and convictions (Kruglanski, 1989; Kunda, 1990; Dunning, 1999; Kruglanski et al., 2012). Kruglanski et al. (2012) describes a directional motivation like a psychological force, similar to a physical force, which is determined by a specific desired goal. People perceive the world through their "motivated colored" glasses – perceiver's goals are crucial for reconciling incoming information.

Past research provided evidence that such directional goals are also important for predicting advisors' behavior in advice-giving processes (Jonas et al., 2005). This findings indicated that advisors, who had to justify their recommendation and therefore had an incentive to appear in a positive light in front of the client (impression motivation), biased their information search in favor of their recommendation and passed on more information supporting the recommendation (Jonas et al., 2005). In contrast, advisors without directional goals were found to act accuracy motivated and were normally directed by finding the best solution for their clients (Jonas and Frey, 2003). Therefore, only in order to reach the directional goal, making a good impression, advisors behaved implicitly strategic by searching information which primarily supported their recommendation.

Implicit strategic behavior can be further illustrated through the previously introduced example of the personnel advisor, who has the directional goal of fast contract fulfilling in order to receive the incentive quickly. An advisor who is preparing a recommendation for a job applicant, always has in mind which jobs are easily available at the moment and can be staffed quickly, allowing the advisor to earn more money. In this scenario earning money can be seen as a main motivation. To reach the directed goal of earning money, the advisor might behave implicitly strategic when evaluating job information. He or she might evaluate the information of an easily available job less critical than that of a difficult available job: the advisor may enhance job relevant information which is in line with his or her goal. Similarly, the advisor might also remember information more easily which supports his or her goal compared to conflicting information. This biased information processing might also implicitly influence the recommendation that is given.

Going beyond to the assumption of PAT (Ross, 1973) people's motivation to reach a goal do not only lead to an actor's obvious

and conscious behavior (e.g., advisors strategic behavior – explicitly hiding information). Past research in motivated reasoning suggests that people engage in biased information processes because they find it more plausible to process information in line with their own beliefs and expectancies (e.g., McDonald and Hirt, 1997) and to remember desired aspects (e.g., Sanitioso et al., 1990). These findings indicated enhanced accessibility of knowledge structures, which are in line with the desired goal, so that, the personnel advisor might be caught in biased information processing, when trying to fulfill the contract quickly in order to receive the incentive. Even when the information processing is not directly and consciously used to receive an incentive, the personnel advisor will evaluate and maybe even remember information in favor of their self-interest. We propose that this phenomenon can be seen as an implicit and unconscious process – or in other words as implicit strategic behavior to get the promised incentive.

Interestingly, receiving incentives has already led to different assumptions regarding information processing and was discussed controversially in past research. On one hand, there is evidence that incentives lead to higher accuracy motivation and that people put more effort in information processing when receiving an incentive (Stone and Ziebart, 1995). On the other hand, incentives can also lead to a higher confirmation bias (Jonas et al., 2008): Participants who were promised an incentive for finding the correct answer showed a preference in searching for supportive rather than conflicting information regarding their preliminary decision. Moreover, they also remembered conflicting information worse. This research showed that incentivized participants were more biased in their information processing than participants without incentives. This poses the question of whether information processing is also biased in order to receive incentives. We assume that the incentives influence the advisor's thinking and convictions in a similar manner as a directional goal. The personnel advisor might bias information processing in favor of their self-interest, or the job alternative which is associated with the incentives.

Independent from incentives, other research showed that self-interest had an influence on the information processing, such as the evaluation and memory of self-interested information (Kunda, 1990). In a study of Ditto et al. (1998) participants were tested by means of a clinical test, which had either positive or negative consequences for the participants' health. Participants had to evaluate the test and their test results. The results showed that they were rather dismissive when they were confronted with negative health consequences, compared to participants who faced a result with a positive health consequence. Additionally, based on the research of Kunda and colleagues (e.g., Kunda, 1987; Kunda and Sanitioso, 1989; Sanitioso et al., 1990) we know that our self-interest also influences our memory search. For instance, participants who were persuaded that introversion (or extraversion) is more desirable for academic success described themselves as more introverted (or extroverted). Furthermore, they were able to report more and faster introverted (or extroverted) behavior pattern than the participants who had been convinced of the opposite (Sanitioso et al., 1990).

Similar results were found in the field of the persuasion research (Petty and Cacioppo, 1990) where people's different involvements and benefits (vested interest) from actions in society led to different

evaluations of information (outcome-relevant involvement, Johnson and Eagly, 1989). Four experimental studies by Darke and Chaiken (2005) showed that self-interest influences the direction of attitudes and the persuasive impact of arguments: participants, who had to pay the costs and did not receive any benefits, devalued the new policy (tuition fees) by processing the information of the arguments in a biased way. In similar vein the research in motivated skepticism of political beliefs found evidence that people used biased information processing as means of finding consistency with their own favored view (Taber and Lodge, 2006). Participants who felt strongly about an issue – even when encouraged to be objective – evaluated supportive arguments more favorably than conflicting arguments. This research indicates that people in social interactions such as discussing new policies are biased through their self-interest. They want to bolster their view and find consistency for the own favored view.

However, this research does not state how people are influenced by their self-interest when processing information for another person and preparing an advice. In the present study, we want to investigate how self-interest in the form of receiving incentives biases people's information processing and implicitly influence the deception in advice-giving. Furthermore, advisors may put more focus and effort in understanding the match between an applicant and an easily available job compared to other job alternatives, which may display in a biased memory. We suppose this also in our following hypothesis.

Hypothesis 3: We propose that self-interested advisors enhance the relevance of supporting information and devalue the conflicting information regarding their self-interest whereas advisors without specific interest do not make this difference.

Hypothesis 4: We suppose that self-interested advisors remember conflicting information regarding their self-interest worse, than advisors without specific interest.

Finally, it remains the question how implicit and explicit strategic behaviors are connected and how advisor's implicit biases shapes self-serving deceptive behavior. A recent work Shalvi et al. (2011) assessed deception in participants who were asked to privately roll a die under a paper cup three times versus only once. Afterward they reported the outcome of the first roll and gain money as a function of their reports ($1 = \$1$, $2 = \$2$, etc.). Results suggested that the degree of lying was significant higher in the condition with three times compared to once rolling the die because of the higher extent of self-justification by referring to the highest outcome of the three rolls. In sum, self-interest led people to view objective information in a biased way which supported their self-justification and enabled them to lie. Specifically, the authors assume that people balance their desire to profit from the lie with a desire to maintain their self-concept as honest individuals. Similar in our study, self-interest bias in information processing (implicit strategic behavior) enables the advisor to justify the later self-interested explicit strategic behavior. Therefore we propose that implicit strategic behavior should help to describe the process of explicit strategic behavior.

Hypothesis 5: We assume that the connection between self-interest and explicit strategic behavior (transfer of

information) is mediated to some extent by implicit strategic behavior (evaluation of information).

PRESENT RESEARCH

In order to assess the outlined hypotheses, in our first study we asked participants to put themselves in the role of either self-interested personnel advisor or a personnel advisor without self-interest. The scenario was similar as already introduced in the beginning. However, in our second study we want to enhance advisor's responsibility and accountability for their recommendation. Therefore, besides self-interest we investigated advisor's accountability and how this influenced explicit and implicit strategic behavior.

STUDY 1

METHOD

Participants and design

Participants were 67 students (54 female, 13 male) at a public university of Austria (University of Salzburg). Psychology students could volunteer in order to receive credits for participation. They participated individually. The study was a two factorial design with two conditions 2(self-interest: yes vs. no) \times 2(type of information: supporting vs. conflicting).

Scenario and task

After participants consented to being in the study, they were placed in a quiet area. The questionnaire started with the description of the personnel advisor's scenario. Part of the scenario was a fictitious client. He was described as a young male high-school graduate that is interested in the field of engineering. The advisor got informed that different tests could already confirm the appropriateness of this client in this field and about some important specific details which should be taken into account for the recommendation (e.g., salary as important criterion, above the average in logical reasoning, loves challenge in logical thinking, below average in the ability to work, and cooperate in teams).

Participants assumed either the role of a freelancer personnel advisor with the possibility to earn an incentive when pursuing self-interest (*self-interested advisor*), or a personnel advisor of an institution for professional training (*advisor without specific interest*). Only the self-interested advisor is also under contract of a company, which commissioned the advisor to find an appropriate candidate for the job of a *product engineer*. Therefore the advisor could pursue self-interest and receive an incentive by recommending this job to the client. Participants also pursued real self-interest – they only participated in lottery to win one of ten 20€-Amazon-voucher if they recommended the product engineer. In the other condition the advisor had no additional interest to fulfill a contract with a company in order to receive an incentive and also participants themselves took part in the lottery independently of their recommendation.

The assignment for the career consultant was then to read information about three vocational trainings for the client: machinery engineer, mechatronic engineer, or product engineer. For preparing the recommendation, the participants had to evaluate the information. Further they expressed their intention of transferring information to the client and finally recommended one job.

In the end, the participants answered questions regarding their perceived self-interest and took part in a quiz regarding the job information.

MATERIAL

Job information – conflicting and supporting information

Information covered six categories for the three different job possibilities (see **Table A1** in Appendix) and there are clear opposed interests between the client and the self-interested advisor. On the one hand, the mechatronic engineer's job description fitted best with the client's demands – it covered most of his needs and wishes (high salary, logic reasoning is important, and working in teams is not mentioned as key competence). On the other hand, the product engineer is the best option for the advisor to meet his/her self-interest. The information about the machinery engineer were similar attractive to the product engineer – both had three pieces of information which covered the wishes of the client and three which were in conflict.

For further analysis, the pieces of information are used as either supportive or conflicting with advisor's self-interest. Conflicting information are all information which weaken the realization of the advisor's self-interested goal (negative arguments for the product engineer, positive arguments for the mechatronic, and machinery engineer). In contrast, supportive information contains all arguments which can bolster the self-interested recommendation of the product engineer (positive arguments for the product engineer, negative arguments for the mechatronic, and machinery engineer). We want to refer to **Table A1** in Appendix where we signed supportive arguments for the advisor with plus and conflicting information of information with minus.

Explicit strategy

Transfer of information. Participants marked on a 10 cm line how likely they would hand this information to their client. For further analysis we divided the scale in the transfer of the supporting (six items, Cronbach's $\alpha = 0.72$) and conflicting information (11 items, Cronbach's $\alpha = 0.90$).

Recommendation. Additionally, after reading all information the participants had to decide for one specific job – product engineer, mechatronic engineer, and machinery engineer – which she or he would recommend to the client.

Implicit strategy

Evaluation of information. The participants in the role of the advisor had to decide how relevant the job information is. They marked their evaluation on a 10 cm line which reached from not relevant to very relevant. For our further analysis we used the evaluated relevance of the supporting (six items, Cronbach's $\alpha = 0.83$) and the conflicting information (11 items, Cronbach's $\alpha = 0.85$) regarding the advisor's self-interest.

Memorizing information. Subsequently the participants were requested to answer six multiple choice quiz questions regarding information of the three different job alternatives to measure how much information was memorized. Actually, we used only

a single-item measure for further analysis; however it is the most conflicting information regarding the advisor's self-interest (salary of the mechatronic engineer).

Manipulation check

Perceived self-interest. Additionally, the perceived own self-interested behavior was measured with the scales *hidden intention* (e.g., "Situations where the client's and my interests were in conflict, I oriented primarily on my interests." Five items, Cronbach's $\alpha = 0.92$), *hidden information* (e.g., "Some important information were not communicated to the client." Four items, Cronbach's $\alpha = 0.83$) and *hidden action* (e.g., "Some actions were more in my interest than in the interest of the client." Two items, $r(66) = 0.65$, $p < 0.01$). Participants answered by responding to the questions on a five-point Likert-scale (1 = strongly disagree to 5 = strongly agree). For further analysis we combined these three scales and measured self-interested behavior in general (11 items, Cronbach's $\alpha = 0.95$).

RESULTS

MANIPULATION CHECK

For checking the influence of the manipulation we used the scale general perceived self-interest. Analysis regarding the influence of the manipulated self-interest revealed a main effect, $F(1,67) = 18.17$, $p < 0.001$. The freelancer advisor perceived him/herself as significant more self-interested compared to the personnel advisor of an institution for professional training without any specific self-interest ($M_s = 2.72$ vs. 1.84 , $SDs = 1.09$ vs. 0.47). Thus, this result suggests that the intended factors were manipulated successfully.

EXPLICIT STRATEGIES

Based on the assumption that the advisor is influenced by the promise to receive an incentive in a self-interested manner, we expected an explicit strategic behavior in the explicit recommendation to the client and in the transfer of information. The self-interested transfer of information should be characterized by withholding conflicting information regarding the self-interest and pushing forward supporting information.

Advice-giving – Hypothesis 1

In line with our first hypothesis a Chi-squared test on advice-giving strategy displayed that participants in the role of self-interested freelancer advisor recommended significant more often the less appropriate option "product engineer" to their client than participants without self-interest, $\chi^2(1, N = 67) = 8.49$, $p = 0.04$ (self-interest: 10 product engineer, 24 no product engineer; without self-interest: 1 product engineer, 32 no product engineer). The result supported our assumption that participants who had a personal self-interest are influenced in advice-giving and recommended significantly more often the product engineer, which is the self-interested alternative for the advisor. Additionally, Chi-squared analysis with all three job alternatives showed that advisors with no specific interest recommended more often the optimal job to their client ("mechatronic engineer") than participants with self-interest, $\chi^2(1, N = 67) = 8.61$, $p = 0.014$ (without self-interest: mechatronic 30, product 1, machinery 2 vs. self-interest:

mechatronic 23, product 10, machinery 1). These results support our hypothesis that self-interested advisors recommend the self-interested alternative of the product engineer more often compared to advisors without specific interest. Additionally advisors without specific interest recommended the optimal job more often than those with self-interest.

Transfer of information – Hypothesis 2

To test this hypothesis we ran a 2 (self-interest: yes vs. no) \times 2 (information: supporting vs. conflicting) analysis of variance with repeated measures on the last factor. This analysis revealed no main effect for the information, $F(1,65) = 1.91$, $p = 0.17$, $\eta_p^2 = 0.029$ (supporting: $M = 7.62$, $SD = 1.48$ vs. conflicting: $M = 7.30$, $SD = 1.89$). However, the analysis displayed a significant interaction effect between job information and self-interest, $F(1,65) = 13.50$, $p < 0.001$, $\eta_p^2 = 0.17$. Subsequent *post hoc* analysis indicated that results are in line with our predictions: Participants in the role of the self-interested advisor transferred less conflicting information ($M = 6.64$, $SD = 2.23$) than the advisor without specific interest ($M = 7.97$, $SD = 1.16$), $F(1,65) = 9.20$, $p = 0.003$. Additionally, self-interested advisors passed significant more information which supported their self-interest to the client than information which conflicted their self-interest ($M = 7.81$, $SD = 1.54$ vs. $M = 6.65$, $SD = 2.23$), $F(1,65) = 12.98$, $p = 0.001$. There was a tendency that advisors without specific interest transferred even less supporting than conflicting information ($M = 7.44$, $SD = 1.41$ vs. $M = 7.97$, $SD = 1.55$), $F(1,65) = 1.04$, $p = 0.112$. Regarding the supporting information there was no significant difference between advisors with self-interest and without specific self-interest ($M = 7.81$, $SD = 1.54$ vs. $M = 7.44$, $SD = 1.41$) $F(1,65) = 1.04$, $p = 0.311$. This indicates that advisors with self-interest primarily withhold conflicting information and did not transfer more supporting, whereas advisors without specific interest transferred information more balanced – with a contrary tendency to transfer more conflicting than supporting information. Results are displayed in **Figure 1**.

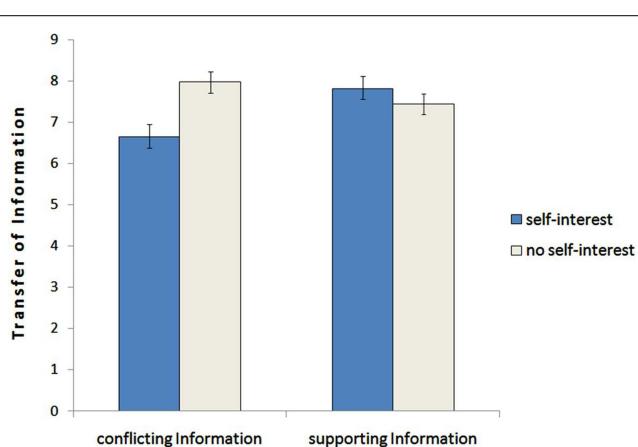


FIGURE 1 | Mean differences representing the transfer of information regarding supportive and conflicting information for advisors with self-interest and without self-interest. The error bars represent SEM (study1).

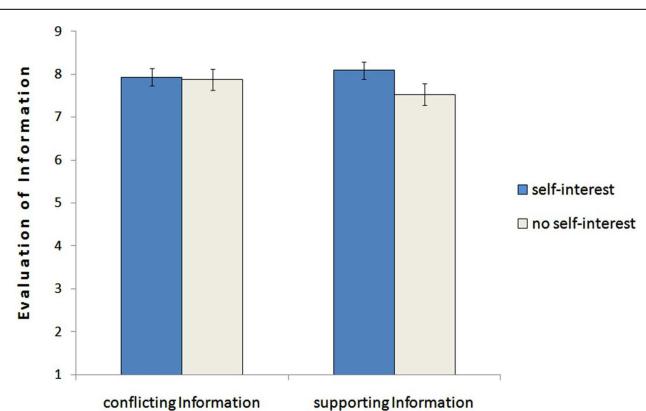


FIGURE 2 | Mean differences representing the evaluation of information regarding supportive and conflicting information for advisors with self-interest and without self-interest. The error bars represent SEM (study 1).

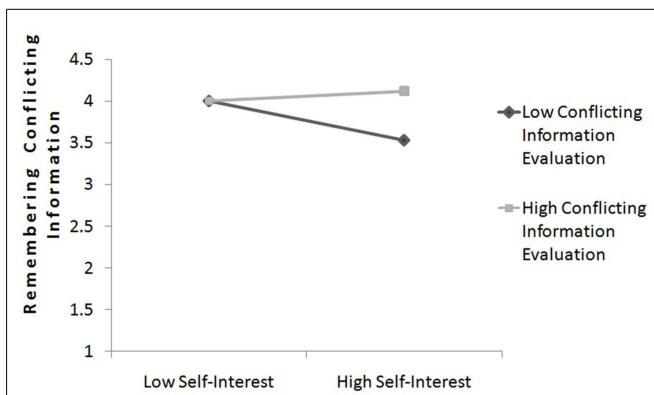


FIGURE 3 | The relationship between self-interest and remembering conflicting information as a function of the evaluation of conflicting information (study 1).

IMPLICIT STRATEGIES

Well, besides explicitly strategic advice-giving we predict that advisor's information processing is also influenced by incentives. The promise of incentives leads to self-interested bias in the evaluation of information and in memorizing the information (supporting vs. conflicting information regarding advisor's self-interest).

Evaluation of information – Hypothesis 3

To examine the effect of the self-interest on the evaluation of the information, we ran a 2 (self-interest: yes vs. no) \times 2 (information: conflicting vs. supporting) analysis of variance with repeated measures on the last factor. This analysis revealed no main effect for the information, $F(1,65) = 0.67, p = 0.42, \eta_p^2 = 0.10$ (conflicting: $M = 7.90, SD = 1.16$ vs. supporting: $M = 7.82, SD = 1.44$). However, it showed a significant interaction effect between job information and self-interest, $F(1,65) = 4.97, p = 0.029, \eta_p^2 = 0.07$. Post hoc analysis verified the pattern that there was a tendency for supporting information to be higher evaluated by the self-interested advisor than by the advisor without special interest ($M = 8.09, SD = 1.41$ vs. $M = 7.53, SD = 1.43$), $F(1,65) = 2.57, p = 0.114$. In contrast, the conflicting information was evaluated similarly in its relevance by the self-interested advisor and the advisor without interest ($M = 7.93, SD = 1.30$ vs. $M = 7.87, SD = 1.02$), $F(1,65) = 0.04, p = 0.835$. However, advisors without specific self-interest devaluated supporting information significantly compared to conflicting information ($M = 7.53, SD = 1.43$, vs. $M = 7.87, SD = 1.02$), $F(1,65) = 4.56, p = 0.036$; advisors with self-interest did not evaluate supporting and conflicting information significantly different ($M = 8.09, SD = 1.41$, vs. $M = 7.93, SD = 1.30$), $F(1,65) = 1.01, p = 0.318$. The hypothesis gets support by the significant interaction between self-interest and type of information, whereas the interaction is mainly driven by the enhanced evaluation of the conflicting information compared to the supporting information within advisors without specific interest. It seems that this distinction regarding the evaluation of the information disappears when pursuing self-interest. Results are displayed in **Figure 2**.

Memorized information – Hypothesis 4

Further, we tested the influence of self-interest on the memorized information. The results indicated that self-interested participants remembered significantly worse than the mechatronic engineer had the best possibilities to receive a good salary (conflicting information), $t(65) = 2.00, p = 0.05$ (self-interest $M = 3.79, SD = 0.59$; without self-interest $M = 4.00, SD = 0.00$). Really remarkable was that each participant without specific could remember the correct answer.

For further exploratory analysis of our data we used the pursued self-interest² together with the devaluation of conflicting information to predict biased memorized information (salary of the mechatronic engineer). We conducted a hierarchical regression analysis in which memorized conflicting information was predicted by main-effect terms (evaluation of the conflicting information and self-interest) and the interaction term simultaneously. Following Aiken and West (1991), the variables evaluation of conflicting information and self-interest were centered (i.e., by subtracting the mean from each score), and the interaction term was based on these centered scores. The interaction between evaluation of the conflicting information and self-interest was significant, $b = 0.32, SE = 0.10, t(63) = 3.18, p = 0.002$. Simple slope analysis was conducted to further analyze this interaction (Aiken and West, 1991). When the relevance of the conflicting information was high (1 SD above the mean), self-interest was not significantly related to memorized information, $b = 0.22, SE = 0.16, t(63) = 1.39, p = 0.170$, which means among participants who evaluated conflicting information high self-interest had no specific influence on the memorized knowledge. However, when the relevance of conflicting information was evaluated low (1 SD below the mean; $b = -0.43, SE = 0.15, t(63) = -2.91, p = 0.005$), self-interest was associated with less memorized information. The slopes are plotted in **Figure 3**.

²The scale perceived self-interest of the manipulation check was used.

Mediation – Hypothesis 5

We assumed that the connection between self-interest and explicit strategic behavior can be explained to some extent by implicit strategic behavior. Therefore we conducted a mediation model with the implicit strategic behavior (evaluation of conflicting information) as a potential mediator, which should help to explain the relation between self-interest and explicit strategic behavior (transfer of conflicting information). The first regression analyses showed that self-interest was significantly associated with the potential mediator implicit strategic behavior, $b = -0.31$, $SE = 0.12$, $t(67) = -2.59$, $p = 0.012$. In the second step we tested whether implicit strategic behavior was significantly associated with the explicit strategic behavior – and indeed, implicit strategic behavior significantly predicted the explicit strategic behavior, $b = 0.38$, $SE = 0.06$, $t(67) = 6.10$, $p < 0.001$. In the final step we examined whether statistical control for the potential mediator reduced the predictive power of the relation between self-interest and explicit strategic behavior. Without the mediator the effect was significant, $b = -0.80$, $SE = 0.08$, $t(67) = -10.68$, $p < 0.001$, however, when controlling for the mediator the relationship was considerably reduced, $b = -0.68$, $SE = 0.06$, $t(67) = -10.83$, $p < 0.001$. Finally, in a bootstrap analysis implicit strategic behavior significantly carried the indirect effect, 95% CI = -0.24 to -0.02. Thus, evidence was found that the direct effect of self-interest on the explicit strategic behavior occurred partly through the implicit strategic behavior, which supports our Hypothesis 5.

DISCUSSION STUDY 1

Our results indicate that advisors with self-interest behaved explicitly strategic by recommending the self-interested alternative of the product engineer more often compared to advisors without specific interest. The self-interested advisor also transferred less conflicting than supporting information to the client, as well as self-interested advisors transferred less conflicting information compared to advisors without self-interest.

Self-interested advisors also behaved implicitly strategic. The evaluation of information led in advisors without specific interest to a significant enhanced evaluation of the conflicting information (supporting for the client, see **Table A1** in Appendix) compared to the supporting information (conflicting for the client). This pattern displays the evaluation of information when having the best interest of the client in mind. The significant differentiation disappeared in advisors with self-interest. They did not take the perspective of the client and his needs and therefore, evaluated conflicting and supporting as similar important. Furthermore, we could confirm direct influence of self-interest on advisors' biased memory. However, the investigation should be improved in Study 2, because in Study 1 we could only refer to one quiz question. Additionally, we could identify evaluation of conflicting information as moderator. Especially when the relevance of conflicting information was devalued self-interest had a significant negative influence on memorizing conflicting information.

With regard to the mediation analysis we found important evidence for the connection between implicit and explicit strategic behavior. Our results indicate that implicit strategic behavior

can partly explain the relation between self-interest and explicit strategic behavior. This finding supports our assumptions that incentives have profound effects which influence people more implicitly and not only explicitly as assumed by the usual practice of incentives.

STUDY 2

In study 1 our hypotheses received support from the experimental data which indicated that advisors with self-interest deceived the client by explicit and implicit strategic behavior. However, because of the hypothetical nature of the experiment participants of Study 1 could not get the impression that their advice would really help or harm a real client. Because of this lack of accountability the results of Study 1 could have been overestimated. Further clarification therefore is needed. In order to do this, we would like to more carefully look at the concept of accountability. Accountability is an expectation (implicit or explicit) that one may be called on to justify ones actions to others (Lerner and Tetlock, 1999). In practice, advisors are in this situation to justify their recommendation and action. But how does enhanced accountability influence advisor's self-interested behavior?

One assumption could be that enhanced accountability leads to reduced self-interested behavior and consequently reduced explicit and implicit strategic behavior. Research findings can indicate that persons who are asked to justify their decisions are more likely to be interested in others outcomes. People with high endowment but having no accountability for group members contributed the same amount to a common system compared to those with few endowments. However, when they were accountable they made higher payments which helped in social dilemma situations (De Cremer and Van Dijk, 2009).

However, based on the review of Lerner and Tetlock (1999) we know it is especially necessary to take a closer look on the conditions of accountability. This review identified different conditions where accountability led to diverse outcomes in decision making and especially identified outcome vs. process accountability as crucial in this context (Lerner and Tetlock, 1999). Especially when people had to justify their outcome, such as their recommendation, they tended to increase their need of self-justification as well as biased information processing (e.g., Simonson and Staw, 1992). Contrarily accountability for decision processes led to more balanced evaluation when confronted with different alternatives. Consequently, advisors' perceived accountability for their decision and expected need to justify this outcome should also lead to enhanced bias in information processing.

An additional closer look on the conditions for accountability in advice-giving situation is provided by research of Jonas et al. (2005). This study investigated the information search and transfer of highly accountable advisors (who assumed to meet the client and have to justify the recommendation) compared to advisors without accountability for their decision. This research found an enhanced confirmation bias for advisors' binding recommendation when they were highly accountable for their decision but not in advisors without accountability. This effect could be explained by the directional goal of impression motivation. This means advisors wanted to appear in a positive light and therefore searched and also transferred primarily that information which

was in line with their preliminary decision. This strategy helped them to provide evidence for their recommendation which supported their wish to present themselves in a positive way in front of their clients.

When referring to a self-interested advisor who has the directional goal to earn money we know already from Study 1 that they will evaluate and transfer information in a biased way in order to favor their self-interest. They also provide mainly information for their self-interested recommendation. However, how do advisors process information when they perceive themselves as both self-interested as well as highly accountable? We know from Study 1 that self-interested advisors who feel motivated by the goal to receive an incentive commit themselves already with the self-interested alternative before searching, evaluating and transferring information. In Study 1 this led to a bias in explicit and implicit strategic behavior. The perception of high accountability for their decision might increase the advisors' wish to bolster their view. However, the salience of accountability might also counteract and reduce the self-interested bias in participants. Yet, given former research this latter alternative seems unlikely because being accountable for an outcome, such as a recommendation, has been shown to increase bias in information processing. Similarly, the presence of a directional goal (impression motivation) has also been shown to increase bias in information processing and information transfer. In Study 2 we investigate the influence of combining the presence of perceived accountability with self-interest on biased information processing and transfer. Therefore we tested the following hypotheses:

Explicit strategic behavior

Hypothesis 6 – Transfer of information: We suppose especially among accountable participants, that self-interested advisors transfer less conflicting information compared to advisors without self-interest, this difference should be weaker within participants who are not accountable.

Implicit strategic behavior

Hypothesis 7 – Evaluation of information: We assume among accountable participants, that self-interested advisors devalue conflicting information compared to advisors without self-interest, this difference should be weaker within participants who are not accountable.

Hypothesis 8 – Memory of information: Again we suppose especially among accountable participants, that self-interested advisors remember conflicting information less compared to advisors without self-interest, this difference should be weaker within participants who are not accountable.

Moderated mediation

Hypothesis 9 – Transfer of information: We propose the indirect effect of self-interest on explicit strategic behavior (transfer of conflicting information) through implicit strategic behavior (evaluation of conflicting information) would be stronger under high than low accountability because accountability moderates the relation between self-interest and the mediator implicit strategic behavior.

METHOD

PARTICIPANTS AND PROCEDURE

Participants were 53 students (36 female, 17 male) at a public university of Austria (University of Salzburg). The present study took place after a social psychology lecture. Psychology students could volunteer in order to receive credits for participation. The procedure in this experiment was similar to Study 1, with the following exception that we tried to manipulate accountability through the following sentence: "Please leave your e-mail address (on an extra sheet), so that the client can contact you for further questions." The condition without accountability did not have this sentence in the questionnaire. Unfortunately our attempt to additionally manipulate accountability failed, $F(3,53) = 0.36, p = 0.552$. The survey took place in a huge lecturer hall where our manipulation was to weak. Although, the manipulation of self-interest was successful, $F(3,53) = 3.16, p = 0.018$, we use perceived self-interest and perceived accountability for further analysis. We discuss this decision later with our findings.

MEASURES

Explicit strategic behavior

Again, we measured the intention to *transfer information* to the client, but used this time a five-point Likert-scale (unlikely to very likely). We used conflicting information (Cronbach's $\alpha = 0.75$) regarding self-interest for further analysis.

Implicit strategic behavior

For the *evaluation of the information* we applied a five-point Likert-scale (not relevant to very relevant). Conflicting information (Cronbach's $\alpha = 0.67$) regarding self-interest (see Table A1 in Appendix) is used for our further analysis. Further, we implemented a quiz to measure the *memorized information*, but we increased the amount of questions from 6 to 11. For further analysis we used only the conflicting information (six items, e.g., career opportunities for the mechatronic engineer, product engineer's problems with the labor market) plus one question where participants had to remember the amount of salary of the mechatronic engineer. However, this question had no correct answer alternative – there was an optimistic (more than 30,000€ per year) vs. two rather pessimistic (not even 30,000€, at the best 30,000€ per year) and one neutral (approximately 30,000€ per year) biased alternative. For our conflicting information scale we added the optimistic alternative as correct answer.

Perceived self-interest

Similar to Study 1 we combined the three subscales of hidden intention, hidden information and hidden action and used one general scale of self-interested behavior for further analysis (nine items, Cronbach's $\alpha = 0.91$).

Perceived accountability

In the past research accountability was often manipulated through justification in front of a real audience. In our case we did not have real audience but some of the participants assumed further contact with the client per e-mail (attempt of manipulation). However, this typical accountability situation should be represented through our

two questions which measures perceived accountability. ("How realistic was the situation to give advice to another person?" and "How accountable did you feel for your advice?" two items; $r = 0.40, p < 0.01$).

RESULTS

EXPLICIT STRATEGIC BEHAVIOR

Transfer of Information – Hypothesis 6

We conducted a hierarchical regression analysis in which the transfer of conflicting information was conducted by perceived accountability and perceived self-interest (main-effect terms) and the interaction term simultaneously. Following Aiken and West (1991), the variables accountability and self-interest were centered (i.e., by subtracting the mean from each score), and the interaction term was based on these centered scores. The interaction between accountability and self-interest was significant, $b = -0.14, SE = 0.06, t(49) = -2.25, p = 0.029$, and as well a main effect for self-interest revealed significance, $b = -0.34, SE = 0.08, t(49) = -4.25, p < 0.001$. Simple slope analysis was conducted to further analyze the interaction (Aiken and West, 1991). When accountability was low (1 SD below the mean), self-interest was significantly negative related to the transfer of conflicting information, $b = -0.20, SE = 0.08, t(49) = -2.37, p = 0.022$. Therefore participants with low accountability were significantly influenced by their self-interest and passed on less conflicting information. However, when accountability was perceived high (1 SD above the mean; $b = -0.49, SE = 0.12, t(49) = -4.08, p < 0.001$), the relation between self-interest and less transfer of conflicting information even increased, which means an enhanced bias when accountability was high. However, the bias already existed when accountability was low, but high accountability increased the bias significantly.

Additional data analysis showed, that this effect was similarly found regarding the general transfer of information (all information – conflicting and supportive), which indicates that among highly accountable advisors self-interest led to general withholding information [self-interest \times accountability: $b = -0.17, SE = 0.69, t(49) = -2.54, p = 0.015$, 1 SD above: $b = -0.60, SE = 0.13, t(49) = -4.66, p > 0.001$, 1 SD below: $b = -0.25, SE = 0.09, t(49) = -2.77, p > 0.001$]. These results provided evidence that among advisors with high accountability, especially high self-interest led to withhold of conflicting information, which supports our Hypothesis 6. Moreover, our results indicate that self-interested advisors withhold general information and provide less information to their clients as advisors without self-interest. In other word self-interested advisors with high accountability do not distinguish between conflicting and supporting information and withhold information in general. The slopes are plotted in Figure 4.

IMPLICIT STRATEGIC BEHAVIOR

Information evaluation – Hypothesis 7

To test the perceived accountability as moderator between self-interest and the evaluation of conflicting information, we applied the same approach as already explained. The interaction between accountability and self-interest was marginally significant, $b = -0.10, SE = 0.05, t(49) = -1.93, p = 0.060$.

Simple slope analysis was conducted to further analyze this interaction (Aiken and West, 1991). When accountability was low (1 SD below the mean), self-interest was not significantly related to the evaluation of conflicting information, $b = 0.04, SE = 0.10, t(49) = 0.53, p = 0.596$, in other words self-interest had no specific influence on the evaluation of conflicting information. However, when accountability was evaluated high [1 SD above the mean; $b = -0.16, SE = 0.10, t(49) = -1.70, p = 0.097$], self-interest was associated negatively with evaluated conflicting information. These results indicate that self-interested people under high accountability devalue information compared to low self-interested participants, whereas participants with low accountability showed a similar level of devaluation regarding conflicting information. Therefore, these results do not suppose an enhanced bias compared to low accountability, however an enhanced bias between low and high self-interest among high accountability which supports the Hypothesis 7. The slopes are plotted in Figure 5.

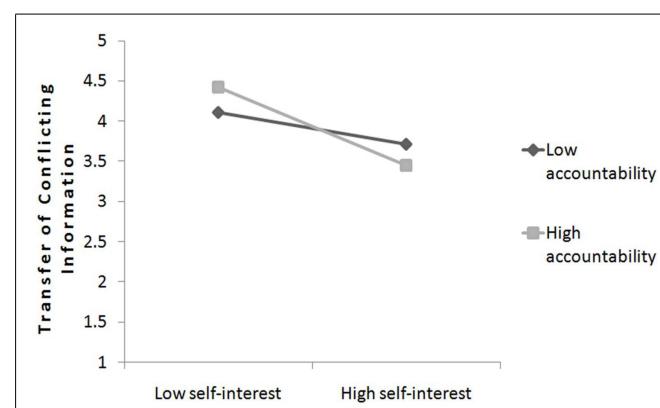


FIGURE 4 |The relationship between self-interest and transferring conflicting information as a function of advisor's perceived accountability (study 2).

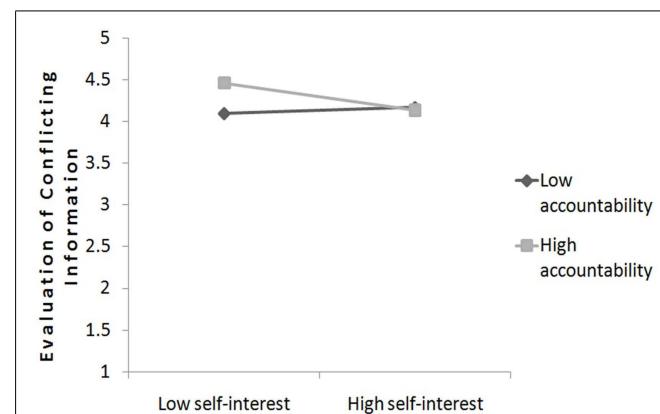
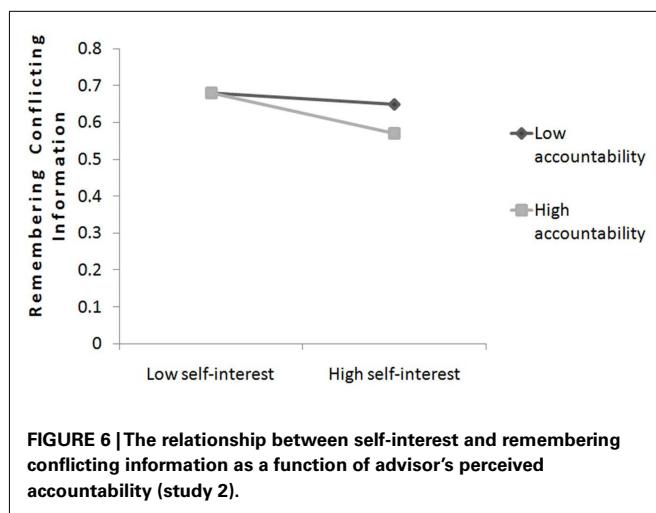


FIGURE 5 |The relationship between self-interest and evaluation of conflicting information as a function of advisor's perceived accountability (study 2).



Remembered information – Hypothesis 8

Accountability should be also tested as moderator between self-interest and the memorized conflicting information. We conducted a hierarchical regression analysis in which the memorized conflicting information was predicted by main-effect terms (perceived accountability and perceived self-interest) and the interaction term simultaneously. There was a significant main effect for self-interest, $b = -0.04$, $SE = 0.01$, $t(49) = -2.52$, $p = 0.015$ and the interaction between accountability and self-interest was marginally significant, $b = -0.02$, $SE = 0.01$, $t(49) = -1.75$, $p = 0.086$. Simple slope analysis was conducted to further analyze this interaction (Aiken and West, 1991). When accountability was low (1 SD below the mean), self-interest was not significantly related to memorized information, $b = 0.02$, $SE = 0.02$, $t(49) = -1.09$, $p = 0.283$, which imply when participants perceived themselves as less accountable self-interest had no specific influence on memory of conflicting information. However, when accountability was perceived high [1 SD above the mean; $b = -0.06$, $SE = 0.02$, $t(49) = -2.64$, $p = 0.011$], self-interest was associated significant negatively with memorized information. Therefore, among advisors with high accountability and high self-interest showed the worst memory regarding conflicting information. Accountability can be identified as marginal significant moderator which increases the self-interested bias in memorized conflicting information and therefore supports our Hypothesis 8. The slopes are plotted in Figure 6.

Moderated mediation – Hypothesis 9

We employed Preacher et al. (2007) (Model 2) bootstrapping procedure to test our moderated mediation hypothesis that the indirect effect of self-interest on explicit strategic behavior (transfer of conflicting information) through implicit strategic behavior (evaluation of conflicting information) would be stronger under high than low accountability because accountability moderates the relation between self-interest and implicit strategic behavior. As we already know the moderated regression analysis confirmed a marginal significant interaction between accountability and self-interest on implicit strategic behavior (see above Hypothesis 7). Using 1000 resample, analyses showed that implicit strategic

behavior significantly mediated the effect of perceived self-interest on explicit strategic behavior under high accountability (90% CI: -0.25 to -0.01) but not under low accountability (90% CI: -0.04 to 0.11).

DISCUSSION STUDY 2

Our results indeed showed an interaction between self-interest and accountability indicating that high accountability enhanced the effect between self-interest and explicit (transfer of conflicting information) as well as implicit strategic behavior (evaluation and memory of conflicting information). More specific, we found that self-interested advisors increased their explicit strategic behavior by withholding information in general, but high accountability in advisors without self-interest led even to a reduced bias. This mean only the combination of high self-interest and high accountability led to increase in self-interested bias. This interaction was also found regarding implicit strategic behavior. Self-interested participants devaluated conflicting information only when they perceived themselves as highly accountable. High accountability without self-interest also led again to a reduced bias. Referring to the memory performance, self-interested advisors showed generally decreased performance regarding conflicting information. However, performance was especially decreased when they also perceived themselves as accountable for the given recommendation. Our moderated mediation analysis indicated that the relation between advisor's self-interest and the explicit strategic behavior (reduced transfer of conflicting information) can be explained by implicit strategic behavior (devaluation of conflicting information). But this was only the case when accountability was high – which confirm accountability again as moderator.

Unfortunately, findings of Study 2 do not exactly replicate findings of Study 1 (self-interest \times type of information). One reason is that in participants with the concern of accountability higher responsibility was salient (attempt of manipulation), which weakened the effect regarding the experimental self-interest and participants did not differentiate between conflicting and supporting information as strong as in Study 1. However, we found convincing findings which showed that participants with high perceived accountability and without self-interest behave especially responsible for their clients regarding conflicting information – they increased transfer and evaluation of conflicting information. Thus, under high accountability without self-interest participant showed especially responsible for the client. But in combination with self-interest, the advisors acted in an even more self-interested way – they withhold and devalue information conflicting with their self-interest. These findings underline in our opinion the weakening effect of the self-interest manipulation.

GENERAL DISCUSSION

The present research examined the effect of incentives on two different forms of strategic behavior. Within two studies we could show that the promise to receive an incentive led to deception through explicit as well as implicit strategic behavior. The aim of Study 1 was to investigate the consequences of self-interest regarding information which are in conflict or in support with the self-interest. The results provided twofold evidence for *explicit strategic behavior*: firstly, self-interested advisors explicitly recommended

the self-serving job option more often compared to those without specific interest. Secondly, we could observe that advisors passed on more supporting information and withhold more conflicting information from their clients compared to participants without self-interest. In Study 2, we measured beside self-interest also advisors' perceived accountability. Our results indeed showed an interaction between self-interest and accountability regarding the transfer of conflicting information. In other words, we found that self-interested advisors increased their explicit strategic behavior by withholding conflicting information compared to advisors without self-interest when accountability was high. This was not the case when accountability was low.

Our findings regarding the explicit strategic behavior were in line with the described strategic behavior of PAT (Ross, 1973) which especially predicts "hidden information" as potential risk in relationships where information is distributed asymmetrically and the two parties have conflicting goals. Similar, Steinel and De Dreu's (2004) findings showed that participants were less accurate when confronted with a competitive counterpart with opposed interests. However, our self-interested participants used withholding conflicting information and passing on supporting information as method to pursue self-interest and to bolster the self-interested decision. Similar in the study of Steinel and De Dreu (2004) this behavior could be observed to primarily handicap the other person and to enrich oneself. Our results indicate that advisors are motivated by the possibility to receive an incentive and therefore transfer information strategically and give strategic recommendation.

Furthermore, we provide evidence for *implicit strategic behavior*, which has so far not been investigated in past research. Thus, it is highly relevant to look at deception in its entirety – this means beside deception as explicit behavior also as bias in information processing. Referring to our results self-interested advisors were biased implicitly which again could be identified twofold: firstly, participants with self-interest evaluated information less in favor of clients' needs compared to the control group. The interaction effect between self-interest and the information type (supporting vs. conflicting regarding self-interest) showed that advisors without self-interest wanted to find the best solution for the client and therefore enhanced evaluation of the conflicting information (supporting for the client) compared to the supporting information (conflicting for the client). This pattern disappeared in advisors with self-interest, who seemed not to take the perspective of the client and his needs into account. Secondly, self-interested advisors even remembered highly conflicting information worse than advisors without self-interest. Interestingly, the biased memory performance can also be explained by the evaluation of conflicting information. Among those participants who especially devalued the conflicting information in advance, high self-interest could significantly predict the bad memory regarding the conflicting information. This means Study 1 could provide first interesting evidence for implicit strategic behavior.

However, in Study 2 the interaction between self-interest and accountability was beside explicit strategic behavior also found in implicit strategic behavior. More specific, we stated in our analysis that self-interested advisors decreased the evaluation of conflicting information compared to advisors with low self-interest when accountability was high. There was no difference between high

and low self-interest when accountability was low. Regarding memory performance of conflicting information also only high accountable participants showed a significant difference between high and low self-interest. Self-interested with high accountability could remember conflicting information worse. Taken together, our results supported the importance of implicit strategic behavior. Deception in advice-giving situation is also driven by biases in information processes like evaluating and remembering information which is even increased when accountability is high.

Our results regarding the implicit strategic behavior provide further support for Kunda's (1990) assumption of biased information processing in favor of one's wishes and desires – or in the current study to earn the incentive and pursue self-interest. So far, research provided evidence for self-interested participants to devalue arguments as less persuasive when its content was against their self-interest (Darke and Chaiken, 2005). We could additionally provide evidence that the self-interested information evaluation interacted with self-interest and led to worse memory performance regarding conflicting information. This means that especially those who devalued already conflicting information in advance remembered this information worse. In other words, these results suggested that bias in memory arises especially when information is not compatible with the self-interest and is evaluated therefore more negatively.

The findings of Study 2 additionally indicate the enhanced influence of self-interest on strategic behavior when accountability is high. So far past research showed under high accountability advisors' search was more confirmation based and in line with preliminary decision (Jonas et al., 2005). And this research also identified impression motivation as directional goal why people are motivated to bias information. Our current results provided evidence that receiving an incentive function as a directional goal and led to a bias. The combination of high self-interest and high accountability enhanced this bias and led to a higher extant of strategic behavior. Highly accountable advisors seem to bias their information transfer in order to convince the client of the self-interested alternative.

Interestingly, in both studies we could confirm that implicit strategic behavior can predict to some extent explicit strategic behavior. Our mediation analysis indicated that the relation between advisor's self-interest and the explicit strategic behavior (reduced transfer of conflicting information) can be explained by implicit strategic behavior (devaluation of conflicting information). In Study 2 this was also the case but only when accountability was high – which highlighted accountability as moderator again. Both mediation analyses are nice evidence that implicit actions explain partly the process of explicit strategic behavior. In other words, advisors might to some extent deceive themselves through biased information processing to justify their explicit strategic behavior afterward. The interesting findings of Shalvi et al. (2011) support this view: they found that the degree of lying depends on the extant of possible self-justification which participants emerge through biased information processing. This means in our study that especially the implicit strategy (the devaluation of conflicting information) justifies in turn the explicit strategic behavior (withholding of conflicting information). According to these results we must suppose that the promise of incentives can lead advisors to implicit strategic behavior which in turn leads to explicit strategic

behavior. We have to take implicit actions more into account in order to understand explicit strategic behavior and deception. We will discuss especially the implications of this finding later.

Additionally to past research our results provide important evidence of the implicit strategic behavior which shows that advisors are influenced by their evaluation and memory. These are processes which advisors themselves can hardly control. Former research of such implicit processes also defined the term "directed forgetting" which especially explains reduced retrieval of unwanted memories or information (e.g., Freud, 1900/1964). However, this phenomenon should not lead to permanent damage of the information. Therefore, for future research it would be essential not only to test the recall of information but also if it can be recognized again (for overview: Baddeley et al., 2009). A further implicit phenomenon is the attention and which might be also directed through our motives. Isaacowitz (2006) exactly discusses this and describes attention as a tool of motivation. Eye-tracking studies provide some evidence that people are often strategic in their attentional preference and he assumed that "people guide their attention to information that can help them to achieve their goals and put away from stimuli that will not" (Isaacowitz, 2006, p. 68). Bias in attention might be also relevant in our study where especially self-interested participants could have used their attention as tool to guide their self-interested intention. For future research, especially eye-tracking studies can help us to understand how much attention self-interested participant pay to supporting vs. conflicting information and with how much effort they try to understand the match between the applicant and the different job alternatives.

With regard to theoretical implications, these findings identified a new aspect of advice-giving, because strategic behavior and deception was hardly discussed in advisor-client research (for overview: Bonaccio and Dalal, 2006). Although we know that clients accept and use advice of self-interested advisors to a lesser extent compared to advisors without specific interest (Jodbauer and Jonas, 2011) and that besides advisor's expertise and confidence also advisor's good intention is highly relevant when evaluating the advisor's recommendation quality (Bonaccio and Dalal, 2009). One exception is the research of Van Swol (2009) who manipulated two different motives – persuasion vs. quality – during the advice-giving process and could show that advisor's motive to persuade manifested in using a high public confidence rating. They did that in a strategic way to convince the client because the private confidence rating differed significantly. Interestingly this attempt was successful in order to pursue clients. Our present research can confirm that advisors behave strategically. However, our results provide an extension of previous research and suggest that strategic behavior has an explicit and an implicit facet. Finally, we can state that the implicit strategic behavior is crucial because it can partly explain the explicit strategic behavior.

LIMITATIONS

The reader should be aware that in Study 2 our manipulation for accountability did not work. Therefore, our simple slope analyses are also based on correlative data (including also

self-interest, experimental self-interest was less convincing). As already discussed, one reason might be that in participants confronted with the manipulation higher responsibility was salient, which weakened the effect regarding the experimental self-interest and differentiation between conflicting and supporting information compared to Study 1. However, we could find convincing findings with simple slope analysis, which is a state-of-the-art analysis for moderation effects. Still, it is a limitation to use correlative data because there can be confounds which we are not aware of. Therefore, in future research it will be essential to manipulate accountability and at the same time self-interest successfully, so that results can be based on experimental manipulation.

A further limitation is that we used students and not real advisors. There is for example evidence, that real experts (physicians) search longer for an alternative explanation and could therefore reduce errors compared to novice (Krems and Zierer, 1994). However, there are also "costs of expertise": Experts who decided for an alternative were more rigid and did not change their decisions easily (Sternberg, 1996). These findings indicate that especially for the practical implications it would be essential to test our hypothesis also with real advisors. Furthermore, regarding incentives it can be further essential to test an incentive that is common in this business and the real field of advisors.

PRACTICAL IMPLICATIONS AND FUTURE DIRECTIONS

In many advisor-client interactions incentives as explicit motivator are part of the business. Companies want to control the interests of the advisors and match them with their interests. For instance, even physicians, who are highly responsible for their clients, are in this situation. When physicians are rewarded with gifts or even get paid when supporting the interests of pharmaceutical companies (e.g., recommending a certain medication, referrals to clinical trials) they are at risk to behave strategically. This approach implies the risk that advisors clearly and explicitly subordinate the needs of the customer to their self-interest. Well, the explicit self-interested behavior aroused by incentives is known and in a way desired in this business sector of the pharmaceutical companies. Furthermore, for physicians this explicit strategic behavior seems maybe controllable and they feel not influenced in their objectivity. But based on our results self-interest is not limited to explicit and conscious acting. Moreover, the influence of incentives goes a step further and already influences their evaluation when searching and thinking about the best medication for their client and moreover they can later remember conflicting information regarding the medication worse. Our findings strongly indicate that advisors do not act independently of their more implicit processes of information processing. The implicit strategic behavior entail a high risk for clients and also for advisors' themselves. It might be especially crucial how incentives are used. The promise of incentive connected with a certain alternative or product showed in our study evidence for deception. Based on the use of incentive within this study, strategic behavior might be especially high because of the connection between the incentive and a certain alternative (product engineer) or product, such as a certain medication. Further research in this field would be necessary to investigate different forms of providing

incentives and how this lead to explicit, and implicit strategic behavior.

CONCLUSION

In order to improve the understanding of deception our results indicated to take explicit and implicit strategic behavior into account. Advisors gave recommendation and transfer of information in self-interested strategic manner to deceive the client. The advisors also biased the information processing which can be seen as an implicit strategic way to deceive the client.

REFERENCES

- Aiken, L. S., and West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park: Sage.
- Baddeley, A., Eysenck, M. W., and Anderson, M. C. (2009). "Motivated forgetting," in *Memory*, eds A. Baddeley, M. W. Eysenck, and M. C. Anderson (New York: Psychology Press), 217–244.
- Boles, T. L., Croson, R. T. A., and Murnighan, J. K. (2000). Deception and retribution in repeated ultimatum bargaining. *Organ. Behav. Hum. Decis. Process.* 83, 235–259.
- Bonacchio, S., and Dalal, R. S. (2006). Advice taking and decision-making: an integrative literature review, and implications for the organizational sciences. *Organ. Behav. Hum. Decis. Process.* 101, 127–151.
- Bonacchio, S., and Dalal, S. D. (2009). Evaluating advisors: a policy-capturing study under conditions of complete and missing information. *J. Behav. Decis. Mak.* 23, 227–249.
- Buller, D. B., and Burgoon, J. K. (1994). "Deception: strategic and nonstrategic communication," in *Strategic Interpersonal Communication*, eds J. A. Daly and J. M. Wiemann (Hillsdale: Erlbaum), 191–223.
- Buller, D. B., and Burgoon, J. K. (1996). Interpersonal deception theory. *Commun. Theory* 6, 203–242.
- Burgoon, J. K., Buller, D. B., Ebisu, A., Rockwell, P., and White, C. (1996). Testing interpersonal deception theory: effects of suspicion on nonverbal behavior and relational messages. *Commun. Theory* 6, 243–267.
- Darke, P. R., and Chaiken, S. (2005). The pursuit of self-interest: self-interest bias in attitude judgment and persuasion. *J. Pers. Soc. Psychol.* 89, 864–883.
- De Cremer, D., and Van Dijk, E. (2009). Paying for sanctions in social dilemmas: the effects of asymmetry of endowments and accountability. *Organ. Behav. Hum. Decis. Process.* 109, 45–55.
- Ditto, P. H., Scepansky, J. A., Munro, G. D., Apanovitch, A. M., and Lockhart, L. K. (1998). Motivated sensitivity to preference-inconsistent information. *J. Pers. Soc. Psychol.* 75, 53–69.
- Dunning, D. (1999). A new look: motivated social cognition and the schematic representation of social concepts. *Psychol. Inq.* 10, 1–11.
- Eisenhardt, K. (1989). Agency theory: an assessment and review. *Acad. Manage. Rev.* 14, 57–74.
- Freud, S. (1900/1964). *The Interpretation of Dreams*. New York: Basic Books.
- Isaacowitz, D. M. (2006). Motivated gaze: the view from the gazer. *Curr. Dir. Psychol. Sci.* 15, 68–72.
- Jodlbauer, B., and Jonas, E. (2011). Forecasting clients' reactions: how does the perception of strategic behavior influence the acceptance of advice? *Int. J. Forecast.* 27, 121–133.
- Johnson, B. T., and Eagly, A. H. (1989). The effects of involvement on persuasion: a meta-analysis. *Psychol. Bull.* 106, 290–314.
- Jonas, E., and Frey, D. (2003). Information search and presentation in advisor-client interactions. *Organ. Behav. Hum. Decis. Process.* 91, 154–168.
- Jonas, E., Schulz-Hardt, S., and Frey, D. (2005). Giving advice or making decisions in someone else's place – the influence of impression, defense and accuracy motivation on the search for new information. *Pers. Soc. Psychol. Bull.* 31, 977–990.
- Jonas, E., Traut-Mattausch, E., Frey, D., and Greenberg, J. (2008). The path or the goal? Decision vs. information focus in biased information seeking after preliminary decisions. *J. Exp. Soc. Psychol.* 44, 1180–1186.
- Jungermann, H. (1999). "Advice-giving and taking," in *Proceedings of the 32nd Hawaii International Conference on System Sciences (HICSS-32)*. [CD-ROM]. (Maui: Institute of Electrical and Electronics Engineers, Inc.). Available at: <http://www.gp.tu-berlin.de/Users/jl/jungermann/Publications/Advice%20giving.html>
- Krems, F., and Zierer, C. (1994). Sind Experten gegen kognitive Täuschungen gefeit? Zur Abhängigkeit des confirmation bias von Fachwissen. *Z. Exp. Angew. Psychol.* 41, 98–115.
- Kruglanski, A. W. (1989). *Lay epistemics and human knowledge: Cognitive and motivational bases*. New York: Plenum Press.
- Kruglanski, A. W., Bélanger, J. J., Chen, X., Köpetz, C., Pierro, A., and Manetti, L. (2012). The energetic of motivated cognition: a force-field analysis. *Psychol. Rev.* 119, 1–20.
- Kunda, Z. (1987). Motivated inference: self-serving generation and evaluation of causal theories. *J. Pers. Soc. Psychol.* 53, 636–647.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychol. Bull.* 108, 480–498.
- Kunda, Z., and Sanitioso, R. (1989). Motivated change in the self-concept. *J. Exp. Soc. Psychol.* 25, 272–285.
- Lerner, J. S., and Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychol. Bull.* 125, 255–275.
- McDonald, H. E., and Hirt, E. R. (1997). When expectancy meets desire: motivational effects in reconstructive memory. *J. Pers. Soc. Psychol.* 72, 5–23.
- Petty, R. E., and Cacioppo, J. T. (1990). Involvement and persuasion: tradition versus integration. *Psychol. Bull.* 107, 367–374.
- Preacher, K. J., Rucker, D. D., and Hayes, A. F. (2007). Addressing moderated mediation hypotheses: theory, methods, and prescriptions. *Multivar. Behav. Res.* 42, 185–227.
- Ross, S. A. (1973). The economic theory of agency: the principal's problem. *Am. Econ. Rev.* 62, 134–139.
- Sanitioso, R., Kunda, Z., and Fong, G. T. (1990). Motivated recruitment of autobiographical memories. *J. Pers. Soc. Psychol.* 59, 229–241.
- Shalvi, S., Dana, J., Handgraaf, M. J. J., and De Dreu, C. K. W. (2011). Justified ethicality: observing desired counterfactuals modifies ethical perceptions and behavior. *Organ. Behav. Hum. Decis. Process.* 115, 181–190.
- Simonson, I., and Staw, B. M. (1992). Deescalation strategies: a comparison of techniques for reducing commitment to loosing courses of action. *J. Appl. Psychol.* 77, 419–426.
- Steinel, W., and De Dreu, C. K. W. (2004). Social motives and strategic misrepresentation in social decision making. *J. Pers. Soc. Psychol.* 86, 419–434.
- Sternberg, R. J. (1996). "Costs of expertise," in *The Road to Excellence*, ed. K. A. Ericsson (Mahwah: Lawrence Erlbaum Associates), 347–354.
- Stone, D. N., and Ziebart, D. A. (1995). A model of financial incentive effects in decision making. *Organ. Behav. Hum. Decis. Process.* 61, 250–261.
- Taber, C. S., and Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *Am. J. Pol. Sci.* 50, 755–769.
- Van Swol, L. M. (2009). The effects of confidence and advisor motives on advice utilization. *Commun. Res.* 36, 857–873.

Furthermore, the fact that the advisor should justify his/her recommendation even increased strategic behavior – explicit as well as implicit.

ACKNOWLEDGMENTS

We are grateful to Susanne Jodlbauer, Isabella Uhl, Jochim Hansen, and Alexandra Brune for their helpful comments on an earlier version of this manuscript. We would also like to thank Dmitrij Agroskin for his helpful support with the moderated mediation analysis.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 17 July 2012; accepted: 07 November 2012; published online: 04 December 2012.

Citation: Mackinger B and Jonas E (2012) How do incentives lead to deception in advisor-client interactions? Explicit and implicit strategies of self-interested deception. Front. Psychology 3:527. doi: 10.3389/fpsyg.2012.00527

This article was submitted to Frontiers in Cognitive Science, a specialty of Frontiers in Psychology.

Copyright © 2012 Mackinger and Jonas. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.

APPENDIX

Table A1 | Information regarding the job alternatives differentiating between supporting and conflicting information regarding the client's and the advisor's interest.

Information regarding...	Mechatronic engineer		Product engineer		Machinery engineer	
	Client	Advisor ^a	Client	Advisor	Client	Advisor
Job characteristics	+	- ^b	-	-	+	-
Career opportunities	Neutral	Neutral	+	+	+	-
Further professional development	+	-	+	+	+	-
Salary	+	-	-	-	-	+
Competence needs	+	-	-	-	-	+
Labor-market situation	+	-	+	+	-	+

^aFor further analysis supporting and conflicting refer always to the advisor's interest;

^b+ supporting, - conflicting.