Acta Psychologica 144 (2013) 224–231

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/ locate/actpsy

Reaction time measures in deception research: Comparing the effects of irrelevant and relevant stimulus–response compatibility

Kristina Suchotzki [a,*], Bruno Verschuere [a,b,c], Geert Crombez [a], Jan De Houwer [a]

[a] Department of Experimental-Clinical and Health Psychology, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium
[b] Department of Clinical Psychology, University of Amsterdam, Weesperplein 4, 1018 XA Amsterdam, The Netherlands
[c] Faculty of Psychology and Neuroscience, Maastricht University, Universiteitssingel 5, 6229 ES Maastricht, The Netherlands

ARTICLE INFO

Article history:
Received 12 October 2012
Received in revised form 17 June 2013
Accepted 20 June 2013
Available online 7 August 2013

PsycINFO classification:
2300 Human Experimental Psychology
2340 Cognitive Processes
4200 Forensic Psychology & Legal Issues

Keywords:
Stimulus–response compatibility
Reaction times
Deception
Criminal intent
Differentiation of Deception

ABSTRACT

Evidence regarding the validity of reaction time (RT) measures in deception research is mixed. One possible reason for this inconsistency is that structurally different RT paradigms have been used. The aim of this study was to experimentally investigate whether structural differences between RT tasks are related to how effective those tasks are for capturing deception. We achieved this aim by comparing the effectiveness of relevant and irrelevant stimulus–response compatibility (SRC) tasks. We also investigated whether an intended but not yet completed mock crime could be assessed with both tasks. Results showed (1) a larger compatibility effect in the relevant SRC task compared to the irrelevant SRC task, (2) for both the completed and the intended crime. These results were replicated in a second experiment in which a semantic feature (instead of color) was used as critical response feature in the irrelevant SRC task. The findings support the idea that a structural analysis of deception tasks helps to identify RT measures that produce robust group effects, and that strong compatibility effects for both enacted crimes as well as merely intended crimes can be found with RT measures that are based on the manipulation of relevant SRC.

1. Introduction

Reaction time (RT) measures are popular in psychological research, partly because they are cheap, quick and easy to apply. Unfortunately, when it comes to measuring deception,[1] the findings concerning the validity of RT measures are mixed. Some researchers have found negative or inconsistent results (e.g., Engelhard, Merckelbach, & van den Hout, 2003; Gronau, Ben-Shakhar, & Cohen, 2005), while others have found results supporting the validity of RT measures (e.g., Allen, Iacono, & Danielson, 1992; Seymour, Seifert, Shafto, & Mosmann, 2000; Verschuere, Crombez, Degroote, & Rosseel, 2010). Verschuere and De Houwer (2011) argued that these inconsistencies might be related to differences in the structural characteristics of the used tasks. More specifically, they pointed out that some, but not all RT deception tasks involve a manipulation of stimulus–response compatibility (SRC).

In SRC tasks, the compatibility between a feature of the stimulus and the response is manipulated over trials. On compatible trials, a stimulus feature is related to the correct response, whereas on incompatible trials, it is related to the incorrect response. This manipulation typically leads to a compatibility effect, such as shorter RTs and fewer errors in compatible compared to incompatible trials (De Houwer, 2003, 2011). A prototypical example of an SRC task is the (spatial) Simon task (Simon, 1990; Simon & Rudell, 1967). In this task, a colored stimulus is presented either on the left or on the right side of a computer screen. Participants are instructed to press a left or right button depending on the color of the stimulus. Although irrelevant for the task, RTs and errors are influenced by the spatial position of the stimulus, resulting in shorter RTs and fewer errors for button presses to stimuli presented on the same side (i.e., compatible trials) compared to button presses to stimuli on the opposite side (i.e., incompatible trials).

In line with the taxonomy of Kornblum (Kornblum, Hasbroucq, & Osman, 1990; Kornblum & Lee, 1995), De Houwer (2003) distinguished between two types of SRC, depending upon whether the incompatibility involves a task-irrelevant or a task-relevant stimulus feature. In an irrelevant SRC task, such as the spatial Simon task described above, the incompatibility is related to a task-irrelevant stimulus feature (i.e., spatial position) that does not need to be processed in order for the task to be performed (i.e., respond to color). In contrast, in a relevant SRC task the incompatibility is related to a task-relevant stimulus

* Corresponding author at: Department of Experimental Clinical and Health Psychology, Ghent University, Henri Dunantlaan 2, B-9000 Ghent, Belgium. Tel.: +32 9 264 94 46; fax: +32 9 264 64 89.
E-mail addresses: Kristina.Suchotzki@UGent.be (K. Suchotzki), B.J.Verschuere@uva.nl (B. Verschuere), Geert.Crombez@UGent.be (G. Crombez), Jan.DeHouwer@UGent.be (J. De Houwer).

[1] Note that with 'measuring deception' we refer in a very broad sense to all methods that aim to detect or measure the truth about statements, facts, or someone's involvement in an event. It does not necessarily mean that the task itself requires that participants have the aim to deceive.

0001-6918/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.actpsy.2013.06.014

feature that always needs to be processed in order for the task to be performed. Extending the previous example of the spatial Simon task, participants may be instructed to respond not to the color, but directly to the spatial position of the stimulus. For instance, they could be asked to press the button on the same side as the stimulus in one block (i.e., compatible trials) and to press the button on the opposite side of the stimulus in another block (i.e., incompatible trials; Fitts & Seeger, 1953; Kornblum & Lee, 1995). The task-relevant feature of the stimuli (i.e., spatial position) is then either compatible or incompatible with the to-be-emitted responses. In this kind of SRC task, it is also typical to observe shorter RTs and fewer errors for the compatible trials compared to the incompatible trials.

Using this framework, Verschuere and De Houwer (2011) reviewed deception studies that used RTs. They found that the available paradigms could be classified as either those that did not manipulate SRC at all, or those that manipulated relevant SRC. Considering the wide use of irrelevant SRC tasks in many areas of psychology (e.g., the aforementioned Simon task or the Stroop task; Stroop, 1935; MacLeod, 1991), it is surprising that they did not find any studies using irrelevant SRC tasks in a deception context. Therefore, their conclusion that RT tasks that manipulate SRC produce robust and large effects was restricted to relevant SRC tasks. They could only speculate about potential differences in the effectiveness of relevant and irrelevant SRC deception tasks. Hence, in our study, the primary objective was to experimentally investigate whether structural differences between RT tasks are related to how effective those tasks are when it comes to capturing deception. We achieved this aim by comparing the effectiveness of relevant and irrelevant SRC tasks.

A secondary objective of the study was to extend deception research from past to intended behavior. Thus far, most research has focused on deception about already executed activities. Only recently have researchers started to examine the possibilities of detecting planned events. For instance, Agosta, Castiello, Rigoni, Lionetti, and Sartori (2011a) used the autobiographical Implicit Association Test (aIAT; Sartori, Agosta, Zogmaister, Ferrara, & Castiello, 2008), a RT task based on the manipulation of relevant SRC, to successfully distinguish between true and false intentions (e.g., 'Tonight I plan to sleep in Padua' vs. 'Tonight I plan to sleep in Milan'). Few studies have addressed intended criminal behavior. For instance, Vrij, Leal, Mann, and Granhag (2011) found that only one of two verbal indices of deception discriminated between true and false verbally reported intentions. Meijer, Verschuere, and Merckelbach (2010) found that skin conductance allowed to detect concealed information in participants who intended to commit a mock crime, with accuracy paralleling that obtained in participants who actually committed the mock crime. Using ERPs, Meixner and Rosenfeld (2011) were able to detect knowledge about a planned mock terrorist attack. These findings suggest that intended criminal behavior can be detected, albeit probably to a lesser extent than actually executed behavior.

In the current study, we used a RT paradigm based on the 'Sheffield Lie Test' (Spence et al., 2001), in which participants give speeded Yes/No answers to simple (mock crime-related) questions. We expected a compatibility effect in both a relevant as well as an irrelevant version of the same task, due to interference on incompatible trials between the response elicited by the stimulus and the response required by the task (Hypothesis 1). Verschuere and De Houwer (2011) argued that the effects of an irrelevant deception SRC task might be smaller than those of a relevant SRC task, because participants might on some trials succeed in ignoring task-irrelevant stimulus features (see also De Houwer, 2009). We therefore predicted bigger compatibility effects in our relevant SRC task compared to the irrelevant task (Hypothesis 2). We also implemented attention control trials, that is, trials on which participants were asked to repeat the question after it had disappeared from the screen. On the one hand, such trials should encourage participants to process the stimulus content in the irrelevant SRC task, which should in turn increase the chances of finding a compatibility effect in this task. On the other hand, the extent to which participants can repeat the questions can function as an indication of the level of stimulus processing in both tasks. As such, these trials allow us to test the idea that a possible difference in the effectiveness of the relevant and irrelevant SRC tasks in capturing the true answer to the questions is due to a difference in the extent to which the meaning of the stimuli is processed in both tasks. Finally, we expected significant compatibility effects for both the enacted as well as the intended mock crime (Hypothesis 3).

2. Experiment 1

2.1. Method

2.1.1. Participants

To obtain a sample of at least 20 participants, we invited 30 undergraduate students of Ghent University. Three participants canceled their appointment, and 27 students participated for partial fulfillment of course requirements. All provided written informed consent. Two participants were excluded from further data-analyses because they exceeded the mean error rate per subject plus 2.5 standard deviations ($M = 7.42\%$, $SD = 10.12$). The mean age of the remaining 25 participants was 18.16 ($SD = 1.07$). Most participants ($n = 19$) were women.

2.1.2. Mock crime procedure

We used two different mock crimes in this experiment. The order of the mock crimes was counterbalanced across participants. The instructions (provided on paper) for the first mock crime were: 'Leave the room, turn left and walk straight through until you reach a glass door. Go through that door, turn left and find the mail room of the department. Look for the post box of professor (…)[2] and steal a CD-ROM with exam questions that was left there.' The instructions for the second mock crime were: 'Leave the room, turn left and walk straight through until you see the elevators on your left side. Take the elevator to the first floor, get out and immediately turn left. Find the printer room of the department, look for a USB stick that was left there on the fridge and steal it.'

After entering the experiment room, participants were informed by the experimenter that they had to plan and commit two mock crimes before completing two RT computer tasks in the context of lie detection research. Participants then received the instructions for the first mock crime. After reading the assignment, participants were instructed to think about how exactly they would proceed in committing the mock crime. They were also instructed to write down the most important steps of their plan.[3] After completing the first crime, participants returned to the laboratory, and received the instructions for the second crime. Once planning was completed and participants were about to leave the laboratory to execute the second crime, they were informed that there was a slight change in the procedure: They first had to complete a computer task before executing the second crime. In fact, participants did not execute the second crime. They were debriefed after the computer task. Thus, all participants planned two crimes, but enacted only one.

2.1.3. Stimuli

For both tasks, the same four categories of crime-related questions were used, with 15 different (positive formulated) questions in each category (see Table 1). The questions concerned both the enacted mock crime and the planned mock crime. In order to avoid the true answer always being 'yes', we also included control questions for both categories concerning mock crimes which they neither planned nor committed.

[2] The actual name of one of the university professors was provided in the instructions.

[3] On average, participants wrote down 4.64 steps ($SD = 1.80$). There were no significant differences between the two mock crimes and also no significant order effect, t's < 1.59. Most participants repeated the most crucial details of the mock crime (e.g., 'Take elevator to the first floor') or general tactics on how to proceed (e.g., 'Try not to look suspicious').