INDIVIDUAL DIFFERENCES AND FALSE CONFESSIONS: A CONCEPTUAL REPLICATION OF KASSIN AND KIECHEL (1996)

ROBERT HORSELENBERG*, HARALD MERCKELBACH and SARAH JOSEPHS

Department of Experimental Psychology and Faculty of Law, Maastricht University, P.O. Box 616, 6200 MD, Maastricht, The Netherlands

(Received 26 February 2001; In final form 9 May 2001)

In their study, Kassin and Kiechel (1996) falsely accused students of causing a computer crash and found that 69% of them were willing to sign a false confession, 28% internalized guilt, and 9% confabulated details to support their false beliefs. The authors interpreted these results to mean that false confessions can be easily elicited. However, in their study, false confessions were explicitly not associated with negative consequences. The current study examined whether false incriminating evidence may elicit false confessions in undergraduate students when such confessions are explicitly associated with financial loss. We also explored whether individual differences in compliance, suggestibility, fantasy-proneness, dissociation, and cognitive failures are related to false confessions. The large majority of participants (82%) were willing to sign a false confession. In about half of the participants, false confessions were accompanied by internalization and confabulation. There was no evidence that individual differences modulate participants' susceptibility to false confessions. Taken together, our study replicates previous findings of Kassin and Kiechel.

INTRODUCTION

Although confession evidence plays an important role in criminal proceedings, this topic has received relatively little attention from experimental psychologists (Kassin, 1997). Meanwhile, especially false confessions warrant systematic study because, as a number of well documented case studies show (e.g., Wright, 1994; Gudjonsson, Kopelman and MacKeith, 1999), their legal ramifications might be far-reaching. Research about the incidence of self-reported false confessions among prison inmates has yielded estimates that vary from 0.6% (Cassell, 1998) to 12% (Gudjonsson and Sigurdsson, 1994). Yet, little is known about the precise situational and interpersonal antecedents of such confessions.

Several authors (Kassin and Wrightsman, 1985; Kassin, 1997; McCann, 1998; Wakefield and Underwager, 1998) have pointed that there are at least three distinct types of false confessions. The first type is the voluntary false confession that involves spontaneous self-incriminating statements made without external pressure. The second type is the coerced-compliant false confession. In these cases, suspects have the private belief that they are inno-
cent, but nevertheless confess in order to secure a benefit (e.g., promises of immunity) or to avoid a threatened harm or an aversive state (e.g., abstinence symptoms). The third type is the coerced-internalized false confession. Here suspects actually come to believe that they are guilty of a non-committed crime. While this type seems rather counter-intuitive, psychiatric literature offers many anecdotal reports illustrating such internalized false confessions (see, for a review, Gudjonsson, 1992). To some extent, internalized false confessions overlap with one particular category of memory illusions, namely illusions known as “false” memories (e.g., Loftus, 1993). However, whereas false memories usually refer to autobiographical recollections of fictitious childhood events that suddenly occur in the context of psychotherapy (e.g., Gudjonsson, 1997), internalized confessions may pertain to relatively recent episodes and, by definition, occur within a legal context. Over the past 5 years or so, a considerable number of laboratory studies have sought to identify factors that may contribute to the development of false memories. In general, this research has been successful in that it has identified certain interventions (e.g., false feedback, imagery, journaling) that may elicit false memories in a non-trivial minority of healthy subjects (e.g., Hyman and Pentland, 1998; Garry, Manning, Loftus and Sherman, 1996; Horserelenberg, Merckelbach, Muris, Rassin, Sijssenaar and Spaan, 2000).

In contrast, little work has been done on the experimental analysis of internalized false confessions. With the exception of the older study by Bem (1966), only the recent experiment of Kassin and Kiechel (1996) addressed the issue of internalized false confessions. These authors invited 75 undergraduates to participate in what was introduced as a computer task. It was stressed that during the task, participants should not hit the ALT-key. After 1 minute, the computer supposedly crashed and the experimenter accused participants of having pressed the forbidden ALT-key. The experimenter then asked participants to sign a standardized confession (“I hit the ALT-key and caused the program to crash. Data were lost”). Signing this confession was interpreted as evidence for the compliant type of false confession. To determine whether internalization of the false confessions had occurred, a confederate who waited outside the lab and who had overheard the commotion, asked participants what happened and recorded their reply. If participants expressed guilt for the computer crash, this was taken as evidence for an internalized false confession. Two variables were systematically manipulated, namely pace of the computer task (slow versus fast) and false incriminating evidence (absent or present), resulting in four experimental conditions. The false incriminating evidence condition consisted of a seemingly independent confederate who claimed that s/he had seen the participant hitting the ALT-key. In the absence of a witness and with a slowly paced task, 35% of the participants signed the confession form (compliance) while none of them expressed guilt for the computer crash (0% internalization). In the absence of a witness, but with a fast task, these percentages increased to 65% and 12%, respectively. When a witness was introduced in the context of a slowly paced task percentages increased even further to 89% and 44%, respectively. Finally, the combination of a witness and a fast paced task induced compliance and internalized false confessions in the large majority of participants, percentages being 100% and 65%, respectively. Accordingly, the authors concluded “that the presentation of false incriminating evidence—an interrogation ploy that is common among the police and sanctioned by many courts—can induce people to internalize blame for outcomes they did not produce” (Kassin and Kiechel, 1996; p. 127).

Although the pioneering work of Kassin and Kiechel (1996) bears relevance to the issue of false confessions, several critical points can be raised. To begin with, in the Kassin and Kiechel (1996) study, a supposedly non-involved and independent peer of the participant provided false incriminating evidence. However, in real life interrogations such an independent witness will be rare. A more plausible scenario is one in which an authority figure claims to possess privileged knowledge that might be incriminating. Secondly, in the Kassin and Kiechel (1996) study,
participants’ confessions had no immediate and explicit negative consequences. Again, this is very unlike the situation during a police interrogation. Here, a confession will produce a cascade of effects, one of which might be conviction, and therefore, punishment.

With these points in mind, the current experiment made an attempt to replicate the Kassin and Kiechel (1996) findings. However, the following procedural changes were made. First, all participants were exposed to an experimenter who claimed to possess incriminating evidence. Second, participants were aware of the fact that if they decided to sign a confession, this would result in a 80% reduction ($2 instead of $10) of the financial fee they would receive for their participation in the experiment.

A subsidiary aim of the present study was to explore to what extent individual difference variables contribute to experimentally induced false confessions. Previous research has found tentative evidence that high scores on the Gudjonsson Compliance Scale (GCS; Gudjonsson, 1989; 1991) are related to false confessions. For example, Sigurdsson and Gudjonsson (1996) noted that prison inmates who claimed to have made false confessions had significantly higher GCS scores than a control group. This suggests that individual differences in compliance contribute to false confessions of the coerced-compliant type (Gudjonsson, 1992). Sigurdsson and Gudjonsson (1996) also found evidence that false confessions of the coerced-internalized type are specifically linked to heightened suggestibility as indexed by the Gudjonsson Suggestibility Scale (GSS; Gudjonsson, 1984). Similarly, in their case study, Santtila, Alkiora, Ekholm and Niemi (1999) described the example of a defendant charged with robbery who made a coerced-internalized false confession, presumably because he was highly suggestible.

Other traits that may make people more susceptible to false confessions are fantasy proneness, dissociation, and self-reported cognitive failures. germane to this issue is research showing that fantasy proneness, dissociation, and/or cognitive failures predict individuals’ tendency to develop pseudomemories (e.g., Hyman and Billing, 1998; Merckelbach, Muris, Rassin and Horselenberg, 2000; Winograd, Peluso and Glover, 1998; but see also Horselenberg et al., 2000; Spanos, Burgess, Burgess, Samuels and Blois, 1999).

In sum, then, the current study examined to what extent false incriminating evidence presented by an experimenter elicits false confessions in undergraduate students, even when such confessions are associated with financial loss. Our study also explored whether compliance, suggestibility, fantasy proneness, dissociation, and cognitive failures predispose individuals to exhibit false confessions.

**METHOD**

**Participants**

Participants were 34 undergraduate psychology students (all women). Their mean age was 18.6 years (range 18–22 years). As a result of previous testing sessions, we had access to the compliance, suggestibility, fantasy proneness, dissociation, and cognitive failures scores of the participants in the current study. These psychometric data had been obtained some 8 months before the current study was conducted. Participants were told that they would participate in an ergonomic study that evaluated the effectiveness of new keyboard configurations. They were paid for their participation.

**Materials**

As part of individual testing sessions that took place some 8 months before the experiment proper, participants completed the Gudjonsson Compliance Scale (GCS; Gudjonsson, 1989),
the Gudjonsson Suggestibility Scale (GSS; Gudjonsson, 1984), the Creative Experiences Questionnaire (CEQ; Merckelbach, Horseelenberg and Muris, in press), the Dissociative Experiences Scale (DES; Bernstein and Putnam, 1986), and the Cognitive Failure Questionnaire (CFQ; Broadbent, Cooper, Fitzgerald and Parkes, 1982).

The GCS ($\alpha = .72$) is a self-report questionnaire consisting of 20 true/false statements that have to do with the way in which individuals may deal with conflicts and confrontation (e.g., “I give in easily to people when I am pressured.”). Responses are totalled with higher scores reflecting a stronger tendency to comply.

The GSS (version 1; $\alpha = .60$) consists of a short story and 20 specific questions about the story. Fifteen of these questions convey misleading information about the story. The extent to which an individual gives in to the leading questions constitutes that individual’s Yield score (maximum score = 15). Following the 20 questions, participants are told in an authoritative manner that they made many mistakes and that all questions have to be answered for a second time. The number of times that participants change their answers constitutes their Shift score (maximum score = 20). Yield and Shift scores are summed to obtain a total suggestibility score (maximum score = 35). A number of other parameters can be derived from the GSS, but for practical reasons (i.e., extensive test sessions), the present study only focused on Yield-, Shift-, and total suggestibility scores.

The CEQ ($\alpha = .70$) is a 25 item yes/no index of fantasy proneness. The CEQ items were derived from the extensive case descriptions of fantasy proneness provided by Wilson and Barber. Typical CEQ-items are: “In general, I spend at least half of the day fantasizing or daydreaming” and “My fantasies are so vivid that they are like a good movie”. Yes-answers are summed to yield a total CEQ score. High scores indicate fantasy proneness.

The DES ($\alpha = .86$) consists of 28 items that pertain to dissociative phenomena (e.g., feelings of derealisation, depersonalisation, disturbances in memory). Respondents indicate on 100 mm Visual Analogue Scales (VASs: $0 = not at all; 100 = very much$) the frequency with which they experience these symptoms. VAS-scores are averaged to obtain a total DES score. Higher total DES scores indicate stronger dissociative tendencies.

The CFQ ($\alpha = .83$) is a 25-item self-report instrument that measures the frequency of everyday lapses and minor blunders (e.g., “Do you forget appointments?”). Subjects indicate on 5-point scales how often they have experienced each cognitive failure in the past month (anchors: $0 = never; 4 = very often$). Scores are summed to obtain a total CFQ score, with higher scores indicating a higher frequency of self-reported cognitive failures.

For the computer task, a DELL PC was used running Experimental Run Time System (ERTS; Beringer, 1996). Stimuli appeared on a 17-inch screen in lower case in yellow colour against a grey background. Responses were recorded on line to correct for actually hitting the SHIFT-key. None of the participants hit the SHIFT-key.

**Procedure**

To increase the credibility of the cover story, participants completed several forms concerning their typing abilities, typing performance, and problems they previously encountered with various keyboard arrangements. Participants were told that they were assigned to a control group and, therefore, had to carry out a task on a standard keyboard. Following this, they were instructed to type as quickly and accurately as possible various letters appearing on the screen. Participants were told that they should not touch the SHIFT-key, because touching that key would result in a fatal computer crash and total loss of data. To make the possibility of a computer crash more plausible the letter Z, X, and A (which are all located near the SHIFT-key) were presented on some trials. In total, 96 stimuli were presented before the computer crashed.
After the computer crash, the experimenter accused participants of having touched the SHIFT-key and insisted that she had seen it happen with her own eyes (false incriminating evidence). Next, the experimenter invited participants to sign a hand-written confession stating—and this was stressed explicitly by the experimenter—that the data were lost because the participant hit the SHIFT-key and that she (i.e., the participant) would therefore lose 80% of her financial fee. If participants refused to sign the confession form, the experimenter once again encouraged them to sign it. The experimenter then left the room and advised the participants to think for a while about what had happened. After 5 minutes, a confederate entered the room. He explicitly stated that he was not associated with the ongoing experiment. The confederate asked participants what was going on. Participants’ answers were recorded verbatim. Finally, the experimenter re-entered and fully debriefed the participants and asked them not to discuss the experiment with their colleagues.

RESULTS

Of the 34 participants, 27 (82%) signed the confession, 14 (42%) exhibited internalization (i.e., expressed the belief that they were fully responsible for the computer crash) and 19 (58%) exhibited confabulation (i.e., gave detailed information about when, why, and how they hit the SHIFT-key). Note that participants more often displayed confabulation than internalization. This has to do with our definitions of these two constructs that allow for the possibility that a participant is not fully convinced of the fact she had touched the SHIFT-key, but does make up a story about what could have happened. Percentages of internalization and confabulation are conservative because only straightforward statements about personal responsibility and about when, why, and how the SHIFT-key was touched were accepted as evidence for internalization and confabulation, respectively.

Mean scores on the GCS, GSS-Yield, GSS-Shift, GSS-total, CEQ, DES, and CFQ were 10.3 (SD = 3.7), 3.4 (SD = 2.3), 5.8 (SD = 2.9), 9.2 (SD = 3.3), 7.8 (SD = 3.7), 20.0 (SD = 9.7), and 46.2 (SD = 10.7), respectively. These scores come close to those reported by previous studies for undergraduate samples (e.g., Gudjonsson, 1997; Merckelbach, Muris and Rassin, 1999; Merckelbach et al., 2000).

To explore the contributions of individual difference variables to coerced confessions (i.e., signing a form), internalizations, and confabulations, 3 separate Multivariate Analysis of Variance (MANOVA) were carried out. Each MANOVA followed a 2 (signed versus not signed or internalized versus not internalized or confabulated versus not confabulated) × 5 (individual difference variables) set-up. No evidence was found to support that people who sign, internalise, and/or confabulate differ from people who do not sign, internalise, and/or confabulate with respect to individual difference measures obtained in the current study: all $F(1,25)$'s < 1.5, all $p$'s > .2 for signing and all $F(1,29)$'s < 2.1, all $p$'s > .15 for internalizing and confabulating. Mean scores on the individual difference variables are given in Table I.

DISCUSSION

The main results of the current study can be catalogued as follows. First, in keeping with the Kassin and Kiechel (1996) study, the large majority of our participants were willing to sign a false confession. Second, in about half of them false confessions were accompanied by internalization (42%) and/or confabulation (58%). Third, no evidence was found to suggest that individual differences in compliance tendencies, interrogative suggestibility, fantasy proneness, or dissociation modulate participants’ susceptibility to false confessions.
Our results provide a fine replication of the Kassin and Kiechel (1996) findings. These authors reported overall percentages of 69%, 28%, and 9% for compliance, internalization, and confabulation, respectively. We obtained somewhat higher percentages (82%, 42%, and 58%, respectively), even though in our study, signing a false confession implicated financial loss. The fact that Kassin and Kiechel found overall lower percentages of false confessions might have to do with their instruction to participants to try to recollect how the computer crash did happen. In contrast, we asked our participants about what could have happened. However, our results demonstrate that normal participants can be manipulated to sign a false confession even when such an act has straightforward negative consequences. Our results also demonstrate that this effect does not depend on participants being confronted with a supposedly independent witness who presents false incriminating evidence. Negative feedback by an experimenter who is clearly involved and who claims to have seen the pertinent act suffices to elicit false confessions in a majority of participants. All in all, our results underline Kassin and Kiechel’s (1996; p.126) words “that the presentation of false incriminating evidence can induce people to internalize blame for outcomes they did not produce”.

A more disappointing aspect of the present study is that we found no evidence that the tendency to exhibit false confessions is modulated by individual difference variables. Inter-correlations between these variables were quite similar to those in previous work. For example, dissociation (DES) was positively and significantly associated with cognitive failures (CFQ; \( r = .53 \)) and fantasy proneness (CEQ; \( r = .31 \)), which replicates findings reported by Meitkelbach et al. (1999). Likewise, cognitive failures and interrogative suggestibility were positively associated with each other (\( r = .38 \)), a finding that was also reported by Merckelbach et al. (2000). These correlations demonstrate that the quality of our individual dierence data was good. Nevertheless, individual dierence variables were unable to predict the various aspects of false confessions. This null finding can be interpreted in several ways. One possibility is that after a time lapse of eight months, our individual dierence data may have become unreliable. However, research shows that the test-retest reliability over longer time periods for the questionnaires used in this study are quite good (GCS: \( r = .88 \); Gudjonsson, 1997; GSS: \( r = .83 \); Gudjonsson, 1997; CEQ: \( r = .95 \); Merckelbach et al., in press; DES: \( r = .78 \); Putnam, Chu and Dill, 1992; and CFQ: \( r = .83 \); Merckelbach, Muris, Nijman and De Jong, 1996). Another possibility is that false confessions in the laboratory are primarily context dependent and have little to do with personality traits. This possibility receives some support from the recently published work of Spanos and co-workers (1999). These authors found that a large majority of participants develop false memories about infancy when they are exposed to certain conditions (e.g., an authoritative person who creates clear expectations about the occurrence of such memories) and that personality traits like dissociation and fantasy proneness do not, or only to a modest degree, predict false

### Table 1

<table>
<thead>
<tr>
<th></th>
<th>Signed (n=21)</th>
<th>Not signed (n=6)</th>
<th>Internalized (n=13)</th>
<th>Not internalized (n=18)</th>
<th>Confabulated (n=18)</th>
<th>Not confabulated (n=13)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCS</td>
<td>9.7 (3.8)</td>
<td>10.7 (2.2)</td>
<td>9.8 (3.8)</td>
<td>10.3 (3.6)</td>
<td>10.6 (3.7)</td>
<td>9.4 (3.5)</td>
</tr>
<tr>
<td>GSS</td>
<td>9.7 (3.6)</td>
<td>8.5 (2.7)</td>
<td>7.4 (3.2)</td>
<td>8.8 (3.4)</td>
<td>7.7 (3.0)</td>
<td>9.0 (3.8)</td>
</tr>
<tr>
<td>CEQ</td>
<td>7.3 (3.7)</td>
<td>9.5 (3.8)</td>
<td>6.9 (3.9)</td>
<td>8.1 (3.5)</td>
<td>8.4 (3.7)</td>
<td>6.5 (3.5)</td>
</tr>
<tr>
<td>DES</td>
<td>19.7 (10.6)</td>
<td>24.7 (9.9)</td>
<td>16.5 (6.4)</td>
<td>22.5 (11.5)</td>
<td>19.1 (10.3)</td>
<td>21.2 (9.9)</td>
</tr>
<tr>
<td>CFQ</td>
<td>46.2 (12.3)</td>
<td>44.2 (9.1)</td>
<td>45.3 (13.1)</td>
<td>46.8 (9.3)</td>
<td>45 (10.5)</td>
<td>47.8 (11.6)</td>
</tr>
</tbody>
</table>

Due to missing data in the individual difference variables, the number of participants differ.
memory reporting. Thus, these findings seem to imply that environmental conditions rather than traits run the show. Of course, such external conditions might also include situationally induced risk factors like drugs and alcohol abuse or withdrawal (Gudjonsson, 1992; see also Santtila et al., 1999).

On the other hand, several case studies about false confessions made during criminal proceedings strongly suggest that personality traits like suggestibility and/or compliance are involved in the development of false confessions (see for examples: Gudjonsson, 1999a,b). Likewise, laboratory studies of, for example, Hyman and Billings (1998) and Winograd and associates (1998) found evidence that, when it comes to the development of false memories, personality traits do matter. We suspect that personality traits come in to play when false confessions have severe consequences for the individual involved. Under such circumstances, one would expect that false confessions become a low-frequency phenomenon. It may well be the case that only under extreme conditions, trait factors predict who is going to display this phenomenon. Clearly, this issue warrants further study. In more general terms, it is important to clarify the antecedents of false confessions as knowledge about these antecedents may inform forensic practice.

References


