The role of confirmation bias in suspect interviews: A systematic evaluation

Carole Hill*, Amina Memon and Peter McGeorge

School of Psychology, University of Aberdeen, Aberdeen, UK

Purpose. The three studies presented in this paper systematically examined the effect of expectations of guilt on interviewer questioning style, confession, denial rates, and suspects' verbal behaviour during interview.

Method. Undergraduate students were recruited to participate in the three studies. In Study 1, 61 participants formulated questions that they wanted to ask a suspect to determine whether or not they cheated on a task. Prior to formulating their questions, participants were led to believe that the suspect was likely to be guilty or innocent. In Study 2, 45 'innocent' and 'guilty' participants were accused of cheating on a task and were interviewed with either guilt-presumptive questions or neutral questions. In the final study, 58 participants listened to a selection of audiotaped interviews from Study 2. They then rated various aspects of the suspects' verbal behaviour.

Results. As hypothesized expectations of guilt resulted in the formulation of more guilt-presumptive questions even when participants were free to generate their own questions (Study 1). A significant association was found between suspect guilt/innocence and whether they confessed or denied, although there was no association between questioning style and confession or denial rates (Study 2). However, as expected, ratings of independent observers who listened to tape recordings of the suspect interviews indicated an influence of questioning style on the suspects' verbal behaviour whereby a self-fulfilling prophecy effect occurred (Study 3).

Conclusions. These results indicate that expectations of guilt can indeed have an effect on questioning style and that this in-turn can lead to a self-fulfilling prophecy effect.

One of the most valuable tools used by investigating officers when gathering information in criminal investigations is arguably the investigative interview. This is particularly so when other forms of evidence against a suspect are weak or non-existent. In some countries, police investigators have been found to use coercive techniques in suspect interviews, which are similar in nature to those recommended

* Correspondence should be addressed to Ms Carole Hill, School of Psychology, William Guild Building, University of Aberdeen, Old Aberdeen AB24 2UB, UK (e-mail: c.hill@abdn.ac.uk).

by Inbau, Reid, Buckley, and Jayne in their interrogation manual *Criminal Interrogation and Confessions* (Leo, 1996). These techniques are designed to break down the suspects' resistance and persuade them to confess (Inbau, Reid, Buckley, & Jayne 2001).

While early research into police interviews with suspects in Britain found that the use of coercive interview techniques was relatively high (Irving, 1980; Irving & McKenzie, 1989), there is some indication that there has been a reduction in the use of such techniques (Irving & McKenzie, 1989; Baldwin, 1992; Pearse & Gudjonsson, 1996). This is likely to be due to the introduction in England and Wales of the Police and Criminal Evidence Act 1984 (PACE) and the first 'National Package on Investigative Interview Training' in 1993. However, despite the introduction of PACE and investigative interview training, research has still found that a noteworthy percentage of suspect interviews are unsatisfactory. For example, Clarke and Milne (2001) examined 177 interviews conducted with suspects. They found that 10% of the interviews observed were highlighted as possibly breaching PACE. One of the reasons identified for these possible breaches was oppressive behaviour including instances of undue pressure, bullying, and continual challenge. It is therefore important to determine why some interviews continue to be conducted in an unsatisfactory manner despite the introduction of guidelines and investigative interview training.

One of the most prominent findings from research into police interviewing is that officers regularly assume suspects to be guilty, even prior to interviewing them, and that the main aim of the interview is to obtain a confession (Baldwin, 1992; Cherryman, Bull, & Vrij, 2000; Mortimer & Shepherd, 1999; Moston, Stephenson, & Williamson, 1992). The literature on confirmation bias provides an insight into the effect that holding a presumption of guilt could have on suspect interviews. Nickerson (1998) describes confirmation bias as 'the seeking or interpreting of evidence in ways that are partial to existing beliefs, expectations, or a hypothesis in hand' (p. 175). This can involve both seeking information that confirms a belief, while not seeking, and even avoiding, information that disconfirms the belief.

However, the role of confirmation bias in the context of police interviews with suspects has received very little attention. If police officers presume suspects to be guilty prior to interviewing them, they may conduct their interviews by seeking information that confirms this hypothesis and avoiding information that disconfirms their hypothesis. Confirmation bias may therefore be one explanation as to why some suspect interviews continue to be conducted in an unsatisfactory manner. The studies described in this paper were therefore designed to systematically examine the role of confirmation bias in interviews with mock suspects.

Previous research has examined the effects of confirmation bias on hypothesis-testing processes in social interaction. For example, Snyder and Swann (1978) provided participants with a hypothesis about the personal attributes of another person (portrayed as an extrovert or introvert). They were asked to choose questions to ask the person in order to test their hypothesis. Participants in the Snyder and Swann study tended to search for information that would support their hypothesis. Furthermore, Snyder and Swann (1978) found that when the interviewers asked hypothesis-confirming questions, the interviewees tended to behave in ways that appeared to confirm the hypothesis. When a target person's behaviour changes in response to the perceiver's actions, this is regarded as a self-fulfilling prophecy effect (Nickerson, 1998).

Kassin, Goldstein, and Savitsky (2003) have tested this in a more forensically relevant context. In their study, suspects (students) committed a mock crime (stole $100) or took part in an innocent but related act. Interviewers were led to believe that either most suspects were guilty or most suspects were innocent prior to interviewing the suspects. Neutral observers then listened to the taped interviews, made judgments about whether the suspect was guilty or innocent, and rated their impressions of both participants. Results indicated that interviewers with guilty expectations chose more guilt-presumptive questions to ask suspects and that they used more techniques during the first minute of the interview than those with innocent expectations. In addition, interviewers used more techniques overall when interviewing innocent than guilty suspects. Post-interview ratings from both the interviewers and the suspects indicated that interviewers tried harder to get a confession and exerted more pressure on innocent than guilty suspects.

Kassin *et al.* (2003) also found that interviewers with guilty expectations were rated by neutral observers as trying harder to obtain confessions and exerting more pressure on suspects than those with innocent expectations. Observers also saw interviewers as more presumptive of guilt, trying harder to obtain confessions, and exerting more pressure when they were paired with actual innocent than with actual guilty suspects. Suspects in the guilty expectation condition were seen as being more defensive, and there was a tendency for observers to judge them as more guilty, than those in the innocent expectation condition.

Kassin *et al.*'s (2003) study demonstrated that presumptions of guilt affected (1) the type of questions and number of techniques which were used to interrogate a suspect, (2) how the suspect behaved during the interview, and (3) how the suspect was subsequently perceived by a neutral observer. One of the limitations of Kassin *et al.*'s (2003) study is that interviewers chose their questions from a pre-set list determined by the experimenters, so it is not clear whether the same level of bias would exist if interviewers were allowed to generate their own questions (see Trope, Bassok, & Alon, 1984). In addition, it is not clear from this study whether the guilt-presumptive questioning led to the change in suspects' behaviour or whether it was the number of techniques used. Given the research indicating that the number of coercive techniques used in police interviews in the United Kingdom is relatively low, it is important to find out whether simply holding an expectation of guilt (as many of them do despite training advising them to keep an open mind) in the absence of coercive techniques can have the same effect on suspects' behaviour.

Prior to conducting time-consuming studies with a police population, these studies were conducted with a student population to examine whether simply holding a presumption of guilt would have an effect on the interview process. The studies presented in this paper extend Kassin *et al.*'s (2003) work by examining three main issues. The first study investigates how an assumption on the part of an interviewer that a suspect is guilty affects the questions they ask a suspect when they are free to generate their own questions. This issue is of particular interest, given that the situation more closely matches that of real police officers. However, in order to purely examine the effect of holding a presumption of guilt, interviewers were not given the additional direction that they were to obtain a confession. We hypothesized that for participants holding presumptions of guilt, (1) the overall proportion of guilt-presumptive questions formulated would be higher and (2) the content of these questions would be more guilt-presumptive than for participants holding presumptions of innocence. The second study examines whether the questioning styles from Study 1 influence confession and

denial rates. The third study examines the effect of these questioning styles on the suspects' verbal behaviour and whether a self-fulfilling prophecy effect occurs. All three experiments follow BPS ethical guidelines.

STUDY 1

Method

Participants

Sixty-one undergraduate students from the University of Aberdeen (60.66% female) participated individually in return for partial course credit or monetary reimbursement for their time. Participants were randomly assigned to one of the following two conditions: the guilty expectation condition ($N = 30$) or the innocent expectation condition ($N = 31$). The mean age of participants was 20.69 years.

Materials and procedure

Each participant received an instruction sheet describing the task. The instruction sheet informed participants that they were to enact the role of a research assistant's supervisor trying to find out whether or not participants cheated on an experimental task. They were provided with the following scenario:

> Two participants were completing an intelligence test in the same room. Their research assistant left the room for 5 minutes leaving the answer sheet for the task on the desk; thus both participants had the opportunity to cheat by looking at the answer sheet. When their research assistant returned to the room, the answer sheet had been moved from its original position on the desk.

The details in the instruction sheet were identical for each participant, except those in the guilty expectation condition were informed that 'approximately four out of five participants in the study (80%) *look* at the answer sheet' and those in the innocent expectation condition were informed that 'approximately four out of five participants in the study (80%) *do not look* at the answer sheet'. Finally, the instruction sheet informed all participants that their task was to interview one of the participants to determine what happened.

Participants then rated how guilty they thought the suspect they were going to interview was, and how confident they were in their rating of guilt or innocence. Both ratings were made on a 10-point Likert-type scale ranging from 1 (not at all) to 10 (extremely). Participants were then asked to write down 10 questions which they wanted to ask the suspect during the interview in order to determine what happened. Once participants had generated their 10 questions, or as many questions as they could think of, they were informed that they did not actually have to conduct the interview.

Coding and scoring

In total, participants generated 558 questions. Five independent judges (postgraduate students) were recruited to rate the questions so that we could determine whether the questions generated were neutral, presumptive of guilt or presumptive of innocence. In order to reduce the number of questions that required rating by independent judges,

those questions that were essentially the same in content (e.g. Did you notice the answer sheet was left in the room? Did you know the answer sheet had been left on the desk?) were grouped together by the first author and were represented by one summary question. This procedure reduced the number of questions to be rated to 234 (although all 558 questions were represented). The judges were informed that participants formulated the interview questions for the purpose of finding out whether mock suspects were innocent or guilty. They were given the same scenario information as the participants had received in order to set the questions in context; however, they were not given any information about the manipulation of guilty or innocent expectations.

The judges then rated the questions, which were randomly ordered, on a scale of 0–6. The points on the scale were labelled from 0 (extremely presumptive of innocence) to 6 (extremely presumptive of guilt). Presumptive of guilt was defined as questions that assume the suspect is guilty, and presumptive of innocence was defined as questions that assume that the suspect is innocent. Neutral was defined as questions that do not assume the suspect is guilty or innocent. For each question, the five judges' ratings of that question were summed and a mean score for each question was produced. For each participant, the mean scores for all of their questions were summed and a total mean question score for each participant was calculated. This produced a total mean question score that ranged from 0 to 6, as described previously. The proportion of guilt-presumptive, innocent-presumptive, and neutral questions that each participant generated was also calculated.

Results

Within-group agreement
Within-group agreement, the extent to which judges agreed on a set of judgments was calculated for the five judges' ratings of each question. This was conducted to determine whether it was appropriate to take an average of their independent ratings to use in subsequent analyses. The formula used to calculate within-group agreement is

$$r_{\text{wg}(1)} = 1 - \left(\frac{s_{\text{xj}}^2}{\sigma_{\text{EU}}^2} \right)$$

where $r_{\text{wg}(1)}$ is the within-group inter-rater reliability for a group of judges on a single item xj. Glick (1985) suggested that levels of reliability should exceed at least .6 in order for the individual data to be consistent enough to be analysed at the group level. Below this level, the data are only suitable to be analysed at the individual level. This current study therefore considered $r_{\text{wg}} \geq .7$ as an acceptable level, which is comparable with other researchers.

Within-group agreement was found to be below the acceptable level of .7 for 24 of the 234 questions rated. Because some of the 24 questions below the acceptable level were summary questions, made up of questions that were essentially the same in content, a total of 71 questions were discarded. This procedure left 210 questions (487 of the original questions prior to producing summary questions for those that were essentially the same in content) that were deemed consistent enough ($\geq .7$) to take an average measure in order to analyse them at the group level. Those participants ($N = 5$) who had less than seven questions remaining after this procedure were excluded from all subsequent analyses.

Analyses by guilty/innocent expectations

Prior to analysing the question data, participants who did not succumb to the initial manipulation of expectation were excluded from the analyses in order to clearly determine the effect of pre-existing expectations. Those participants in the guilty expectation condition who rated the suspect in the range of 1–5 (not guilty) and those participants in the innocent expectation condition who rated the suspect in the range of 6–10 (guilty) were excluded ($N = 16$). This resulted in 20 participants in the guilty expectation and 20 in the innocent expectation conditions.

The question data were analysed according to the guilty/innocent expectation conditions using a series of independent sample t tests. As would be expected, participants in the guilty expectation condition rated the suspect as significantly more guilty ($M = 7.35$, $SD = 1.09$) than those in the innocent expectation condition ($M = 4.00$, $SD = 1.12$; $t(38) = 9.57$, $p < .001$, $d = 3.02$). Participants' confidence in their ratings of guilt/innocence was significantly higher for those in the guilty expectation condition ($M = 6.10$, $SD = 1.45$) than those in the innocent expectation condition ($M = 4.40$, $SD = 2.72$; $t(28.95) = 2.47$, $p = .02$, $d = 0.68$), equal variances not assumed.

As hypothesized, those participants in the guilty expectation condition formulated a significantly higher proportion of guilt-presumptive questions to ask the suspects ($M = 0.28$, $SD = 0.24$) than those in the innocent expectation condition ($M = 0.16$, $SD = 0.12$; $t(28.04) = 1.95$, $p = .031$ (one-tailed), $d = 0.46$), equal variances not assumed.

The difference in the proportion of innocent-presumptive questions between the guilty ($M = 0.01$, $SD = 0.02$) and innocent expectation conditions ($M = 0.01$, $SD = 0.04$) was non-significant ($t(38) = -0.80$, $p = .43$). Additionally, the difference in the proportion of neutral questions between the guilty ($M = 0.71$, $SD = 0.25$) and innocent expectation conditions ($M = 0.82$, $SD = 0.13$) was non-significant ($t(28.74) = -1.68$, $p = .11$), equal variances not assumed.

As expected, those participants in the guilty expectation condition formulated questions that were given significantly higher ratings for guilt-presumptiveness ($M = 3.62$, $SD = 0.53$) than those in the innocent expectation condition ($M = 3.33$, $SD = 0.20$; $t(24.45) = 2.30$, $p = .03$, $d = 0.59$), equal variances not assumed.

Discussion

This study demonstrates that interviewers holding expectations of guilt generate significantly more guilt-presumptive questions to ask mock suspects than interviewers holding expectations of innocence even when interviewers are free to generate their own questions. These results indicate the presence of confirmation bias where the interviewers in the guilty expectation condition tend to seek information confirming their expectation.

In order to further examine the effects that holding a presumption of guilt could have on mock suspect interviews, Study 2 was designed to examine whether guilt-presumptive questions, as opposed to neutral questions, influence confession and denial rates. Previous studies have examined the effect of various interrogation techniques on false confession rates (e.g. Kassin & Kiechel, 1996; Horselenberg, Merckelback, & Josephs, 2003; Klaver, Gordon Rose, & Lee, 2003). However, more recently Russano, Meissner, Narchet, and Kassin (2005) have developed a new paradigm

that allows examination of both confession and denial rates for guilty and innocent suspects. In this paradigm, participants are given the opportunity to cheat on a problem-solving task and are then interviewed about what occurred. This paradigm therefore involves an intentional act where participants decide whether to cheat or not, rather than the negligent act (pressing the wrong key on a computer causing it to crash) that has been used in the previous studies. Using this procedure, Russano *et al.* (2005) found that minimization techniques (e.g. minimizing the seriousness of the act), the offer of a deal (an explicit offer of leniency if they confessed), and the use of both these techniques together increased both genuine and false confession rates. Study 2 used a slightly modified version of this paradigm in order to examine the effect of guilt-presumptive questioning on confession and denial rates in the absence of interrogation techniques. The questions used in Study 2 were 10 guilt-presumptive questions and 10 neutral questions from Study 1.

STUDY 2

Method

Participants

Sixty-four undergraduate students from the University of Aberdeen were recruited to participate in this study. They all participated individually in return for partial course credit. Due to the potentially stressful nature of the experiment, strict procedures were put in place in order to minimize the likelihood of participants experiencing adverse effects. One of procedures that were put in place was to screen out participants who scored above the 70th percentile on the State Trait Anxiety Inventory (Form Y-1; Spielberger, 1983) (see Materials and Procedure section). The purpose of this procedure was to ensure that highly anxious individuals would not be exposed to the deception involved in the study. In addition, a rating scale (on a scale of 1–10) was designed to measure how agitated, calm, angry, happy, or sad participants felt. Those who scored >7 on the scales measuring agitated, angry, and sad, and those scoring <4 on the scales measuring calm and happy, were also screened out from the study.

In total, 12 participants were screened out and not invited to continue with the study. A further six participants were excluded from the analyses because they indicated during the debriefing that they had been suspicious about the nature of the experiment. Despite careful ethical consideration prior to commencing the study, and the screening procedure that was in place, one participant was also excluded from the analyses because they became upset during the study. This case highlighted that it is not possible to screen out all participants who may become upset and as a result of this the study was terminated. This left 45 participants in the final sample, 53.33% of whom were female. The mean age of these participants was 20.69 years.

Materials and procedure

A confederate arrived to take part in the study at the same time as the participant. They were both informed that they had to complete various personality and emotion measures, and that if their responses were suitable then they would be asked to continue the experiment and complete an intelligence task. It was explained to the participants that the task was quite difficult and therefore it may cause some distress, but that they were free to stop the experiment at any time and leave. While the

participants were completing the emotion and personality measures, the experimenter received a phone call, informed the caller that she was busy but that she would call them back as soon as she could. Once the experimenter had scored the completed personality and emotion measures, those participants who were screened out were informed that they were not required to continue with the rest of the study and were debriefed. Those who were not screened out were informed that they were now to complete the intelligence task and that it was important that they complete it individually.

After giving the participants the instructions for the task, the experimenter said that she had to leave to return the phone call but that she would be back in 5 minutes. During these 5 minutes, the confederate found the answer sheet in a pile of papers that the experimenter had left on the table and asked the participant whether they wanted any answers. For the purpose of this experiment, those participants who accepted the answers were classed as 'guilty' and those who did not as 'innocent'. After 5 minutes, the experimenter re-entered the room, noticed that the answer sheet had been found and moved, and accused the participants of looking at the answer sheet. The experimenter proceeded to make a phone call to her supervisor, a second confederate, who interviewed the participant on the phone about what had happened. The confederate was asked to wait outside at this point and was told that he would be interviewed following the other participant. The participant was asked either 10 guilt-presumptive questions (e.g. Did you feel guilty cheating? Are you ashamed of what you did?) or 10 neutral questions (e.g. What happened when the researcher left the room? Could you tell me more about when you were completing the task?). This design resulted in four conditions: guilty suspect – guilt-presumptive questions ($N = 9$), guilty suspect – neutral questions ($N = 9$), innocent suspect – guilt-presumptive questions ($N = 13$), and innocent suspect – neutral questions ($N = 14$).

Following the interviews that were audio-recorded, participants were fully debriefed. They were also shown a short happy video-clip designed to induce a positive mood state before they left the experimental room. Participants were informed that the telephone interview had been recorded and were asked whether they wanted to consent to their data being used for research purposes. All participants consented to its use.

In addition, they were asked to rate on a scale of 1–10 how much pressure they felt to confess, how forceful they felt they were in their denials, how hard the interviewer tried to get them to confess, and how intimidated and nervous they felt during the interview.

Results

Confession rates

Fisher's exact tests showed that there was a significant association between suspect status (guilty/innocent) and whether the suspects confessed or denied ($p < .01$), but that there was no significant association between interviewer presentation (guilt-presumptive/neutral) and whether the suspects confessed or denied ($p = .19$). Table 1 shows the frequency of confessions and denials according to condition.

Participant ratings of their interview experience

A 2 (suspect status: guilty, innocent) × 2 (questioning style: neutral, guilt-presumptive) between-subject ANOVA revealed that those participants who were asked

Table 1. Study 2: Frequency of confessions and denials according to condition

	Innocent suspect and neutral questions	Innocent suspect and guilt-presumptive questions	Guilty suspect and neutral questions	Guilty suspect and guilt-presumptive questions
Denial	14	13	8	5
Confession	0	0	1	4

guilt-presumptive questions, when compared with those asked neutral questions, felt significantly more pressurized ($F(1, 41) = 10.22, p = .01, \eta_p^2 = .20$), during the interview and that they felt the interviewer tried harder to get a confession ($F(1, 41) = 2.45, p < .01, \eta_p^2 = .51$). There was no effect for suspect status and no interaction. There were no significant effects for interviewer presentation or suspect status in how nervous or intimidated participants felt or how forceful they thought they were in their denials.

Discussion

While the guilt or innocence of suspects was associated with whether they confessed or denied cheating on the task, the style of questioning was not associated with confession or denial rates. Therefore, it appears that a guilt-presumptive questioning style is not associated with more confessions than a neutral questioning style.

In terms of the participants' experiences during the telephone interview, those who were asked guilt-presumptive questions reported feeling significantly more pressurized, and feeling that the interviewer was trying harder to get a confession than those who were asked neutral questions. However, there were no self-reported differences in experiences between guilty and innocent suspects. One of the reasons why there may have been null effects is that we screened out the most anxious individuals due to the nature of the study.

The third study explored whether the two different questioning styles utilized in Study 2 (guilt-presumptive or neutral) had an effect on independent observers' ratings of the suspects' behaviour during the interview. In particular, Study 3 examined (1) whether the presumption of guilt led to a self-fulfilling prophecy effect and (2) whether it led to certain behaviours associated with deception (plausibility, nervousness, and defensiveness) that could lead observers to perceive suspects as suspicious (DePaulo *et al.*, 2003; Vrij, Akehurst, & Knight, 2006).

STUDY 3

Method

Participants

Sixty-nine undergraduate students from the University of XXX were recruited to participate in this study. They participated in groups in return for partial course credit. In total, 11 participants were excluded from the analyses ($N = 58$) because they indicated that they had recognized one or more of the voices on the audiotaped interviews (see Materials and procedure section).

Materials and procedure

In order to assess the extent to which the answers of the mock suspects provided behavioural confirmation for the initial expectation of the interviewer, participants listened to a selection of tape recordings of the interviews. Participants were informed that they were to listen to eight brief interviews with individuals who were being interviewed about the possibility that they cheated on an experimental task. In addition, they were told that they would hear only the individuals' responses to the interviewer, and not the interviewer's questions. Participants were therefore blind to the previous manipulation of whether the suspects were asked guilt-presumptive or neutral questions.

Immediately after each interview, participants were asked to provide ratings of their perception of the suspects' behaviour during the interview. The following measures were obtained on a scale of 1–10: how guilty/innocent they believed the suspect was and how confident they were in this judgment of guilt/innocence. This was to determine whether the suspect reacted to the questions in a way that confirmed the interviewer's initial expectation of guilt/innocence, that is resulting in a self-fulfilling prophecy. In terms of the suspects' behaviour, participants also rated how plausible the suspect was in their denials, and how nervous and defensive the suspect was during the interview.

Eight audiotaped interviews from Study 2 were randomly selected for use in this study based on the following criteria: (1) interviewees had consented for their audiotapes to be used in this study; (2) they did not contain a confession, as this would obviously influence the participants' responses; and (3) there had to be two audio-recordings from each of the following four conditions: guilty suspect – guilt-presumptive questions, guilty suspect – neutral questions, innocent suspect – guilt-presumptive questions, and innocent suspect – neutral questions. Although the audiotapes contained no verbal information from which identification of the person on the tape could be made, it was possible that their voices would be recognized. Participants were therefore asked whether they recognized the voices on the audiotape. Any who did were excluded.

Results

Participants' perceptions of the suspects' behaviour

In order to examine participants' perceptions of the suspects' behaviour, a series of 2 (interviewer presentation: guilt-presumptive and neutral questions) × 2 (suspect status: guilty, innocent) repeated measure ANOVAs were conducted.

Suspect guilt/innocence

In terms of how guilty/innocent participants perceived the suspects, there was no effect of suspect status ($F(1, 57) = 0.82$, $p = .37$, $\eta_p^2 = .01$). However, interviewer presentation had an effect ($F(1, 57) = 66.56$, $p < .01$, $\eta_p^2 = .54$) indicating that suspects who responded to guilt-presumptive questions were perceived by the participants to be guiltier than those who responded to neutral questions. There was also a significant interaction ($F(1, 57) = 12.01$, $p < .01$, $\eta_p^2 = .17$) indicating that innocent suspects who responded to guilt-presumptive questions were perceived as significantly more guilty ($M = 4.85$, $SD = 1.62$) than guilty suspects who responded to

guilt-presumptive questions ($M = 5.65$, $SD = 1.71$; $t(57) = 2.59$, $p = .01$, $d = 0.34$). There was no effect of how confident the participants were in their judgment of how guilty/innocent they perceived suspects to be for either interviewer presentation ($F(1, 57) = 2.60$, $p = .11$, $\eta_p^2 = .04$) or suspect status ($F(1, 57) = 0.06$, $p = .80$, $\eta_p^2 = .001$). Therefore, participants were equally confident that the suspect was guilty/innocent, regardless of whether they were actually guilty or innocent.

Suspect nervousness during interview
In terms of how nervous suspects were during the interview, interviewer presentation had an effect ($F(1, 57) = 101.33$, $p < .01$, $\eta_p^2 = .64$). There was no effect for suspect status ($F(1, 57) = 1.17$, $p = .28$, $\eta_p^2 = .02$) and no interaction ($F(1, 57) = 0.44$, $p = .51$, $\eta_p^2 = .01$). Thus, individuals were perceived as more nervous when they responded to guilt-presumptive than to neutral questions.

Suspect defensiveness during interview
Interviewer presentation had an effect on how defensive suspects were perceived to be during the interview ($F(1, 57) = 258.96$, $p < .01$, $\eta_p^2 = .82$) as did suspect status ($F(1, 57) = 71.47$, $p < .01$, $\eta_p^2 = .56$). There was no interaction ($F(1, 57) = 2.14$, $p = .15$, $\eta_p^2 = .04$). Thus, those suspects who responded to guilt-presumptive questions were perceived by participants to be more defensive than those who responded to neutral questions, and innocent suspects were perceived by participants to be more defensive than guilty suspects.

Suspect plausibility during interview
In terms of how plausible suspects' denials were perceived to be, there was an effect for interviewer presentation ($F(1, 57) = 34.95$, $p < .01$, $\eta_p^2 = .38$) and suspect status ($F(1, 57) = 4.92$, $p = .03$, $\eta_p^2 = .08$). An interaction was also found ($F(1, 57) = 13.32$, $p < .01$, $\eta_p^2 = .19$) indicating that innocent suspects were viewed as significantly less plausible ($M = 5.57$, $SD = 1.56$) than guilty suspects ($M = 4.59$, $SD = 1.51$) when responding to guilt-presumptive questions ($t(57) = -0.38$, $p < .01$, $d = 0.49$), and that innocent suspects were seen as less plausible when responding to guilt-presumptive questions ($M = 5.57$, $SD = 1.56$) than when responding to neutral questions ($M = 3.86$, $SD = 1.39$; $t(57) = 6.89$, $p < .01$, $d = 0.9$).

Discussion

The results from Study 3 highlight that the mock suspects' answers to the interviewer's questions provided behavioural confirmation for the interviewer's initial expectations. Even when independent observers are blind to the type of questions asked (i.e. guilt-presumptive or neutral), they judge those suspects who responded to guilt-presumptive questions as more guilty than those to neutral questions. In addition, independent observers were just so confident in their judgments of guilt/innocence, regardless of whether suspects were actually guilty or innocent indicating that they were confident in their judgment even when this was in error.

Another notable finding in Study 3 was that when innocent suspects (when compared with guilty suspects) responded to guilt-presumptive questions, the self-fulfilling

prophecy effect was more pronounced indicating that some aspect of their verbal response was perceived as indicative of guilt.

With regard to the characteristics associated with deception, suspects were perceived to be more nervous, more defensive, and less plausible when responding to guilt-presumptive than to neutral questions. In addition, innocent suspects were perceived to be more defensive and less plausible than guilty suspects when responding to guilt-presumptive questions. These findings indicate that guilt-presumptive questioning does indeed lead suspects to be viewed as more deceptive, particularly when they are actually innocent.

GENERAL DISCUSSION

The first study extended Kassin *et al.*'s (2003) study by demonstrating that even when interviewers are free to generate their own questions, rather than selecting them from a pre-set list, interviewers with expectations of guilt ask more guilt-presumptive questions than those with innocent expectations indicating that confirmation bias is evident. In addition, interviewers were found to be more confident in their judgments of guilt/innocence when they presumed the suspect to be guilty than when they presumed the suspect to be innocent.

While Study 1 was conducted with student interviewers, the results indicate that police officers judging an innocent suspect to be guilty may search for hypothesis-confirming information while being highly confident that their judgment of guilt is correct. If police officers could accurately discriminate between behaviour that was indicative of guilt or innocence in suspects, then this situation would not be so problematic. However, psychological research indicates that when attempting to detect deception, either by verbal or non-verbal cues, laypeople and police officers perform only slightly better than chance (Vrij, 2000; Bond & DePaulo, 2006).

A number of studies have identified an *investigator bias effect*, where trained and experienced police officers have a bias towards judgements of deceit. For example, Meissner and Kassin (2002) found that greater prior experience and training was significantly correlated with a tendency to judge suspects as deceitful rather than truthful, but that it was not correlated with the accuracy of their judgments. A further study by Kassin, Meissner, and Norwick (2005) found that law enforcement officers were predisposed to believe both true and false confessions, and that the bias was particularly evident in investigators with more law enforcement experience and those who had received special training in interrogation. In addition, law enforcement experience and training elevated law enforcement officers' confidence in these judgments but failed to improve their accuracy in detection. Therefore, police officers with more experience may be particularly susceptible to confirmation biases.

The second study was designed to examine whether the two different questioning styles would have an effect on confession or denial rates. The findings indicated that there was no association between the questioning style and the tendency for suspects to confess or deny cheating. Given previous research findings of false confession rates between 8% and 20.3% in studies utilizing similar paradigms (Russano *et al.*, 2005; Horselenberg *et al.*, 2006), it appears that the use of guilt-presumptive questions in the absence of other interrogation techniques is not sufficient to elicit false confessions. It also appears that the use of guilt-presumptive questions does not increase genuine confessions.

The results from the third experiment indicated that mock suspects reacted to guilt-presumptive questions in a way that appeared to confirm the guilty expectation. Our results may be due to the finding that those responding to guilt-presumptive questions appeared to be more nervous, more defensive, and less plausible in their denials than those responding to neutral questions. These are cues that have been found to be associated with deception to varying extents. For example, liars have been found to be less plausible, less co-operative, make more negative statements and complaints, and to be more nervous and tense than truth tellers (DePaulo *et al.*, 2003; Vrij, Akehurst, & Knight, 2006).

Another notable finding in the third experiment was that when innocent suspects responded to guilt-presumptive questions, the self-fulfilling prophecy effect was more pronounced indicating that some aspect of their verbal response was perceived as indicative of guilt. Again the results indicate that this could be because innocent suspects responding to guilt-presumptive questions were perceived to be more defensive and less plausible in their denials than guilty suspects. These findings indicate that the use of guilt-presumptive questions may lead suspects to behave in a manner consistent with deceptive behaviour, and that this is particularly the case when the suspect is innocent. These findings are particularly concerning given that independent observers were just so confident in their judgments of guilt/innocence, regardless of whether the suspect was actually guilty or innocent. In other words, erroneous judgments of guilt were made with confidence.

The apparent differences in verbal behaviour are robust enough to be picked up by independent observers blind to the suspects' actual guilt or innocence and the interviewers questioning style. This result indicates that the suspect must respond differently to the interviewer, in terms of their verbal behaviour, depending on the type of questions asked, therefore indicating a self-fulfilling prophecy effect. This finding is supportive of Kassin *et al*.'s (2003) results where a self-fulfilling prophecy effect was found in suspect interviews. However, this study additionally demonstrates that guilt-presumptive questions produce a self-fulfilling prophecy effect on their own without the use of any interrogation techniques or coercion. In terms of suspect interviews conducted by police officers, these results indicate that if police officers presume the suspect to be guilty and conduct interviews that are guilt-presumptive in nature, then the suspect is likely to unwittingly behave in a way that appears to confirm their guilt, particularly if they are actually innocent.

Due to the difficulty of conducting ethically and scientifically sound research into actual suspect interviews, the current research was carried out with undergraduate students as participants and researchers playing the role of interviewers. There are a number of reasons why these studies may underestimate the role of confirmation bias in suspect interviews. Police officers are highly motivated to solve crimes and obtain confessions whereas the participants in Study 1 were unlikely to have had the same level of motivation and were told that their aim was to find out what happened rather than obtain a confession. Also, police officers are often under extreme pressure from their organization, as well as the public, to solve crimes and bring the perpetrator to justice as quickly as possible. There was no such pressure for participants in Study 1. Furthermore, the participants in Study 1 received information that did not contain any evidence against the suspect they were to interview. Police officers, on the other hand, will often have some form of evidence indicating the suspects' guilt before questioning them, regardless of how weak the evidence is, or they would be interviewed as a witness rather than a suspect. Research has also indicated that experienced police

officers tend to have higher rates of suspicion and tend to believe they can make judgments of deception and guilt with a higher degree of confidence than police recruits or undergraduate students (Kassin, Meissner, & Norwick, 2005; Masip, Alonso, Garrido, & Anton, 2003).

Another factor that could potentially exacerbate the effect of confirmation bias in a police setting, when compared with the current study, is the way in which information is gathered. For example, Jonas, Schulz-Hardt, Frey, and Thelen (2001) found that there is a significantly stronger preference for hypothesis-confirming information when the information is gathered sequentially as opposed to simultaneously. In addition, the stronger preference for hypothesis-confirming information is due to the seeking of more supporting information rather than less conflicting information.

The research carried out by Jonas *et al.* (2001) has potentially important implications for suspect interviews. In the current study, interviewers generated their questions in a simultaneous manner, that is they generated all their questions prior to the interview. However, in suspect interviews, officers typically gather information or evidence in a sequential manner, that is they formulate a question, process the answer to the question, and then formulate another question. If the interview is being commenced with a presumption of guilt, then it is likely that the bias to seek information consistent with this expectation could be exacerbated due to the sequential nature of the information gathering. In order to explore these hypotheses, it is necessary to conduct further research on the role of confirmation bias in suspect interviews.

References

Baldwin, J. (1992). *Video taping police interviews with suspects: A national evaluation. Police Research Series Paper 1*. London: Home Office Police Department.

Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review, 10*, 214–234.

Cherryman, J., Bull, R., & Vrij, A. (2000, April). *How police officers view confessions: Is there still a confession culture*. Paper presented at the European Conference in Psychology and Law, Cyprus.

Clarke, C., & Milne, R. (2001). *National evaluation of the PEACE investigative interviewing course. Police Research Award Scheme. Report No. PRAS/149*: Institute of Criminal Justice Studies, University of Portsmouth.

DePaulo, B. M., Lindsay, J. L., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129*, 74–118.

Glick, W. H. (1985). Conceptualising and measuring organisational and psychological climate: Pitfalls in multilevel research. *Academy of Management Review, 10*, 601–616.

Horselenberg, R., Merckelbach, H., & Josephs, S. (2003). Individual differences and false confessions: A conceptual replication of Kassin and Kiechel (1996). *Psychology, Crime and Law, 9*, 1–8.

Horselenberg, R., Merckelbach, H., Smeets, T., Franssens, D., Ygram Peters, G., & Zeles, G. (2006). False confessions in the lab: Do plausibility and consequences matter? *Psychology, Crime and Law, 12*, 61–75.

Inbau, F. E., Reid, J. E., Buckley, J. P., & Jayne, B. C. (2001). *Criminal interrogation and confessions* (4th ed.). Gaithersberg, MD: Aspen.

Irving, B. (1980). *Police interrogation: A case study of current practice. Research studies no. 2*. London: HMSO.

Irving, B. L., & McKenzie, I. K. (1989). *Police interrogation: The Effects of the Police and Criminal Evidence Act 1984*. London: The Police Foundation.

Jonas, E., Schulz-Hardt, S., Frey, D., & Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: An expansion of dissonance theoretical research on selective exposure to information. *Journal of Personality and Social Psychology, 80*, 557–571.

Kassin, S. M., Goldstein, C. C., & Savitsky, K. (2003). Behavioural confirmation in the interrogation room: On the dangers of presuming guilt. *Law and Human Behaviour, 27*, 187–203.

Kassin, S. M., & Kiechel, K. L. (1996). The social psychology of false confessions: Compliance, internalisation, and confabulation. *Psychological Science, 7*, 125–128.

Kassin, S. M., Meissner, C. A., & Norwick, R. J. (2005). 'I'd know a false confession if I saw one' A comparative study of college students and police investigators. *Law and Human Behaviour, 29*, 211–227.

Klaver, J., Gordon Rose, V., & Lee, Z. (2003, July). *Differential effects of minimisation and maximisation interrogation techniques and the role of plausibility in an experimental false confession paradigm*. Paper presented at the International and Interdisciplinary Psychology and Law Conference, Edinburgh, Scotland.

Leo, R. A. (1996). Inside the interrogation room. *Journal of Criminal Law and Criminology, 86*, 266–303.

Masip, J., Alonso, H., Garrido, E., & Anton, C. (2003). *Generalised communicative suspicion among police officers*. Paper presented at the International and Interdisciplinary Psychology and Law Conference, Edinburgh, Scotland.

Meissner, C. A., & Kassin, S. M. (2002). 'He's guilty!' Investigator bias in judgments of truth and deception. *Law and Human Behaviour, 26*, 469–480.

Mortimer, A., & Shepherd, E. (1999). Frames of mind: Schemata guiding cognition and conduct in the interviewing of suspected offenders. In A. Memon & R. Bull (Eds.), *Handbook of the psychology of interviewing* (pp. 293–315). Chichester: Wiley.

Moston, S., Stephenson, G. M., & Williamson, T. M. (1992). The effects of case characteristics on suspect behaviour during police questioning. *British Journal of Criminology, 32*, 23–40.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*, 175–220.

Pearse, J., & Gudjonsson, G. H. (1996). Police interviewing techniques at two south London police stations. *Psychology, Crime and Law, 3*, 63–74.

Russano, M. B., Meissner, C. A., Narchet, F. M., & Kassin, S. M. (2005). Investigating true and false confessions within a novel experimental paradigm. *Psychological Science, 16*, 481–486.

Snyder, M., & Swann, W. B. (1978). Hypothesis-testing processes in social interaction. *Journal of Personality and Social Psychology, 36*, 1202–1212.

Speilberger, C. D. (1983). *Manual for the State-Trait Anxiety Inventory (STAI)*. Palo Alto, CA: Consulting Psychologists Press.

Trope, Y., Bassok, M., & Alon, E. (1984). The questions lay interviewers ask. *Journal of Personality, 52*, 90–105.

Vrij, A. (2000). *Detecting lies and deceit: The psychology of lying and the implications for professional practice*. Chichester: Wiley.

Vrij, A., Akehurst, L., & Knight, S. (2006). Police officers', social workers', teachers' and the general public's beliefs about deception in children, adolescents and adults. *Legal and Criminological Psychology, 11*, 297–312.