Reaction Time Measures in Deception Research:
Comparing the Effects of Irrelevant and Relevant Stimulus-Response Compatibility

Kristina Suchotzki[1], Bruno Verschuere[2,1,3], Geert Crombez[1], and Jan De Houwer[1]

[1]Department of Experimental-Clinical and Health Psychology, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium

[2]Department of Clinical Psychology, University of Amsterdam, Weesperplein 4, 1018 XA Amsterdam, The Netherlands

[3]Faculty of Psychology and Neuroscience, Maastricht University, Universiteitssingel 5, 6229 ES Maastricht, The Netherlands

Email: Kristina.Suchotzki@UGent.be, B.J.Verschuere@uva.nl, Geert.Crombez@UGent.be, Jan.DeHouwer@UGent.be

Corresponding author:

Kristina Suchotzki, Department of Experimental Clinical and Health Psychology, Ghent University, Henri Dunantlaan 2, B-9000 Ghent, Belgium, Email: Kristina.Suchotzki@UGent.be, Phone: 0032 9 264 94 46, Fax: 0032 9 264 64 89

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.

Keywords: Stimulus-Response Compatibility; Reaction Times; Deception; Criminal Intent; Differentiation of Deception

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 et al., 2003; Gronau et al. 2005), while others have found results supporting the validity of RT measures (e.g., Allen et al., 1992; Seymour et al., 2000; Verschuere et al., 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 and Rudell, 1967; Simon, 1990). 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).

[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.

In line with the taxonomy of Kornblum (Kornblum et al., 1990; Kornblum and 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 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 and Seeger, 1953; Kornblum and 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 et al. (2011a) used the autobiographical Implicit Association Test (aIAT; Sartori et al., 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 et al. (2011) found that only one of two verbal indices of deception discriminated between true and false verbally reported intentions. Meijer et al. (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 task 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

[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').

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.

- INSERT TABLE 1 ABOUT HERE -

The questions were presented in Arial font in the center of a black 17-inch computer screen using Inquisit Software 3.0.1, which recorded RTs with millisecond accuracy (De Clercq et al., 2003). Reminder labels for the possible 'yes' and 'no' responses were presented left and right below the questions.

2.1.4. RT measures: Relevant and irrelevant SRC tasks

The questions were initially presented in white. After 700 milliseconds, the color of the questions changed to one of four different colors (purple or grey in the irrelevant SRC task, yellow or blue in the relevant SRC task)[4]. Within each task, questions were presented randomly intermixed. The inter-trial interval varied randomly between intervals of 500 ms, 700 ms, 900

[4] Different colors were used in the two tasks, in order to avoid transfer effects from one task to another. Within the tasks, the assignment of colors to responses was counterbalanced across participants.

ms, 1100 ms, and 1300 ms (*M* = 770 ms). In both tasks, participants gave left or right key presses (keys 4 and 6 on a standard AZERTY keyboard) with their right hand to indicate 'yes' or 'no' responses. The assignment of the keys was counterbalanced across participants. To ensure that the 'yes' and 'no' meaning of the keys was learned properly, we included catch trials in both tasks, in which participants only had to react to the words 'Yes' or 'No' presented centrally on the screen by pressing the corresponding key. To assess the degree to which participants processed the meaning of the questions, we included attention control questions. In these trials, the questions did not change to another color, but disappeared instead. Immediately thereafter, the question 'What was the last question?' appeared on the screen. Participants had to write down their answer, and continue with a key press. In total, each of the two tasks consisted of 135 trials: 120 crime and control questions, 10 Yes/No catch trials, and 5 attention control questions.

In the *irrelevant* SRC task, participants responded to the color of the stimuli. For example, they had to respond 'yes' to grey questions and 'no' to purple questions. This is an irrelevant SRC task, because it involves a manipulation of the compatibility between an irrelevant stimulus feature (i.e., the true answer to the question) and the meaning of the response (i.e., 'yes' or 'no'). Hence, when participants are instructed to respond 'yes' to grey questions and 'no' to purple questions, compatible trials were trials in which true questions were presented in grey and false questions in purple. Incompatible trials were those in which true questions were presented in purple or false questions in grey.

In the *relevant* SRC task, participants responded to the truth value of the stimuli. For example, they answered truthfully when the question was presented in yellow and untruthfully when the question was presented in blue. This is a relevant SRC task, because it involves a manipulation of the compatibility between a relevant stimulus feature (i.e., the true answer to the

question) and the meaning of the responses (i.e., 'yes' or 'no'). Compatible trials were those in which participants answered truthfully (i.e., responded 'yes' to true questions and 'no' to false questions) while incompatible trials were those in which they lied (i.e., responded 'yes' to false questions and 'no' to true questions).

The order of the two RT tasks was counterbalanced across participants: 14 participants first completed the relevant SRC tasks and 11 participants first completed the irrelevant SRC task.

2.2. Results

Mean error rates and RTs are shown in Figure 1. Both were analyzed with a 2 (Task: relevant vs. irrelevant SRC task) x 2 (Compatibility: compatible vs. incompatible trials) x 2 (Enactment: enacted vs. intended crime) repeated measures ANOVA.[5] For the RT analysis, trials with errors were discarded. To reduce the impact of outlying values, RTs more than 2.5 standard deviations from the mean per subject and task (Ratcliff, 1993) were removed from the RT analysis (2.75 % of all trials with correct responses). As a measure of effect size the standardized mean difference *d* was calculated. As a rule of thumb, Cohen (1988) proposed .20, .50 and .80 as thresholds for 'small', 'moderate' and 'large' effects.[6] As we used a repeated measures design, we corrected *d* for intercorrelations (Cohen's *d* for paired data; e.g., Dunlap et al., 1996; Morris & Deshon, 2002)[7].

- INSERT FIGURE 1 ABOUT HERE -

[5] Because there was no significant interaction of Question type (enacted/intended crime versus control crime) with Task and Compatibility, this factor was not included in the reported analyses.

[6] For the main effects, a positive Cohen's *d* indicates an effect in the following directions: $M_{\text{incompatible trials}} > M_{\text{compatible trials}}$, $M_{\text{relevant SRC task}} > M_{\text{irrelevant SRC task}}$, $M_{\text{intended crime}} > M_{\text{enacted crime}}$.

[7] We calculated the effect size *d* using the following formula: $d = M_1\text{-}M_2/\sqrt{SD_1^2 + SD_2^2 - 2*r_{12}* SD_1* SD_2}$

2.2.1. Error rate

Most importantly, the ANOVA on error percentage showed significant main effects of Task, F (1, 24) = 71.44, $p < .001$, $d = 1.69$, and Compatibility, $F(1, 24) = 25.89$, $p < .001$, $d = 1.02$, and a significant interaction of Task x Compatibility, $F(1, 24) = 28.19$, $p < .001$, $d = 1.06$. A significant compatibility effect (less errors in compatible compared to incompatible trials) was observed in the relevant SRC task, $t(24) = 5.39$, $p < .001$, $d = 1.08$, but not in the irrelevant SRC task, $t(24) = 1.06$, $p = .30$, $d = 0.21$.

Furthermore, we observed a main effect of Enactment, $F(1, 24) = 6.40$, $p = .02$, $d = 0.51$. The interaction of Task x Enactment, $F(1, 24) = 11.48$, $p < .01$, $d = 0.68$, indicated that participants made more errors in the intended crime condition compared to the enacted crime condition in the relevant SRC condition, $t(24) = 3.06$, $p < .01$, $d = 0.61$, but more errors in the enacted crime condition compared to the intended crime condition in the irrelevant SRC condition, $t(24) = 2.50$, $p < .05$, $d = 0.50$. There were no other significant interaction effects, F's < 1.65.

Follow-up analyses were conducted to investigate whether the compatibility effects were present in both the enacted and the intended crime condition. We performed a 2 (Task: relevant vs. irrelevant SRC task) x 2 (Compatibility: compatible vs. incompatible trials) repeated measures ANOVA's for each condition of the Enactment manipulation. The ANOVA in the enacted crime condition showed significant main effects of Task, $F(1, 24) = 27.50$, $p < .001$, $d = 1.05$, and Compatibility, $F(1, 24) = 25.20$, $p < .001$, $d = 1.01$, but most importantly, the predicted interaction of Task x Compatibility, $F(1, 24) = 29.94$, $p < .001$, $d = 1.09$. A significant

compatibility effect was found in the relevant SRC task, $t(24) = 5.70$, $p < .001$, $d = 1.14$, but not in the irrelevant SRC task, $t(24) = 0.51$, $p = .62$, $d = 0.10$.

In the intended crime condition, the ANOVA showed significant main effects of Task, $F(1, 24) = 67.96$, $p < .001$, $d = 1.65$, and Compatibility, $F(1, 24) = 6.94$, $p < .05$, $d = 0.53$, and the crucial interaction of Task x Compatibility, $F(1, 24) = 9.59$, $p < .01$, $d = 0.62$. Again a significant compatibility effect was found in the relevant SRC task, $t(24) = 2.89$, $p < .01$, $d = 0.58$, but not in the irrelevant SRC task, $t(24) = 0.91$, $p = .32$, $d = 0.18$.

2.2.2 Reaction times

The ANOVA on RTs showed significant main effects of Task, $F(1, 24) = 115.94$, $p < .001$, $d = 2.15$, and Compatibility, $F(1, 24) = 61.02$, $p < .001$, $d = 1.56$, and the crucial interaction of Task x Compatibility, $F(1, 24) = 49.19$, $p < .001$, $d = 1.40$. Furthermore, we found an effect of Enactment, $F(1, 24) = 10.67$, $p < .01$, $d = 0.65$, as well as interactions of Task x Enactment, $F(1, 24) = 5.47$, $p = .03$, $d = 0.46$, and Compatibility x Enactment, $F(1, 24) = 6.37$, $p = .02$, $d = 0.50$. The significant three-way interaction of Task x Compatibility x Enactment, $F(1, 24) = 4.99$, $p < .05$, $d = 0.45$, together with an inspection of Figure 1, revealed a larger compatibility effect (shorter RTs in compatible compared to incompatible trials) in the intended crime condition compared to the enacted crime condition in the relevant SRC task, $t(24) = 2.49$, $p < .05$, $d = 0.50$, but not in the irrelevant SRC task, $t < 1$.

To further explore the presence of the compatibility effect in both conditions of the Enactment manipulation, we performed separate 2 (Task: relevant vs. irrelevant SRC task) x 2 (Compatibility: compatible vs. incompatible trials) repeated measures ANOVA's. The ANOVA in the enacted crime condition showed significant main effects of Task, $F(1, 24) = 126.66$, $p <$

.001, d = 2.25, and Compatibility, $F(1, 24) = 15.07$, $p < .01$, d = 0.77, and the crucial interaction of Task x Compatibility, $F(1, 24) = 12.41$, $p < .01$, d = 0.70. A significant compatibility effect was found in the relevant SRC task, $t(24) = 3.78$, $p < .01$, d = 0.76, but not in the irrelevant SRC task, $t(24) = 0.61$, $p = .55$, d = 0.12.

The ANOVA in the intended crime condition showed significant main effect of Task, $F(1, 24) = 103.48$, $p < .001$, d = 2.03, and Compatibility, $F(1, 24) = 58.18$, $p < .001$, d = 1.53, and the predicted interaction of Task x Compatibility, $F(1, 24) = 50.63$, $p < .001$, d = 1.42. Again a significant compatibility effect was found in the relevant SRC task, $t(24) = 7.69$, $p < .001$, d = 1.54, but not in the irrelevant SRC task, $t(24) = 1.30$, $p = .21$, d = 0.27.

2.2.3. Catch trials and attention control questions

Errors in catch trials were defined as ‘yes’ responses to the word ‘No’, and ‘no’ responses to the word ‘Yes’. There was no significant difference in the error rate to the Yes/No catch trials between the relevant (M = 8.00 %, SD = 17.80) and the irrelevant SRC task (M = 2.40 %, SD = 5.92), $t(24) = 1.48$ $p = .15$, d = 0.30.

Attention control question were regarded as incorrect if the written question differed semantically from the question that was presented on the screen. Small grammatical or syntactic differences were not considered as incorrect. There was a significantly higher error rate to the attention control questions in the irrelevant SRC task (M = 19.20 %, SD = 22.72), compared to the relevant SRC task (M = 4.80 %, SD = 10.46), $t(24) = 3.52$, $p < .01$, d = 0.70, suggesting that participants processed the content of the questions in the irrelevant SRC task less than in the relevant SRC task.

2.2 Discussion

In line with Hypothesis 1, we found the predicted main effect of Compatibility. However, the significant interaction of Task and Compatibility and the follow-up tests revealed that the compatibility effect was significant only in the relevant SRC task and not in the irrelevant SRC task. Although we did expect to find a compatibility effect in the irrelevant SRC task, the results nevertheless supported Hypothesis 2: Greater compatibility effects were obtained in the relevant SRC task compared to the irrelevant task. The results of the attention control questions showed that participants made more errors when asked about the content of the last question in the irrelevant SRC task, compared to the relevant SRC task. Our data provided the first indication that compatibility effects may be smaller in irrelevant SRC tasks than in relevant SRC tasks because stimuli are processed less deeply in irrelevant compared to relevant SRC tasks.

The pattern of results was similar in both crime conditions (the enacted and the intended crime), supporting Hypothesis 3 that also intended criminal behavior can be measured, albeit only with the relevant SRC task. The RT compatibility effect in the relevant SRC task was even bigger in the intended compared to the enacted crime condition.

3. Experiment 2

Experiment 2 was conducted for three reasons. First, it allowed us to test the robustness of the findings of the relevant SRC effects. Second, in contrast with our hypothesis, we did not find a significant compatibility effect in the irrelevant SRC task. The attention control questions revealed differences in the depth of processing of the stimuli. If the irrelevant SRC effect was indeed absent in Experiment 1 because participants did not process the meaning of the questions (and could therefore not determine the true answer to the question), then the magnitude of the compatibility effect should increase and perhaps mirror that of the relevant SRC effect when the

semantic meaning of the question becomes task relevant. Third, we wanted to replicate the findings regarding the merely intended crime.

3.1. Method and participants

The number and presentation of stimuli, categories and timing were identical to Experiment 1. However, we reformulated some of the questions, and immediately presented the questions in color. With the exception of these changes, the relevant SRC task was identical to the one of Experiment 1. In the irrelevant SRC task, participants now responded to a semantic feature of the questions. More specifically, they had to determine whether the questions involved a person or a room. For instance, they had to say 'yes' to questions involving a person (e.g., 'Did you steal something from Bruno?'), and 'no' to questions involving a room (e.g., 'Did you look for the printer room?'). Hence, compatible trials were trials in which true questions involved a person, and false questions involved a room. Incompatible trials were those in which true questions involved a room, and false questions involved a person. To make it more difficult for participants to use strategies to avoid processing the meaning of the stimuli (e.g., only processing the last word or the verb), we created sentences in which the position of the indicative word was not fixed, and the verb was not indicative of the correct response (of course within the possibilities of Dutch grammar). The assignment of the keys, semantic features, colors, and tasks was again counterbalanced across participants

To obtain a sample of at least 20 participants, we invited 25 undergraduate students of Ghent University. Three participants canceled their appointment, and 22 students participated for a monetary reward. All provided written informed consent. One participant was excluded from further data-analyses because she exceeded the mean error rate per subject plus 2.5 standard

deviations (*M* = 8.27 %, *SD* = 7.92). The mean age of the remaining 21 participants was 20.47 (*SD* = 1.12). Most participants (*n* = 16) were women.

3.2. Results

Mean error rates and RTs are shown in Figure 2. Both were analyzed with a 2 (Task: relevant vs. irrelevant SRC task) x 2 (Compatibility: compatible vs. incompatible trials) x 2 (Enactment: enacted vs. intended crime) repeated measures ANOVA. For the RT analysis, trials with errors were discarded. RTs more than 2.5 standard deviations from the mean per subject and task were removed from the RT analysis (2.90 % of all trials with correct responses).

- INSERT FIGURE 2 ABOUT HERE -

3.2.1. Error rate

The ANOVA on error percentage showed a significant main effect of Task, $F(1, 20) = 28.85$, $p < .001$, $d = 1.17$, with a higher error rate in the relevant compared to the irrelevant SRC task, and a significant main effect of Compatibility, $F(1, 20) = 4.38$, $p < .05$, $d = 0.46$, with more errors in incompatible compared to compatible trials. The crucial interaction of Task x Compatibility was not significant, $F(1, 20) = 2.57$, $p = .13$, $d = 0.35$. Furthermore, we observed a main effect of Enactment, $F(1, 20) = 6.72$, $p < .05$, $d = 0.57$, with a higher error rate for the intended crime compared to the enacted crime. There were no other significant interaction effects, F's < 2.20.

As in Experiment 1, we performed a separate 2 (Task: relevant vs. irrelevant SRC task) x 2 (Compatibility: compatible vs. incompatible trials) repeated measures ANOVA for each condition of the Enactment manipulation. The ANOVA in the enacted crime condition revealed a

main effect of Task, F (1, 20) = 20.73, $p < .001$, $d = 0.99$, with a higher error rate in the relevant compared to the irrelevant SRC task. The main effect of Compatibility was not significant, F(1, 20) = 0.97, $p = .34$, $d = 0.22$, as well as the interaction of Task x Compatibility, F(1, 20) = 2.95, $p = .10$, $d = 0.37$.

The ANOVA in the intended crime condition showed a significant main effect of Task, F(1, 20) = 20.98, $p < .001$, $d = 1.00$, with a higher error rate in the relevant compared to the irrelevant SRC task, and a significant main effect of Compatibility, F(1, 20) = 5.35, $p < .05$, $d = 0.50$, with more errors in incompatible compared to compatible trials. The interaction of Task x Compatibility was not significant, F(1, 20) = 1.81, $p = .19$, $d = 0.29$.

3.2.2. Reaction times

The ANOVA on RTs showed significant main effects of Compatibility, F(1, 20) = 7.05, $p < .05$, $d = 0.58$, and Task, F(1, 24) = 58.87, $p < .001$, $d = 1.67$, and we observed the crucial interaction of Task x Compatibility, F(1, 20) = 6.23, $p < .05$, $d = 0.54$. This interaction was due to a significant compatibility effect in the relevant SRC task, t(20) = 2.96, $p < .01$, $d = 0.64$, but not in the irrelevant SRC task, t(20) = 0.55, $p = .59$, $d = 0.12$.[8] Furthermore, we observed a significant interaction of Task x Enactment, F(1, 20) = 12.13, $p < .01$, $d = 0.76$, indicating longer RTs in the intended crime condition compared to the enacted crime condition in the relevant SRC condition, t(20) = 2.84, $p < .05$, $d = 0.62$, but longer RTs in the enacted crime condition compared to the intended crime condition in the irrelevant SRC condition, t(20) = 2.63, $p < .05$,

[8] It is known that large overall RT differences between two conditions can cause an artificial increase in difference measures. To make sure that the interaction of Task x Compatibility in both experiments was not due to the main effects of task (significantly longer RTs in the relevant SRC task compared to the irrelevant SRC task), we log-transformed the mean RTs for each participant (Chapman et al., 1994). The Task x Compatibility x Enactment repeated measures ANOVA revealed a significant interaction of Task x Compatibility, F(1, 24) = 36.56, $p < .001$ in Experiment 1 and a marginally significant interaction of F(1, 20) = 4.14, $p = .06$ in Experiment 2.

d = 0.57. The main effect of Enactment and the other interaction effects were not significant, F's < 1.79.

Again, we performed a separate 2 (Task: relevant vs. irrelevant SRC task) x 2 (Compatibility: compatible vs. incompatible trials) repeated measures ANOVA for each condition of the Enactment manipulation. The ANOVA in the enacted crime condition showed a significant main effect of Task, F (1, 20) = 47.60, p < .001, d = 1.51, with longer RTs in the relevant compared to the irrelevant SRC task, and a marginally significant main effect of Compatibility, F (1, 20) = 3.99, p = .06, d = 0.43, with longer RTs for incompatible compared to compatible trials. The interaction of Task x Compatibility was not significant, F(1, 20) = 1.87, p = .19, d = 0.30. Nevertheless, a tendency towards a significant compatibility effect was found in the relevant SRC task, t(20) = 1.86, p = .08, d = 0.40, but not in the irrelevant SRC task, t(20) = 0.06, p = .95, d = 0.01.

The ANOVA in the intended crime condition showed significant main effects of Task, F (1, 20) = 67.67, p < .001, d = 1.79, and Compatibility, F(1, 20) = 5.53, p < .05, d = 0.53, and a significant interaction of Task x Compatibility, F(1, 20) = 8.70, p < .01, d = 0.64. A significant compatibility effect was found in the relevant SRC task, t(20) = 2.92, p < .01, d = 0.64, but not in the irrelevant SRC task, t(20) = 1.42, p = .17, d = 0.31.

3.2.3. Catch trials and attention control questions

Errors were defined as in Experiment 1. There was no significant difference in the error rate to the Yes/No catch trials between the relevant (M = 9.52 %, SD = 15.26) and the irrelevant SRC task (M = 15.50 %, SD = 15.81), t(20) = 0.84, p = .41, d = 0.18.

There was a significant difference in error rate to the attention control questions between the relevant (M = 6.00 %, SD = 13.47) and the irrelevant SRC task (M = 21.40 %, SD = 24.10), $t(20) = 2.65$, $p < .05$, $d = 0.58$, suggesting that participants processed the content of the questions in the irrelevant SRC task less than in the relevant SRC task.

3.3. Discussion

The key finding of Experiment 2 was that the task that manipulated relevant SRC again outperformed the task that manipulated irrelevant SRC. Although the interaction of Task and Compatibility was significant only in the RTs, the same trends were observed in the error data. The fact that the interaction was somewhat less evident in Experiment 2 than in Experiment 1 might be due to the larger variance in Experiment 2 compared to Experiment 1, most likely because the irrelevant SRC task in Experiment 2 was more difficult. The stimuli were also somewhat different due to the changes that had to be made for the irrelevant SRC task. This might have led to slightly longer overall RTs in Experiment 2 in comparison with Experiment 1.

A possible reason for not finding a compatibility effect in the irrelevant SRC task might be that, despite several precautions, participant nevertheless found strategies that enabled them to perform the irrelevant SRC task without processing the meaning of the entire question. The results of the attention control questions in Experiment 2 indeed indicated that participants processed the question content in the irrelevant SRC task less deeply than in the relevant SRC task. This idea suggests that differences between both tasks may disappear once the level of stimulus processing is similar. With this said however, we were not able to test this prediction because our measure to increase the level of processing in the irrelevant SRC task (i.e., use a semantic rather than non-semantic relevant feature) did not eliminate the task differences.

The results regarding the intended crime were replicated, with a trend towards the expected effect in the error data and a significant compatibility effect in the RTs. The higher error rate and longer RTs in the relevant compared to the irrelevant SRC task indicated that the relevant SRC task was more difficult than the irrelevant SRC task.

4. General Discussion

Following up on a suggestion made by Verschuere and De Houwer (2011), we investigated whether the structural properties of RT tasks determine how suitable these tasks are for capturing deception. We therefore compared a relevant with an irrelevant SRC task. Our first prediction was that a compatibility effect would arise in both tasks. Although the irrelevant SRC tasks did not reveal significant compatibility effects, our results indicated that robust group effects can be found with RT deception measures that are based on the manipulation of relevant SRC.

The relevant SRC task that we used in our experiments is an adaptation of the 'Differentiation of Deception' paradigm (Furedy et al., 1988; see also Spence et al., 2001). Developed as an experimental manipulation of truth telling and lying, the paradigm has mainly been used in fundamental research about the cognitive processes underlying lying, and provides reliable lie-truth differences (e.g., Debey et al. 2012; Farrow et al. 2010; Fullam et al. 2009; Hu et al., 2012a; Spence et al., 2001; Van Bockstaele et al. 2012; Verschuere et al. 2011, 2012). This relevant SRC task has, however, rarely been applied to discriminate between guilt and innocence. For that purpose, the relevant SRC task would have to be tested in the more realistic situation in which guilty participants would not comply with the 'truth telling' and 'lying' instructions, but would try to deceive the experimenter. Using a relevant SRC task identical to

the one we used, one would expect that the instruction that leads to less errors and faster RTs represents the actual truth for a given subject. To our knowledge, the only study that tested a variant of this task in an applied context is the case study by Spence et al. (2008). These authors rephrased the instructions of 'truth telling' vs. 'lying' to 'admitting the crime' vs. 'denying the crime'. Greater RTs and stronger BOLD activity in ventrolateral prefrontal and anterior cingulate cortices in the 'admitting the crime' condition compared to the 'denying the crime' condition were taken as an indication of the examinee´s innocence. Ground truth was not known in this real life case, which is why this conclusion could not be verified. Moreover, it remains to be determined whether the same pattern could have been obtained in case the examinee was guilty. Testing the 'Differentiation of Deception' paradigm under more realistic situations also implies an investigation of its fakeability. Faking is a problem of almost all deception measures (e.g., Ben-Shakhar, (2011), but it had been argued that RTs in particular might be under voluntary control (e.g., Farwell & Donchin, 1991). Although research indicated that RTs are not as vulnerable to faking attempts as is often assumed (Seymour et al., 2000), providing participants with specific faking instructions did significantly decrease test accuracy (e.g., Hu et al., 2012c; Verschuere et al., 2009). The application of response deadlines to prevent strategic RT slowing has been proposed and the first algorithms to detect faking have been developed (Agosta et al., 2011b). Nevertheless, much more research is needed to evaluate the circumstances under which RT tasks are fakeable and to compare their faking vulnerability with that of other methods. In sum, whereas it is premature to apply the 'Differentiation of Deception' paradigm for the applied purpose of lie detection, the medium to large effect sizes found with this relevant SRC task make it worthwhile to explore its applied potential.

The error rate and RT results in Experiment 1 supported our second hypothesis that RT tasks that manipulate relevant SRC produce bigger effects than RT tasks that manipulate irrelevant SRC. These findings are in line with a study of Field et al. (2011), who found that only a relevant SRC task and not an irrelevant SRC task produced significant compatibility effects in heavy drinkers. These authors speculated that the difference between the tasks could be attributed to depth of stimulus processing. An advantage of our study was that by implementing attention control trials, we were able to quantify and compare the degree of stimulus processing in both tasks. As revealed by the attention control trials, participants did indeed process the content of the questions less deeply in the irrelevant SRC task than in the relevant SRC task. When participants do not process the meaning, it is unlikely that they can determine the true answer to this question, and thus that there can be an effect of the compatibility between that true answer and the response that the participants are required to emit. To enhance stimulus processing in the irrelevant SRC task, we used semantic content instead of color as a critical response feature in Experiment 2. Nevertheless, there was again a substantial compatibility effect in the relevant SRC task (at least in the RTs) and no significant compatibility effect in the irrelevant SRC task. The attention control questions again indicated a shallower processing of the content in the irrelevant SRC task than in the relevant SRC task. It seems reasonable to conclude that the fact that participants did not process the meaning of the questions and therefore did not determine the correct answer to the question, led to the absence of a compatibility effect in the irrelevant SRC task. Without the processing of the crucial incompatibility-related stimulus feature, irrelevant SRC effects may be unlikely to emerge.

There may be other reasons why the irrelevant SRC task in the current study produced smaller compatibility effects than the relevant SRC task. One of the key features of irrelevant

SRC tasks is that they are, in general, less controlled than relevant SRC tasks. Participants might come up with different strategies to recode responses or task instructions to avoid processing the incompatibility-related feature. Although results of our Yes/No catch trials suggest that we successfully prevented participants from recoding the meaning of the response buttons in the irrelevant task (e.g., 'left' and 'right' or 'person' and 'room' instead of 'yes' and 'no'), we did not control for other strategies. Developing an irrelevant SRC task that forces participants to process the content to an extent comparable to the relevant SRC task would help to clarify whether such a task would produce comparable compatibility effects, or whether differences would still remain. Although our results indicate that developing such a task is challenging, the use of less complex stimuli to make it more difficult for participants to ignore incompatibility-related feature might be a first step. Moreover, additional means could be implemented to prevent participants from ignoring the content (e.g., include trials on which the semantic content of the stimuli is task relevant, similar to the case in the Identification Extrinsic Affective Simon Task; ID-EAST; De Houwer & De Bruycker, 2007). In sum, although we cannot exclude the possibility that some optimal version of an irrelevant SRC task would be suitable for capturing deception, perhaps even to the same extent than relevant SRC tasks, our studies clearly demonstrate the difficulties in constructing effective irrelevant SRC tasks and document the reasons underlying these difficulties (e.g., the fact that it is difficult to ensure that meaning is processed). More generally, our research reveals that structural properties of RT task are relevant for the effectiveness of these tasks in capturing deception.

With regard to our third hypothesis, relevant SRC tasks, but not irrelevant SRC tasks, seem to be able to measure intended criminal behavior to the same degree as deception about an enacted mock crime - as indicated by the error rate and RTs in Experiment 1 and the RTs in

Experiment 2. This finding is in line with previous research on intended criminal behavior (Clemens et al., 2011; Meijer et al., 2010; Meixner & Rosenfeld, 2011; Vrij et al., 2011). This also brings about the challenge for future research to examine whether deception tasks in general and RT-based deception task in particular can differentiate between individuals with actual criminal intentions and those who are only fantasizing about a crime (Agosta et al., 2011a).

As we mentioned above, a limitation of our study is that we did not control for other strategies that participants might have used to avoid processing the incompatibility-related feature. Another limitation of our study is that our irrelevant SRC task might have been underpowered, because compatibility effects in irrelevant SRC tasks are known to be smaller than compatibility effects in relevant SRC tasks. However, considering the large effect sizes that are needed in the context of deception detection in order to enable sufficiently precise individual classification, this would not invalidate our conclusion that irrelevant SRC tasks are less suited than relevant SRC tasks in the deception context. A limitation of our mock crime design is that the intended crime was always closer in time to the RT tasks than the enacted one. This complicates a comparison of effects for the two types of crime. Finally, it may be worthwhile to mention that although Verschuere and De Houwer (2011) proposed a crucial role of the manipulation of SRC in order to obtain robust RT effects, they did not exclude the possibility that paradigms without SRC manipulation can sometimes also produce significant RT effects. For instance, Rosenfeld and Labkovsky (2010) found significant RT differences in the Complex Trial Protocol (CTP; Rosenfeld et al., 2008), a paradigm without any obvious manipulation of SRC (see also Hu et al., 2012b; Labkovsky and Rosenfeld, 2012). The main dependent variable measured in this paradigm is an event-related potential (i.e., the P300) known to be evoked by infrequently presented meaningful stimuli (e.g., crime-related information) within more

frequently presented irrelevant stimuli. This so-called oddball effect might also lead to a RT slowing for personally relevant stimuli. However, in agreement with Verschuere and De Houwer (2011), we would hypothesize that the RT effects in such paradigms will not be as large and robust as the effects of relevant SRC tasks and therefore less suited for deception detection purposes.

To sum up, the results of the our experiments confirm that robust compatibility effects can be found with RT deception measures that are based on the manipulation of relevant SRC, for both executed as well as merely intended crimes. In line with the idea that structural properties of RT task need to be considered when designing RT measures of deception, the irrelevant SRC tasks did not measure deception. The difference between the relevant and the irrelevant SRC task seems to be related to the fact that participants processed the stimuli to a lesser degree during the irrelevant SRC tasks than during the relevant SRC tasks. Future research may try to develop an irrelevant SRC task that ensures processing of the crucial incompatibility-related feature. More promising perhaps, future research is needed to exploit the applied potential of the relevant SRC task.

Acknowledgments

Kristina Suchotzki is supported by an ECRP Grant (09-ECRP-025; FWO Grant ESF 3G099310).

References

Agosta, S., Castiello, U., Rigoni, D., Lionetti, S., Sartori, G., 2011a. The Detection and the Neural Correlates of Behavioral (Prior) Intentions. J. Cogn. Neurosci. 23, 3888-3902. doi:10.1162/jocn_a_00039

Agosta, S., Ghirardi, V., Zogmaister, C., Castiello, U., Sartori, G., 2011b. Detecting Fakers of the autobiographical IAT. Appl. Cogn. Psychol. 25, 299-306. doi: 10.1002/acp.1691

Allen, J., Iacono, W. G., 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(5), 504-522. doi: 10.1111/j.1469-8986.1992.tb02024.x

Ben-Shakhar, G., 2011. Countermeasures, in: Verschuere, B., BenShakhar, G., Meijer, E.H. (Eds.), Memory detection: theory and application of the concealed information test. Cambridge University Press, Cambridge, pp. 200-214.

Chapman, L.J., Chapman, J.P., Curran, T.E., Miller, M.B., 1994. Do Children and the Elderly Show Hightened Semantic Priming? How to Answer the Question. Dev. Rev. 14, 159-185. doi:10.1006/drev.1994.1007

Clemens, F., Granhag, P.A., Strömwall, L.A., 2011. Eliciting cues to false intent: A new application of strategic interviewing. Law Human Behav., 35, 512-522. doi:10.1007/s10979-010-9258-9

Cohen, J., 1988. Statistical power analysis for the behavioural sciences. Lawrence Erlbaum, Hillsdale.

Debey, E., Verschuere, B., Crombez, G., 2012. Lying and executive control: an experimental investigation using ego depletion and goal neglect. Acta Psychol. 140(2), 133-141. doi:10.1016/j.actpsy.2012.03.004

De Clercq, A., Crombez G., Buysse, A., Roeyers, H., 2003. A simple and sensitive method to measure timing accuracy. Beh. Res. Meth. Ins. C. 35, 109-115.

De Houwer, J., 2003. A structural analysis of indirect measures of attitudes, in: Musch, J., Klauer, K.C. (Eds.), The psychology of evaluation: Affective processes in Cognition and Emotion. Lawrence Erlbaum, Mahwah, pp. 219-244.

De Houwer, J., 2009. Comparing measures of attitudes at the procedural and functional level, in: Petty, R., Fazio, R.H., Brinol, P. (Eds.), Attitudes: Insights from the New Implicit Measures. Erlbaum, Hillsdale, NJ.

De Houwer, J., 2011. Why the Cognitive Approach in Psychology Would Profit From a Functional Approach and Vice Versa. Perspect. Psychol. Sci. 6, 202-209. doi:10.1177/1745691611400238

De Houwer, J., De Bruycker, E., 2007. The identification-EAST as a valid measure of implicit attitudes toward alcohol-related stimuli. J. Behav. Ther. Exp. Psy. 38, 133-143. doi:10.1016/j.jbtep.2006.10.004

Dunlap,W.P., Cortina, J.M., Vaslow, J.B., Burke, M.J., 1996. Meta-analysis of experiments with matched groups or repeated measures designs. Psychol. Methods 1, 170–177. doi:10.1037//1082-989X.1.2.170

Engelhard, I. M., Merckelbach, H., van den Hout, M. A., 2003. The guilty knowledge test and the modified Stroop task in detection of deception: An exploratory study. Psychol. Rep. 92(2), 683-691. doi:10.2466/PR0.92.2.683-691

Farrow, T.F.D., Hopwood, M.-C., Parks, R.W., Hunter, M.D., Spence, S.A., 2010. Evidence of mnemonic ability selectively affecting truthful and deceptive response dynamics. The American Journal of Psychology, 123, 447-453. doi: 10.5406/amerjpsyc.124.4.0447

Farwell, L.A., Donchin, E., 1991. The Truth Will out - Interrogative Polygraphy (Lie Detection) with Event-Related Brain Potentials. Psychophysiology 28, 531-547. doi:10.1111/j.1469-8986.1991.tb01990.x

Field, M., Caren, R., Fernie, G., De Houwer, J., 2011. Alcohol approach tendencies in heavy drinkers: Comparison of effects in a Relevant Stimulus-Response Compatibility Task and an approach / avoidance Simon task. Psychol. Addict. Behav. 25, 697-701. doi:10.1037/a0023285

Fitts, P.M., Seeger, C.M., 1953. SR compatibility: Spacial characteristics of stimulus and response codes. J. Exp. Psychol. 46, 199-210. doi:10.1037/h0062827

Fullam, R.S., McKie, S., Dolan, M.C., 2009. Psychopathic traits and deception: functional magnetic resonance imaging study. Br. J. Psychiatry, 194, 229-235. doi:10.1192/bjp.bp.108.053199

Furedy, J.J., Davis, C., Gurevich, M., 1988. Differentiation of Deception as a Psychological Process - a Psychophysiological Approach. Psychophysiology 25, 683-688. doi:10.1111/j.1469-8986.1988.tb01908.x

Gronau, N., Ben-Shakhar, G., Cohen, A., 2005. Behavioral and physiological measures in the detection of concealed information. J. Appl. Psychol. 90, 147-158. doi:10.1037/0021-9010.90.1.147

Hu, X., Chen, H., Fu, G., 2012a. 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

Hu, X., Hegeman, D., Landry, E., Rosenfeld, J.P., 2012b. 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. doi:10.1111/j.1469-8986.2011.01286.x.

Hu, X., Rosenfeld, J.P., & Bodenhausen, G.V., 2012c. Combatting 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:10.1177/0956797612443834

Kornblum, S., Hasbroucq, T., Osman, A., 1990. Dimensional Overlap - Cognitive Basis for Stimulus-Response Compatibility - a Model and Taxonomy. Psychol. Rev. 97, 253-270. doi:10.1037//0033-295X.97.2.253

Kornblum, S., Lee, J.W., 1995. Stimulus-Response Compatibility with Relevant and Irrelevant Stimulus Dimensions That Do and Do Not Overlap with the Response. J. Exp. Psychol. Human 21, 855-875. doi:10.1037//0096-1523.21.4.855

Labkovsky, E., Rosenfeld, J.P., 2012. The P300-Based, Complex Trial Protocol for Concealed Information Detection Resists Any Number of Sequential Countermeasures Against Up to Five Irrelevant Stimuli. Appl. Psychophys. Biof., 37, 1-10. doi:10.1007/s10484-011-9171-0

MacLeod, C.M.,1991. Half a century of research on the Stroop effect: An integrative review. Psychol. Bull. 109, 163–203. doi:10.1037//0033-2909.109.2.163

Meijer, E.H., Verschuere, B., Merckelbach, H., 2010. Detecting Criminal Intent with the Concealed Information Test. The Open Criminology Journal 3, 44-47. doi:10.2174/1874917801003020044

Meixner, J.B., 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

Morris, S.B., Deshon, R.P., 2002. Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. Psychol. Method, 7, 105-25. doi:10.1037/1082-989X.7.1.105

Ratcliff, R., 1993. Methods for Dealing with Reaction-Time Outliers. Psychol. Bull. 114, 510-532. doi:10.1037//0033-2909.114.3.510

Rosenfeld, J.P., 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. doi:10.1111/j.1469-8986.2010.01024.x

Rosenfeld, J.P., Labkovsky, E., Winograd, M., Lui, M.A., Vandenboom, C., 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. doi:10.1111/j.1469-8986.2008.00708.x

Sartori, G., Agosta, S., Zogmaister, C., Ferrara, S.D., Castiello, U., 2008. How to accurately detect autobiographical events. Psychol. Sci. 19, 772-780. doi:10.1111/j.1467-9280.2008.02156.x

Seymour, T. L., Seifert, C. M., Shafto, M. G., Mosmann, A. L., 2000. Using response time measures to assess "guilty knowledge". J. Appl. Psychol. 85(1), 30-37. doi:10.1037//0021-9010.85.1.30

Simon, J.R., 1990. The effect of an irrelevant directional cue on human information processing, in: Proctor, R.W., Reeve, T.G. (Eds.), Stimulus-response compatibility:an integrated perspective. Amsterdam: North-Holland, pp. 31-86. doi:10.1016/S0166-4115(08)61218-2

Simon, J. R., Rudell, A. P., 1967. Auditory S-R compatibility: the effect of an irrelevant cue on information processing. J. Appl. Psychol. 51, 300-304. doi:10.1037/h0020586

Spence, S.A., Farrow, T.F.D., Herford, A.E., Wilkinson, I.D., Zheng, Y., Woodruff, P.W.R., 2001. Behavioural and functional anatomical correlates of deception in humans. Neuroreport 12, 2849-2853. doi:10.1097/00001756-200109170-00019

Spence, S.A., Kaylor-Hughes, C.J., Brook, M.L., Lankappa, S.T., 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. Psychiat. 23, 309-314. doi:10.1016/j.eurpsy.2007.09.001

Stroop, J.R., 1935. Studies of Interference in Serial Verbal Reactions. J. Exp. Psychol. 18, 643-662. doi:10.1037//0096-3445.121.1.15

Van Bockstaele, B., Verschuere, B., Moens, T., Suchotzki, K., Debey, E., Spruyt, A., 2012. Learning to lie: effects of practice on the cognitive cost of lying. Front. Psychol., 3, 526-552. doi:10.3389/fpsyg.2012.00526

Verschuere, B., Crombez, G., Degrootte, T., Rosseel, Y., 2010. Detecting concealed information with reaction times: Validity and comparison with the polygraph. Appl. Cognitive Psych. 23, 1-11. doi:10.1002/acp.1601

Verschuere, B., De Houwer, J., 2011. Detecting concealed information in less than a second: response latency-based measures, in: Verschuere, B., BenShakhar, G., Meijer, E.H.

(Eds.), Memory detection: theory and application of the concealed information test. Cambridge University Press, Cambridge, pp. 46-63.

Verschuere, B., Prati, V., De Houwer, J., 2009. Cheating the lie detector: Faking in the Autobiographical IAT. Psychol. Sci. 20, 410-413. doi:10.1111/j.1467-9280.2009.02308.x

Verschuere B., Schuhmann, T., Sack, A.T., 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

Verschuere, B., Spruyt, A., Meijer, E.H., Otgaar, H., 2011. The ease of lying. Conscious Cogn. 20, 908-911. doi:10.1016/j.concog.2010.10.023

Vrij, A., Leal, S., Mann, S.A., Granhag, P.A., 2011. A Comparison between Lying about Intentions and Past Activities: Verbal Cues and Detection Accuracy. Appl. Cognitive Psych. 25, 212-218. doi:10.1002/Acp.1665

Table 1

Trial types for the irrelevant and relevant SRC task. The examples listed correspond to the condition in which participants are asked to answer Yes to grey questions and No to purple questions (irrelevant SRC task) or to answer truthfully when the question is presented in yellow and lie when the question is presented in blue (relevant SRC task).

Trial types			**Irrelevant SRC task**			**Relevant SRC task**		
N	**Question type**	**Example**	**Color**	**Correct Response**	**Compatibility**	**Color**	**Correct Response**	**Compatibility**
15	Enacted crime	Did you steal a CD-ROM from Bruno?	Grey	Yes	Compatible	Yellow	Yes	Compatible
15			Purple	No	Incompatible	Blue	No	Incompatible
15	Control crime	Did you steal a laptop from Agnes?	Grey	Yes	Incompatible	Yellow	No	Compatible
15			Purple	No	Compatible	Blue	Yes	Incompatible
15	Intended crime	Will you steal an USB stick from Ernst?	Grey	Yes	Compatible	Yellow	Yes	Compatible
15			Purple	No	Incompatible	Blue	No	Incompatible
15	Intended control crime	Will you steal car keys from Geert?	Grey	Yes	Incompatible	Yellow	No	Compatible
15			Purple	No	Compatible	Blue	Yes	Incompatible

Note. N = number of trials

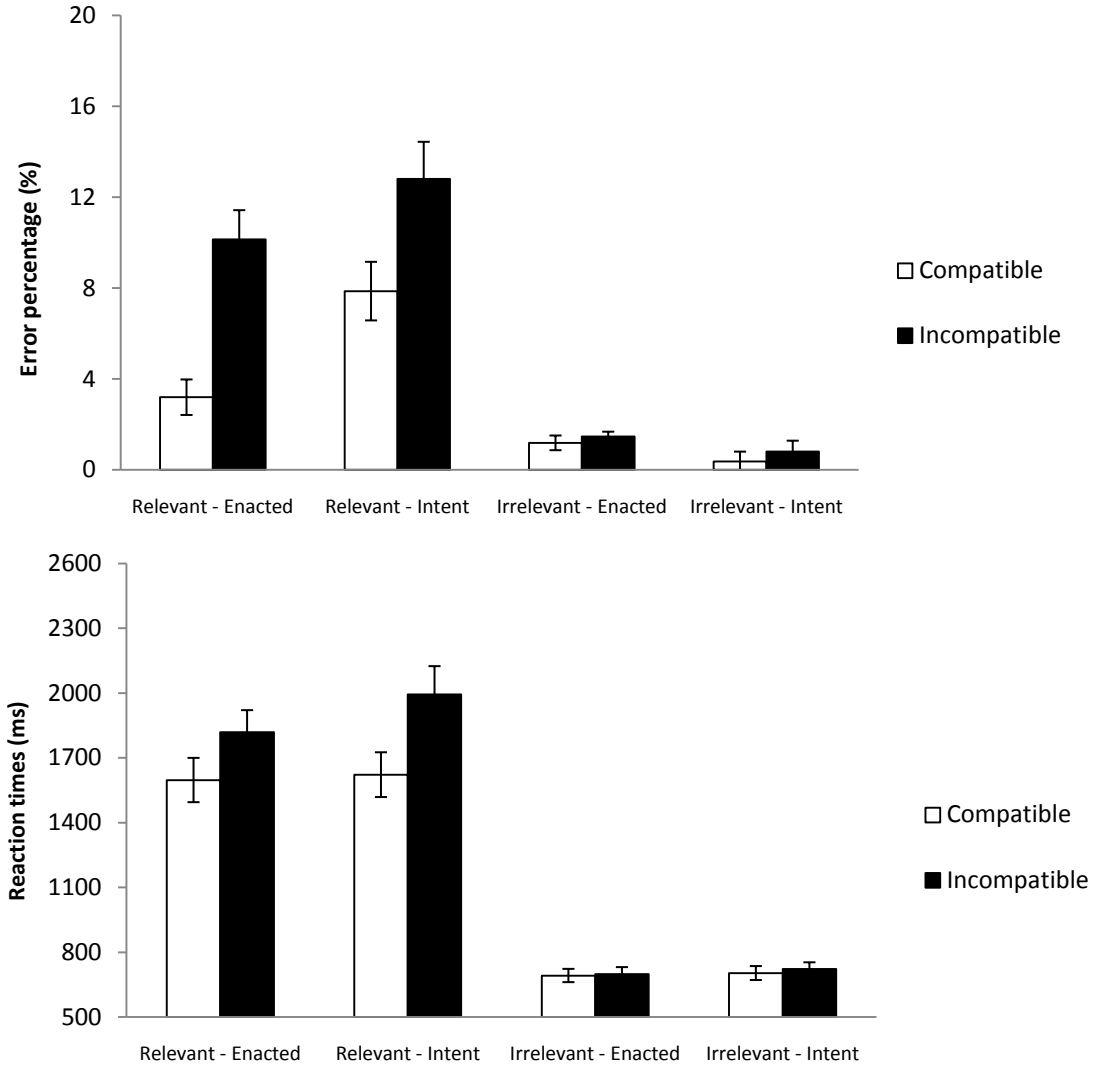

Figure.1. Mean error percentage (Panel A) and reaction times (Panel B) in Experiment 1, with standard error of the mean, for compatible and incompatible trials in the relevant and irrelevant SRC task, for both the enacted and the intended crime conditions.

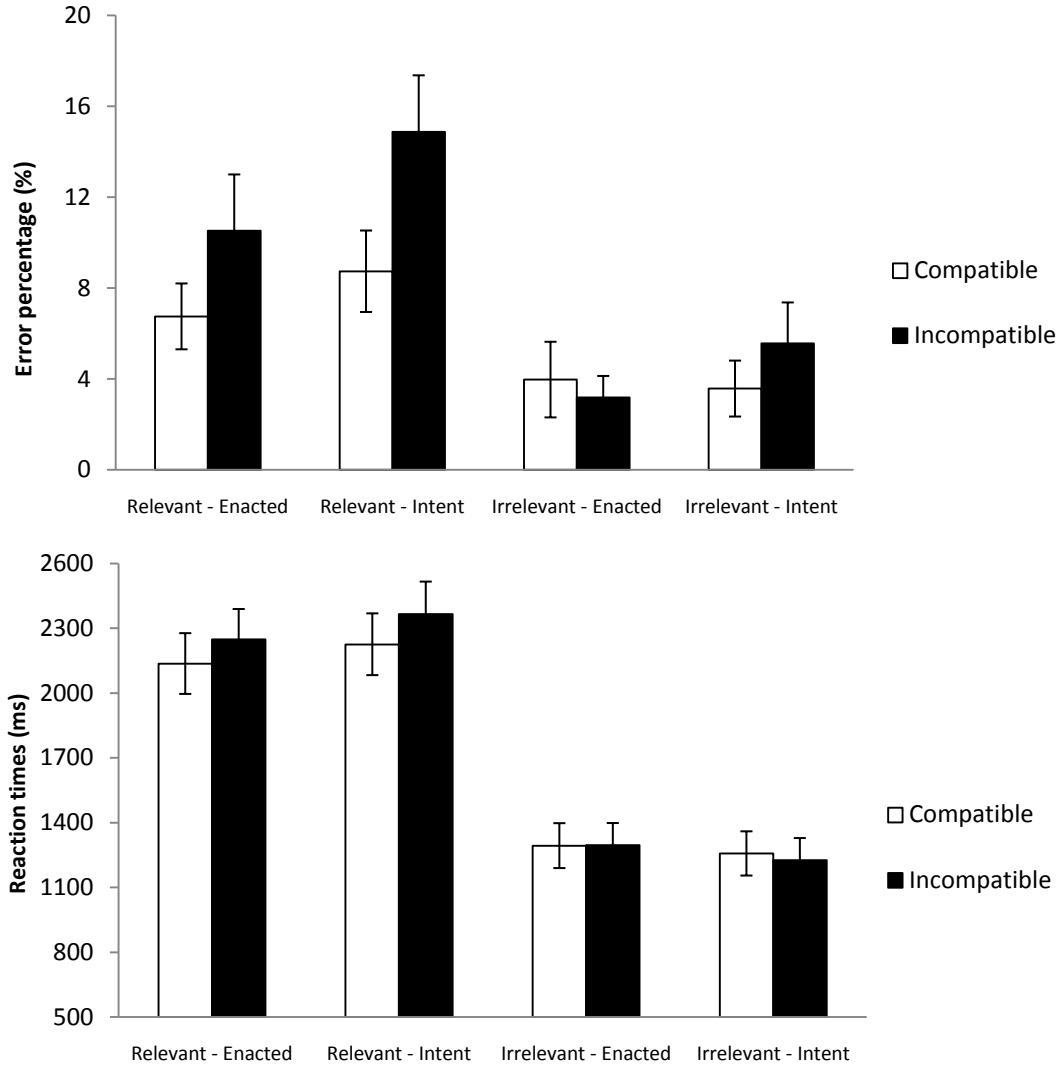

Figure.2. Mean error percentage (Panel A) and reaction times (Panel B) in Experiment 2, with standard error of the mean, for compatible and incompatible trials in the relevant and irrelevant SRC task, for both the enacted and the intended crime conditions.