

# The Presumption of Guilt in Suspect Interrogations: Apprehension as a Trigger of Confirmation Bias and Debiasing Techniques

Moa Lidén, Minna Gräns, and Peter Juslin  
Uppsala University

This research tests whether a police officer's decision to apprehend a suspect triggers confirmation bias during an interrogation. The study also tests two strategies to reduce confirmation bias: (1) decoupling decision to apprehend from interrogation and (2) reducing cognitive load for the interrogating police officer. In Experiment 1, Swedish police officers ( $N = 60$ ) were faced with 12 scenarios in which they either had to decide for themselves whether to apprehend a suspect or were informed about the corresponding decision by another police officer or a prosecutor. Participants then prepared questions for a suspect interrogation and evaluated the trustworthiness of the suspect's denial or confession. The same method was used in Experiment 2 but with law and psychology students ( $N = 60$ ) as participants. In Experiment 3, psychology students ( $N = 60$ ) prepared interrogation questions either by freely producing their own or by choosing questions from a preset list. Overall, apprehended suspects were interrogated in a more guilt presumptive way and rated as less trustworthy than non apprehended suspects. However, the tested debiasing techniques, primarily reducing cognitive load for the interrogating police officer, hold some potential in mitigating this bias.

## Public Significance Statement

Even though suspects have a right to be presumed innocent until they have been convicted, this research suggests that, already during the criminal investigation, police officers ask more guilt presumptive questions if the suspects are apprehended. Possible ways to reduce this guilt presumption is to have another police officer conduct the interrogation than the one who apprehended and to make the interrogation less cognitively demanding for the interrogating police officer.

**Keywords:** confirmation bias, police, apprehension, interrogation, investigation

**Supplemental materials:** <http://dx.doi.org/10.1037/lhb0000287.supp>

It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts. (Doyle, 2001, p. 14)

Twisting and turning of facts to suit already existing theories describes the behavior referred to as *confirmation bias*, that is, a cognitive tendency to search for and evaluate information in ways that are partial to an already formed hypothesis (Nickerson, 1998). Such partiality clearly deviates from legal standards in criminal cases which require legal actors to be objective and impartial. In cases in which the dominating hypothesis is incorrect, confirmation bias is associated with wrongful suspicions or even wrongful convictions, if legal actors cannot free themselves of the hypoth-

esis in due time. Because a wrongful conviction is the worst conceivable type of error in criminal cases, examining potential triggers of confirmation bias as well as ways to prevent or mitigate the bias are two important tasks for scientific research.

The purpose of this study is to examine the role of confirmation bias in suspect interrogations, how this bias is affected by the decision (and by whom the decision is made) to apprehend a suspect, and evaluate two possible techniques to reduce confirmation bias in suspect interrogations. It adds to the already existing research in this area by identifying a trigger of confirmation bias that is inherent in criminal investigation (the apprehension) and also tests two novel potential debiasing techniques. In the following text, we first describe the phenomenon of confirmation bias and what is known about its causes. Thereafter, we discuss how confirmation bias may enter into the judicial process and ways to mitigate the confirmation bias through so-called debiasing techniques. Finally, we report three experiments that investigate the role of confirmation bias in suspect interrogations and evaluate two possible debiasing techniques.

## Confirmation Bias

Confirmation bias denotes a partiality in relation to an already formed hypothesis (Nickerson, 1998). This partiality expresses

This article was published Online First July 2, 2018.

Moa Lidén and Minna Gräns, Faculty of Law, Uppsala University; Peter Juslin, Department of Psychology, Uppsala University.

The data from Experiment 1 were reported during a 15-min oral presentation at the American Psychology–Law Society 2016 Annual Conference in Atlanta, Georgia (March 10–12, 2016). However, the article is original and not previously published.

Correspondence concerning this article should be addressed to Moa Lidén, Trädgårdsgatan 20, 753 09 Uppsala, Sweden. E-mail: [moa.liden@jur.uu.se](mailto:moa.liden@jur.uu.se)

itself in a limited search for and evaluation of information that confirms that the hypothesis is correct, even if it is seriously flawed. Information opposing the hypothesis is either ignored or interpreted in ways that conform to the predetermined conclusion. This one-sided reasoning happens subconsciously, which means that even a person who consciously attempts to be objective may in fact be biased.

Confirmation bias has been recognized in a range of settings (Wason, 1966, regarding general problem solving and reasoning; Pines, 2006, regarding medicine; Fugelsang, Stein, Green, & Dunbar, 2004, regarding science) but its driving mechanisms are still relatively poorly researched. Some researchers point to cognitive explanations. For instance, in his heuristic, analytical model of reasoning, Evans (2006) emphasized capacity limits that simply make it too cognitively demanding to seriously consider more than one hypothesis at the time. Similarly, it has been suggested that limitations in working memory capacity undermine the ability to test multiple hypotheses as only one hypothesis (one interpretation of evidence) can be held in working memory at the same time (Doherty & Mynatt, 1990; Mynatt, Doherty, & Dragan, 1993; Mynatt, Doherty, & Sullivan, 1991). More recent research indicates a confirmation bias in visual search because stimuli matching a template was prioritized and attended to even when such a strategy was not optimal for the task at hand (Carrasco, 2011; Rajsic, Wilson, & Pratt, 2015). Others emphasize social explanations, for example that humans do not reason to find the truth but instead to convince others that they are right (Mercier, 2016; Mercier & Sperber, 2011). There are also motivational explanations according to which humans are motivated to defend their ideas and behaviors in order to maintain control and self-esteem (Pyszczynski & Greenberg, 1987). Whereas the social and motivational explanations support the notion of confirmation bias as a self-enhancing bias, the cognitive explanations point to more general causes and suggest that confirmation bias is not dependent on whether and how much an individual him- or herself has invested in a hypothesis.

### Confirmation Bias in Criminal Investigations

In criminal cases, researchers have identified several different manifestations of confirmation bias, for instance in *suspect-driven investigations* where someone becomes a suspect for reasons that are not, in objective terms, explained by the evidence in the case, and the inquiry is then limited to the evidence that links that suspect to the crime (Wagenaar, van Koppen, & Crombag, 1993). Furthermore, prior information regarding an alleged crime seems to influence the forms of evidence investigators secure during crime scene investigations (van den Eeden, de Poot, & van Koppen, 2016), and a growing body of research describes the influence of contextual information on forensic experts' analysis of fingerprints and DNA-mixtures (Dror & Cole, 2010; Kassin, Dror, & Kukucka, 2013). Given that forensic evidence is considered the most reliable type of evidence, this so-called forensic confirmation bias poses a particularly potent risk of wrongful convictions.

In the interrogation context, research indicates that interrogators' expectations that a suspect is guilty are associated with a guilt presumptive questioning style. Kassin, Goldstein, and Savitsky (2003), who manipulated guilt expectations by providing different base rates of guilt and innocence, found that interrogators with

guilty expectations selected more guilt-presumptive questions and used more interrogation techniques aimed at eliciting confessions, particularly when interrogating innocent suspects. Furthermore, suspects that were interrogated with a guilt presumptive questioning style were perceived as more defensive and as somewhat more guilty. In a similar study by Hill, Memon, and McGeorge (2008), guilt expectations, also manipulated using different base rates of guilt and innocence, set in motion a process of behavioral confirmation that influenced not only the interrogators' behavior, but also that of the suspects. In addition, the guilt presumptive questioning style seemed to influence the suspects' verbal behavior, making them appear to be significantly more nervous, more defensive, less plausible, and therefore more likely to be guilty. The effects were more pronounced for innocent than for guilty suspects. Furthermore, expectations of guilt potentiate the risk of confrontational or even manipulative interrogation, which is associated with false confessions (Davis & Leo, 2006; Leo & Drizin, 2010) and nonstrategic use of evidence, such as revealing the evidence to the suspect before he or she has provided his or her account (Hartwig, Granhag, Strömwall, & Vrij, 2005; Kassin, 2005; Leo & Drizin, 2010). This makes it even more difficult to decide whether the suspect is lying or telling the truth.

This study is conducted in the Swedish legal context where there is no nationally accepted model for how suspect interrogations should be carried out (unlike, e.g., in Norway, the Netherlands, and the United Kingdom; Granhag, Strömwall, & Cancino Montecinos, 2013). Experimental and interview studies indicate that there is a great variability in the techniques used by Swedish police officers, where Reid-inspired techniques (i.e., techniques that are aimed at eliciting a confession, such as asking closed ended questions and interrupting the suspect) are quite common (Hartwig et al., 2005; Kronkvist, 2013). Thus, the risk of a guilt presumptive questioning style is likely to be enhanced in the Swedish legal context.

Because the biasing effect of manipulating guilt expectations on interrogation style has already been tested and documented (Hill et al., 2008; Kassin et al., 2003), an important next step is to study in which situations guilt expectations are formed naturally during criminal investigations. Detecting triggers of confirmation bias that are inherent in criminal cases is necessary not only to understand how this bias can express itself, but also to identify realistic and functional debiasing techniques. A common situation in criminal investigations that may trigger confirmation bias is when police officers are forced to hypothesize about a suspect's guilt at an early stage, that is, "before one has data" (Doyle, 2001, p. 14). For instance, when deciding whether to apprehend a suspect, a police officer has to assess whether there is at least reasonable suspicion for the suspect's guilt. In the Swedish legal context, the apprehension is part of an interrelated construction of law that, from an early stage of the inquiry, focuses on the conviction of a suspect. An apprehension explicitly requires that there are grounds for an arrest, and the arrest requires that there are grounds for detention. Similarly, but implicitly, a detention requires that the suspect is expected to be found guilty. Because decisions to apprehend suspects are often formed on scant and ambiguous information, there is a risk of *misclassification errors* (i.e., erroneous decisions that result in an innocent person being found guilty; Leo & Drizin, 2010). Thus, apprehensions pose a risk of both confirmation bias and wrongful convictions, particularly

when combined with other risk factors of selectivity in information processing, such as time pressure (Ask & Granhag, 2007; Fahsing & Ask, 2013) and *police culture* (i.e., thought patterns and ways of working that are common to police officers; Ask & Granhag, 2005; Mortimer & Shepherd, 1999).

### Debiasing Techniques

Even though avoiding wrongful convictions ought to be quite a strong incentive for examining debiasing techniques, the existing research is still scarce and the findings are mixed (Kassin et al., 2013; Zenker & Dahlman, 2016; Zenker, Dahlman, Bååth, & Sarvar, 2016). Understanding how to mitigate bias requires an understanding of why it occurs. Thus, a reasonable starting point is the different explanations of confirmation bias and what they indicate regarding potential debiasing techniques. In the following, such a review results in the identification of two potential debiasing techniques that are tested within the frames of this study, namely (1) changing the decision maker and (2) changing interrogation questioning mode (production and detection vs. detection only).

The first potential debiasing technique, changing the decision maker is motivated mainly by the notion of confirmation bias as a self-enhancing bias (social/motivational explanations). If confirmation bias can be explained as a way of protecting the individual's self-esteem, differences in mindset would be expected in relation to own and others' decisions. In previous research, there are some indications of stronger confirmatory tendencies in relation to self-generated hypotheses compared with hypotheses generated by others (Klahr, Dunbar, & Fay, 1990; Schunn & Klahr, 1993 regarding hypothesis testing when trying to discover the function of an unknown command in a computer device; Haverkamp, 1993 regarding the counseling context). However, because these studies concern contexts distinct from criminal investigations, it is uncertain whether the findings are at relevant. In fact, research on decision making in groups also points in the opposite direction; that is, changing the decision maker can exacerbate confirmation bias because of group processes such as groupthink and group polarization (Isenberg, 1986; Janis, 1982; Myers & Lamm, 1975). If Swedish police culture involves high levels of internal solidarity (Kjöller, 2016), police officers may be less critical (consciously or unconsciously) in relation to their colleague's hypotheses. This is also aligned with Darley and Latané's (1968) theory on *diffusion of responsibility*, that is, that an individual is less likely to take responsibility for action or inaction when others are present as compared with when the individual is alone. Also, the extent of this influence is likely to vary depending on the form of the relationship between the police officers. For instance, if the officer who decides to apprehend a suspect has a higher occupational status than the interrogating officer, the influence might be greater compared with that in the opposite situation.

In real-life criminal investigations, it is hardly feasible that a colleague's (or a prosecutor's) apprehension of a suspect can be kept secret from the interrogating police officer and therefore the technique tested here does not involve blinding either. Instead, we test whether a police officer who knows about an apprehension (or a non apprehension) acts differently in relation to the suspect, depending on whether the police officer him- or herself decided

about the apprehension compared with if their police colleague or the prosecutor made that decision.

The second potential debiasing technique, changing interrogation mode (production and detection vs. detection only), is motivated primarily by cognitive explanations. If confirmation bias is due to cognitive limitations (Carrasco, 2011; Doherty & Mynatt, 1990; Evans, 2006), then a stronger confirmation bias would be expected for more cognitively demanding tasks. This debiasing technique suggests that there are (at least) two cognitively demanding tasks in a standard suspect interrogation: (1) to identify and formulate the most relevant questions and (2) to monitor and screen the formulated questions for being too leading and guilt presumptive (assuming that most interrogators are aware of the potential risk of posing leading questions).

In a similar vein, it is often proposed that one reason why recall is a more cognitively demanding task than recognition (Tulving, 2000) is that recall requires a person both to generate the response option and to monitor it for being relevant and appropriate, while recognition only calls on the latter capacity (Jacoby, Toth, & Yonelinas, 1993; Jacoby, Toth, Yonelinas, & Debnier, 1993; Mandler, 1980). Similarly, recognizing the difference between guilt presumptive and neutral questions and only choosing the neutral ones from a checklist (detection of guilt presumption) is a cognitively less demanding task than freely producing questions, which simultaneously have to be screened for guilt presumption (question production and detection of guilt presumption). Assuming that police officers are typically aware of the risk of posing guilt presumptive questions but have limited cognitive capacities, reducing or eliminating the cognitive load required for producing the questions will make more cognitive resources available to better achieve the second task of detecting guilt presumption. Thus, in this study, cognitive load is operationalized through a less cognitively demanding interrogation questioning mode that uses a predetermined list (detection of guilt presumption) as compared with a more cognitively demanding questioning mode that uses self-generated questions (question production and detection of guilt presumption). If using the predetermined list is associated with no or lower levels of guilt presumption, it would indicate that reducing police officers' cognitive load could be a strategy to reduce confirmation bias.

### The Experiments

This research sets out to test whether police officers who have apprehended a suspect carry out more guilt presumptive interrogations than do police officers who have not apprehended the suspect. The notions that confirmation bias might be a self-enhancing bias, a general cognitive phenomenon, or both motivate testing of several different types of debiasing techniques. Thus, this research tests two techniques: In Experiment 1, we change the decision maker between apprehension and interrogation, and in Experiment 3, we change the interrogation mode (production vs. detection: i.e., freely self-generated questions or choosing questions from a preset list, respectively). In Experiment 1, police officers were presented with 12 scenarios inspired by real criminal cases. In four of these scenarios, the police officers had to decide whether a suspect should be apprehended; in four other scenarios, police officers were informed about another police officer's apprehension decision; in yet another four scenarios, police officers

were informed about the prosecutor's decision. The 12 scenarios were rotated across the conditions so that each scenario appeared the same number of times in each condition. Police officers were then asked to prepare two different types of questions (i.e., freely self-generated questions and questions chosen from a preset list) for an interrogation with the suspect to assess the trustworthiness of his or her confession or denial. As a result of the findings in Experiments 1 and 2, Experiment 3 systematically tested the change of interrogation mode (production and detection vs. detection) as a debiasing technique. Apart from that, the same method was used in all three experiments.

In line with prior research (Hill et al., 2008; Kassir et al., 2003; Leo & Drizin, 2010), it is expected that suspects who have been apprehended and those who have not been apprehended will be interrogated differently, even though the described scenarios and evidentiary bases are identical. Specifically, because of the stronger belief that the suspect is guilty after apprehension, it is hypothesized, as follows, that the police officers would formulate more guilt presumptive questions.

*Hypothesis 1:* Participants will display a stronger guilt presumption (in interrogation questions and trustworthiness ratings) with regard to apprehended than to non apprehended suspects.

The hypotheses for the aforementioned potential debiasing techniques are as follows:

*Hypothesis 2:* Participants will demonstrate a stronger guilt presumption in relation to their own decision about apprehension than to another police officer's or a prosecutor's decision.

*Hypothesis 3:* Participants will show a stronger guilt presumption when freely producing their own questions (production and detection) as opposed to when they choose questions from a preset list (detection).

If confirmation bias can correctly be described as both a self-enhancing bias and a general cognitive tendency, then Hypotheses 2 and 3 will be supported.

## Experiment 1

### Method

**Participants.** Study participants were 60 police officers (21 women, 39 men) from Swedish urban and rural districts. They were approached through personal contact with representatives of the different districts, and they volunteered to participate in the experiment. Officers' ages varied between 25 years and 61 years ( $M = 41.84$ ,  $SD = 10.14$ ), and their length of experience as criminal investigators ranged from 1 year to 39 years ( $M = 14.37$ ,  $SD = 11.39$ ). The study (including all three experiments) was approved by the Regional Ethical Review Board in Uppsala, Sweden, before the commencement of data collection.

**Design.** The experiment had a 3 (decision maker: self vs. colleague vs. prosecutor)  $\times$  2 (decision: apprehension vs. no apprehension) within-subjects design. Note that because of the issue addressed—the effects of whether the choice is made by oneself or another person—the independent variable, *decision*, is

experimentally controlled only in the four “other-decision” cells of the design, whereas this variable is controlled by the participant's own decision in the two “self” cells. This means that although the scenarios are perfectly counterbalanced with a Latin Square in the four cells where the decision is made by someone else (i.e., a police colleague or a prosecutor), by necessity of the experimental design, the scenarios in the cells with decision by oneself may potentially be affected by self-selection effects (e.g., certain scenarios may become overrepresented in the category “apprehended” relative to “non apprehended” because of the decisions made by the participants). We address the potential problems that might arise from self-selection in the cells with decision by oneself with additional analyses in the Results section and the Appendix. Overall, the additional analyses provide no evidence of self-selection effects.

**Materials.** The materials used in the experiment consisted of 12 scenarios concerning 12 different crimes for which apprehension is permissible, provided that the other legal requirements are fulfilled. The scenarios were inspired by, but not identical to, real criminal cases because such scenarios probably would be taken more seriously than would completely fictitious cases. At the same time, the outcome of the case would be unknown to the participants so that their answers would not be influenced by the outcomes in the real cases. Because the 12 scenarios all varied with respect to the suspect's characteristics (gender, ethnicity, etc.), the nature of the alleged crime and the more specific course of events, it enabled testing of the hypotheses in a range of different and relatively realistic settings, minimizing the risk that the results were due to some unusual aspect of one specific scenario. In each scenario, we provided a one-page case summary with information about what happened (the alleged crime) and who was involved (the suspect, plaintiff, witnesses, etc.). The scenarios were deliberately described in a way that allowed different interpretations about whether the remaining legal requirements for apprehension were met (whether there was at least reasonable suspicion, special grounds for apprehension, etc.). The scenarios were also pretested to guarantee their ambiguity, that is, that they generated approximately the same numbers of decisions to apprehend and decisions to not apprehend.

With the aim of creating scenarios as representative of real crime suspicions as possible, statistics from [The National Council for Crime Prevention \(2014b\)](#) concerning the proportion of different kinds of crimes reported in Sweden during 2014 as well as statistics regarding the suspects' personal characteristics ([The National Council for Crime Prevention, 2014a](#)) were consulted. To reflect the proportions of the different crime categories as accurately as possible, five scenarios concerned crimes of stealing (41.67%); two concerned crimes against a person (16.67%); and one concerned fraud (8.33%), inflicting damage (8.33%), narcotic drug offenses (8.33%), traffic offenses (8.33%) and other crimes (8.33%) respectively. Furthermore, to mirror real suspects' personal characteristics, the suspect was a man in 10 (83.33%) and a woman in two (16.67%) of the 12 scenarios. For men, the most common crime category was narcotic drug offenses (21.00%) and for women it was theft (24.00%; [The National Council for Crime Prevention, 2014a](#)), and the scenarios were created accordingly. For some aspects of the scenarios, proper statistics were lacking and distributions concerning those aspects are therefore our own approximations of real life conditions. For instance, the suspect



had a criminal record in 6 out of 12 scenarios, and 6 of the suspects had names that indicated a completely or partially different ethnicity than Swedish whereas the remaining 6 suspects had Swedish-sounding names. To illustrate, Scenario 1 (for own decision to apprehend) is transcribed in full in the following text, and the remaining 11 scenarios are available in the [online supplemental material](#).

You and a colleague are called to an apartment after a woman, Linda Berggren, reported her partner, Mats Alvarsson, for assault. When you get to the apartment building you meet a man in the stairwell which, on your question, states that he is Mats Alvarsson. You ask him to go back into the apartment again, which he agrees with. In the apartment, your colleague and Linda goes inside the bedroom to talk about what happened and you are talking to Mats in the living room. Mats denies that he assaulted Linda. According to him, a verbal conflict has occurred between him and Linda when he was packing his things in an ongoing separation. After a rather heated exchange of words, Linda locked herself in the bathroom and called the police to report him falsely for assault. He became annoyed and tried to get Linda to withdraw her report, but when he realized that she would not do that, he became so angry that he felt he had to leave the apartment. That's when he met you in the stairwell. Mats finds it all very unfortunate. He does not know why Linda claims that he has been assaulting her but believes it has to do with their separation. Mats had an affair with another woman while he and Linda were still in a relationship and he thinks Linda does this as a kind of vengeance for his betrayal. He will go to his brother to live there until further notice.

Your colleague summarizes Linda's version of what happened. According to Linda, she and Mats began to argue about the reason for their separation and various economic disagreements. Mats owed her money and when she pointed that out, he went crazy. He pushed her hard into the bedroom wall, then into the bed and then down on the floor. She got hurt really bad, for instance in her right shoulder. After Mats left the room, Linda got up and brought her cell phone to the bathroom where she called the police. Linda does not have any marks or bruises from when Mats pushed her. Mats has previously been aggressive and violent to her on two occasions. On the first occasion, he hit her in the stomach so that she couldn't breathe and kicked her on her legs. On the second occasion, he threw her down on the floor and dragged her on the floor by pulling her hair. She had to call in sick to her employer for a few days after both occasions since she had difficulties moving and bruises on her legs. Linda also picked up her cell phone and showed pictures of her legs where there are big bruises on both shins.

The participants then responded to the questions in the following steps.

Step 1: Should Mats be apprehended? [*yes/no*]

Step 2: For what crime [free text], what level of suspicion [probable cause/reasonable suspicion] and what is the reason for the apprehension [risk of recidivism/removing evidence or hindering the investigation/fleeing]?

Step 3: According to The Code of Judicial Procedure 24 ch. 8 §, anyone apprehended shall be interrogated as soon as possible. Because you have chosen to apprehend Mats, you will prepare the interrogation in the next step (assuming that the decision was *yes* in Step 1).

Step 4: Formulate the six questions that you consider the most relevant to ask Mats during the interrogation. [free text]

Step 5: Choose six of the following questions that you consider most relevant to ask Mats during the interrogation (12 questions were available to choose from).

Step 6: When you've asked Mats all your questions and the interrogation is coming to an end, Mats denies having committed any crime. How trustworthy do you consider his denial? Answer by indicating a number from 1 to 7 where 1 means *not at all trustworthy* and 7 means *completely trustworthy*. On the scale from 1 to 7, I think the denial is. . . . (Seven choices were provided.)

Step 7: Scenario 1 out of 12 is completed. Click *Next* below to continue.

**Procedure.** Participants were instructed that they would get to read 12 scenarios and then make several judgments and decisions regarding the scenarios. The instructions also stated that the case materials had been inspired by authentic criminal cases but that the circumstances as well as the names and so forth had been changed to preserve confidentiality.

After reading the instructions, participants were presented with the 12 scenarios. All participants read four scenarios in each condition before the next condition was introduced. In order to avoid confounders, the order of the conditions as well as scenarios were systematically counterbalanced using a Latin square. For the conditions, there were in total 6 condition orders and for the scenarios there were in total 12 orders. As the scenarios were systematically counterbalanced across the four conditions with the colleague or prosecutor's decisions, each scenario appeared the same number of times in each condition ( $12 \times 5 = 60$  participants), to avoid that the conditions would systematically coincide with potentially biasing scenario information such as the suspect's gender, ethnicity or criminal record. Table S1 in the [online supplemental material](#) displays how the condition and scenario orders were counterbalanced across trials as well as how the scenarios were aggregated in the analysis.

In the condition for their own decisions, participants were asked, "Shall the suspect be apprehended?", and they could choose either *yes* or *no*. If they chose *yes*, they went through Steps 2 to 7. If participants instead chose *no*, they were informed that "According to The Code of Judicial Procedure 23 ch. 6 §, anyone who is reasonably likely to possess information relevant to the inquiry may be questioned. Although you have decided not to apprehend X, you find it appropriate to interrogate him/her. X is summoned to the interrogation and agrees to appear the next day. You will get to prepare the interrogation in the next step." Then, they went through Steps 4 to 7. In the conditions with a colleague's or a prosecutor's decision, the procedure was the same, except the participants did not make own decisions but were simply informed about the other person's decision. In both the colleague and prosecutor conditions, the decisions were preset so that in two of the scenarios the decisions were to apprehend, whereas the other two were to not apprehend. Just like for their own decisions, participants were informed about the legal ground for the interrogation; that is, The Code of Judicial Procedure 24 ch. 8 § in the appre-

hension condition and The Code of Judicial Procedure 23 ch. 6 § in the non apprehension condition. The order of the decisions was counterbalanced. In all conditions, participants freely self-generated six questions and then chose six questions from an Interrogation Questions Checklist (ICQ; Kassin et al., 2003) and assessed how trustworthy the suspect's confession or denial was.

#### Measures.

**Freely self-generated questions.** After having read each scenario, participants were asked to freely self-generate the six questions that they found most relevant to ask the suspect during the subsequent interrogation. The questions were rated by the first author and by four independent evaluators (i.e., law students) who were all blind to the condition in which the questions had been formulated; that is, they were unaware of what the decision was (apprehension or no apprehension) and who was the decision maker (participant, colleague, or prosecutor). Also, a senior researcher who is also the second author only rated 10% of the questions to assess whether the students and the senior researcher's ratings differed. The ratings were done on a scale, ranging from 1 to 7 where 1 = *completely innocence presumptive*, 4 = *neutral*, and 7 = *completely guilt presumptive*. *Presumptive of innocence* questions were defined as questions that assume that the suspect is innocent, and *presumptive of guilt* questions were defined as questions that assume that the suspect is guilty. *Neutral* questions were defined as questions that do not assume that the suspect is guilty or innocent. An example of a question that was considered presumptive of innocence was as follows: "Are you only confessing to protect your friends?" (in a case concerning inflicting damage where the suspect was the only one left at the crime scene from which his friends just left). An example of a neutral question was: "Can you please tell me what happened before the police arrived?" (in a case concerning assault against a partner in their home). A guilt presumptive question was as follows: "Why did you point a gun at the plaintiff?" (in a case regarding unlawful threat where the pointing of the gun was disputed by the suspect). The independent evaluators were informed that participants had formulated the questions with the purpose of finding out whether suspects were innocent or guilty, and they were also provided with the same scenario descriptions as the participants. For each question, all evaluators' ratings were averaged to produce a mean score. For each participant, the mean scores for all of their questions in each condition were averaged. This produced a total mean questions score that ranged from 0 to 7 in each condition.

The interrater reliability for the five raters (four independent raters and the first author) of the freely generated questions was  $\kappa = .68, p < .001, 95\% \text{ CI } [.66, .70]$ . The ratings of the four independent judges were also compared with the second author's ratings resulting in an interrater reliability of  $\kappa = .64, p < .001, 95\% \text{ CI } [.58, .69]$ .

**IQC.** Apart from freely generating their own questions, participants were also asked to choose six questions from a preset list of questions. A modified version of Kassin et al.'s (2003) IQC was used. The checklist contained 12 questions specifically tailored to each scenario and constructed as six pairs, with one question of each pair being guilt-presumptive, for example, "Why did you push Linda into the wall?", and one neutral, for example, "What happened in the apartment?" To check on the effectiveness of the pairings, two independent judges rated the degree to which each item was presumptive of guilt on a scale, ranging from 1 to 7 with

higher numbers indicating higher levels of guilt presumption. Both during evaluation of the questions and in the experiment, the questions were randomly ordered within a single list, not paired. Participants were asked to select the six questions from the list that they found most relevant to ask the suspect during the subsequent interrogation. Choosing a guilt-presumptive question gave 1 point and choosing a neutral question gave 0 points. Thus, depending on the nature of the questions chosen, every participant obtained a score between 0 and 6, where 0 represents all neutral questions and 6 represents all guilt-presumptive questions. Because the 12 scenarios concerned different crimes and suspects, a different checklist was used for each scenario.

As designed, the two independent judges' mean ratings were higher for the guilt presumptive questions ( $M = 5.84, SD = 1.02, 95\% \text{ CI } [4.60, 6.43]$ ) than for the more neutral alternatives ( $M = 2.69, SD = .53, 95\% \text{ CI } [3.17, 1.33]$ ),  $t(142) = -19.31, p < .001$ , and the size of this difference was large (Cohen's  $d = 3.88, 95\% \text{ CI } [3.12, 4.27]$ ). The interrater reliability for the two judges who rated all questions in the IQC was  $\kappa = .66, p < .001, 95\% \text{ CI } [.55, .77]$ . This suggests that the checklists did not limit participants in their responses to only one of the categories but instead allowed both guilt-presumptive and neutral choices.

**Trustworthiness ratings.** When participants had chosen questions from the checklist, they were informed that the suspect either denied or confessed to the crime. Participants were asked to rank how trustworthy the denial or confession was on a scale, ranging from 1 to 7 where 1 = *not at all trustworthy* and 7 = *extremely trustworthy*. In 10 out of 12 scenarios, the suspect denied, and in the remaining two scenarios, the suspect confessed. The scenarios in which the suspects denied and confessed were always the same. Because of lack of knowledge about the general proportions of denials and confessions in Swedish criminal cases, it is difficult to estimate to what extent this proportion is representative of real life criminal investigations. To avoid effects of base-rate expectations, we ensured that the proportion of denials and confessions was 10 to 2, in line with the real cases that inspired the scenarios.

**Social desirability.** After having completed all 12 scenarios, participants were asked to answer 14 questions from a short version of the Marlowe–Crowne Social Desirability Scale (M-C SDS; Crowne & Marlowe, 1960; Robinson, Shaver, & Wrightman, 1991). This was done to assess response bias, that is, the degree to which individuals attempt to present themselves in a favorable light. In this specific context, such a response bias could be at hand if participants realize what the purpose of the study is and try to present themselves as someone who does not ask suspects leading or guilt presumptive questions. Respondents answered true or false to the 14 items, resulting in a score ranging from 0 to 14, with higher scores reflecting a greater degree of socially desirable responding. The M-C SDS scores were correlated with the dependent measures.

There were no significant correlations between social desirability and the level of guilt presumption in the freely generated questions ( $r = .17, p = .248$ ), the questions chosen from the IQC ( $r = .16, p = .297$ ), or trustworthiness ratings ( $r = -.24, p = .114$ ). This indicates that participants' responses were not significantly influenced by response bias.

**Assumed purpose of the study.** Finally, participants were also asked to state what they thought was the purpose of the study. Participants' answers were classified into four groups based on

their content. Most participants believed that the study was about evaluating objectivity and different kinds of biases (51.06%) or evaluating interrogation techniques (36.17%). A few police officers stated that they had no idea about the purpose of the study (6.38%) and the remaining police officers stated another purpose (such as telling the difference between truths and lies or participants' patience with carrying out the survey, where only one participant stated each purpose, 6.38%). The assumed purpose of the study did not significantly influence the questions that participants freely generated,  $F(18, 120) = 0.75, p = .768$ , chose from the IQC,  $F(18, 120) = 1.36, p = .165$  or trustworthiness ratings,  $F(18, 120) = 1.061, p = .392$ . This suggests that even when police officers were aware and prepared that the prevalence of bias was being studied, that did not significantly reduce confirmatory tendencies, emphasizing the more or less subconscious nature of confirmation bias.

## Results

Table 1 displays descriptive statistics (means and standard deviations) and post hoc comparisons for significant effects found for the main dependent variables.

**Background variables.** Exploratory analyses were conducted to examine whether participants' background characteristics, that is, gender, age and years of experience, were related to any of the dependent variables.

Looking first at participants' gender, there were no significant differences between the proportion of guilt presumptive questions that male ( $M = 4.82, SD = .81, 95\% CI [3.73, 5.55]$ ), and female participants freely generated ( $M = 4.92, SD = .34, 95\% CI [3.97,$

$5.03]$ ),  $t(58) = -.74, p = .462$ , or chose from the IQC ( $M = 2.25, SD = .23, 95\% CI [2.01, 3.20]$ ) and ( $M = 2.20, SD = .34, 95\% CI [1.97, 3.12]$ ),  $t(58) = .21, p = .832$ . Male and female participants did not differ significantly in their trustworthiness ratings ( $M = 3.01, SD = 1.13, 95\% CI [1.98, 4.30]$ ) and ( $M = 2.98, SD = 1.43, 95\% CI [2.21, 4.74]$ ),  $t(58) = .18, p = .860$ .

Furthermore, participants' ages were not significantly correlated with the level of guilt presumption in the freely generated questions,  $r = .21, p = .109$ , the choices of questions from the IQC ( $r = .077, p = .553$ ) or the trustworthiness ratings ( $r = -.065, p = .621$ ). Similarly, the number of years of experience was not significantly related to the level of guilt presumption in the freely generated questions ( $r = .23, p = .070$ ), the choices of questions from the IQC ( $r = .095, p = .464$ ), or the trustworthiness ratings ( $r = .025, p = .847$ ).

**Freely generated questions.** With a 3 (decision maker: self vs. colleague vs. prosecutor)  $\times$  2 (decision: apprehension vs. no apprehension) repeated-measures analysis of variance (ANOVA), a main effect of decision on the type of questions was found,  $F(1, 59) = 31.35, p < .001$ . Suspects who had previously been apprehended were asked significantly more guilt presumptive questions than suspects who had not been apprehended and the size of this effect was  $r = .59$ .

There was no significant effect of decision maker on the questions that participants generated,  $F(2, 58) = 1.29, p = .281$ . However, there was a significant interaction effect between the decision and the decision maker,  $F(2, 58) = 3.80, p = .025, r = .25$ . Post hoc tests between all six cells in the design using Bonferroni revealed significant differences in guilt presumption

Table 1  
Means, Standard Deviations, and Post Hoc Comparisons for Significant Effects by Dependent Variable (DV), Decision, and Decision Maker in Experiment 1

DV/Decision maker	Decision				Total		Post hoc
	Apprehension		No apprehension		M	SD	
	M	SD	M	SD	M	SD	
Freely self-generated (FSG) questions							$p < .001$ for all A vs. all NA
Self (S)	5.30	.82	4.22	1.06	4.76	.76	
Colleague (C)	5.17	.70	4.47	.80	4.82	.49	
Prosecutor (P)	5.22	.67	4.57	.73	4.90	.46	
Total	5.23	.066	4.42	.18	4.83	.46	
Questioning mode							
FSG							$p < .001$ for all FSG vs. all IQC
S	4.30	.82	3.22	1.06			$p < .001$ FSG: SA/SNA
C	4.17	.70	3.47	.80			$p < .001$ FSG: CA/CNA
P	4.22	.67	3.57	.73			$p < .001$ FSG: PA/PNA
Total	4.23	.073	3.42	.86			$p < .001$ FSG: SA/SNA
IQC							
S	2.15	.93	2.33	1.12			
C	2.09	.80	2.02	1.02			
P	2.23	1.01	2.00	.88			
Total	2.16	.91	2.12	1.00			
Trustworthiness rating							
S	2.83	.92	3.66	1.23	3.25	.59	$p = .003$ for S/A vs. S/NA
C	3.22	1.09	3.30	1.33	3.27	.071	
P	3.22	1.00	2.96	1.15	3.09	.18	
Total	3.09	.23	3.32	.35	3.20	.29	$p = .022$ for S/NA vs. P/NA

Note. IQC = Interrogation Questions Checklist.

between all apprehension conditions and all no apprehension conditions ( $p < .001$ ; for all differences). Also, there was a nearly significant difference in guilt presumption when police officers themselves had decided to not apprehend and when prosecutors had decided to not apprehend ( $p = .062$ ). When police officers themselves had decided to not apprehend the suspect, their questions were somewhat less guilt presumptive compared to when the prosecutor had decided to not apprehend.

The significant interaction that is illustrated in Figure 1 thus amounts to the finding that suspects who have been apprehended are asked more guilt presumptive questions, and this effect appears to be somewhat larger when the police officers made the decision themselves (although this effect fails to reach statistical significance in the subsequent more conservative Bonferroni post hoc tests).

**IQC.** With a 3 (decision maker: self vs. colleague vs. prosecutor)  $\times$  2 (decision: apprehension vs. no apprehension) repeated-measures ANOVA no significant main effect of decision on the type of questions participants chose from the IQC was found,  $F(1, 59) = 0.39, p = .531$ . Neither was there a significant main effect of decision maker on the questions that participants choose from the IQC,  $F(2, 58) = 1.51, p = .231$ , or any significant interaction effect,  $F(2, 58) = 2.30, p = .110$ . Hence, whereas the findings clearly point to a confirmation bias in the freely generated questions, there are no such indications for the questions that participants chose from the IQC. Note that for the independent judges' ratings the means of 5.84 and 5.12 for the questions in the IQC that were intended to be guilt presumptive lies relatively close to the means for the self-generated questions in all conditions demonstrated in Figure 1, indicating a relatively high guilt presumption in the freely generated questions overall.

**Comparing questioning modes.** Since guilt presumption was initially indicated on different scales, ranging from 1 to 7 for the freely generated questions and 0 to 6 for the IQC, all ratings of the

freely generated questions were subtracted with 1 to allow relevant comparisons of the questioning modes. The effect on guilt presumption of freely generated questions versus IQC is difficult to estimate rigorously with the current design, because questioning mode is confounded with order (the freely generated questions were always produced before the IQC). With this caveat in mind, we nonetheless tested if the difference in guilt presumption was significantly higher with freely generated questions ( $M = 3.83, SD = .91, 95\% \text{ CI } [3.72, 3.92]$ ) as compared with the IQC ( $M = 2.14, SD = .97, 95\% \text{ CI } [2.03, 2.24]$ ), which proved to be the case,  $t(59) = -23.83, p < .001$ , and the size of this difference was large (Cohen's  $d = 1.80, 95\% \text{ CI } [1.54, 2.43]$ ).

Figure 2 displays the level of guilt presumption for freely generated questions and the IQC for all levels of the decision (apprehension vs. no apprehension) and decision maker (self vs. colleague vs. prosecutor) variables.

**Trustworthiness ratings.** With a 3 (decision maker: self vs. colleague vs. prosecutor)  $\times$  2 (decision: apprehension vs. no apprehension) repeated-measures ANOVA, a main effect of decision on participants' trustworthiness ratings was found,  $F(1, 59) = 5.061, p = .028$ . Suspects who had previously been apprehended were considered significantly less trustworthy than suspects who had not been apprehended and the size of this effect was  $r = .28$ . There was no significant effect of decision maker on the trustworthiness ratings,  $F(2, 58) = 1.26, p = .287$ . However, there was a significant interaction effect between the decision and the decision maker,  $F(2, 58) = 6.37, p = .002, r = .31$ . Post hoc tests using Bonferroni revealed significant differences in trustworthiness ratings when police officers themselves had apprehended and not apprehended ( $p = .003$ ). Furthermore, there were significant differences in trustworthiness ratings between police officers when they had decided to not apprehend a suspect and when the prosecutor had decided to not apprehend the suspect ( $p = .022$ ). Figure 3 shows that apprehended suspects were rated as less trustworthy

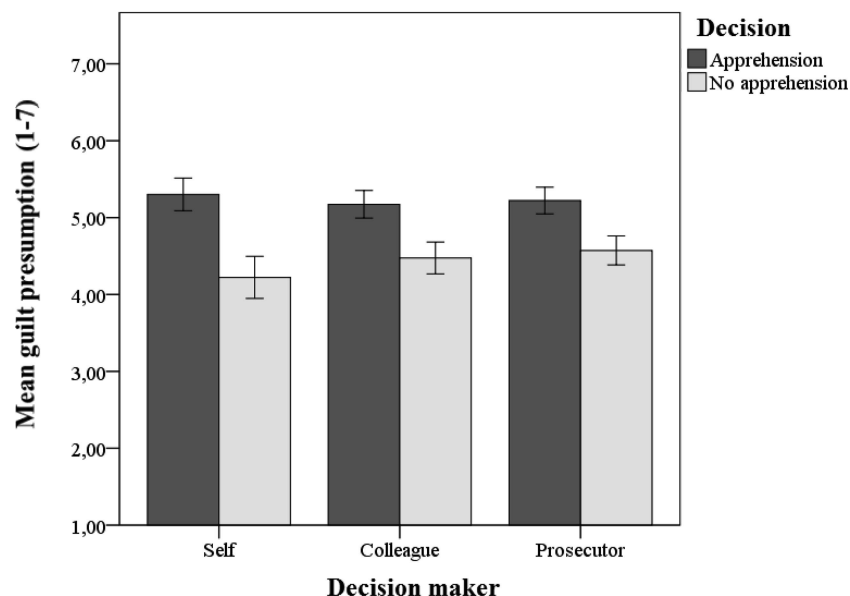


Figure 1. Mean guilt presumption in freely generated questions by decision and decision maker in Experiment 1. Error bars represent 95% CI.



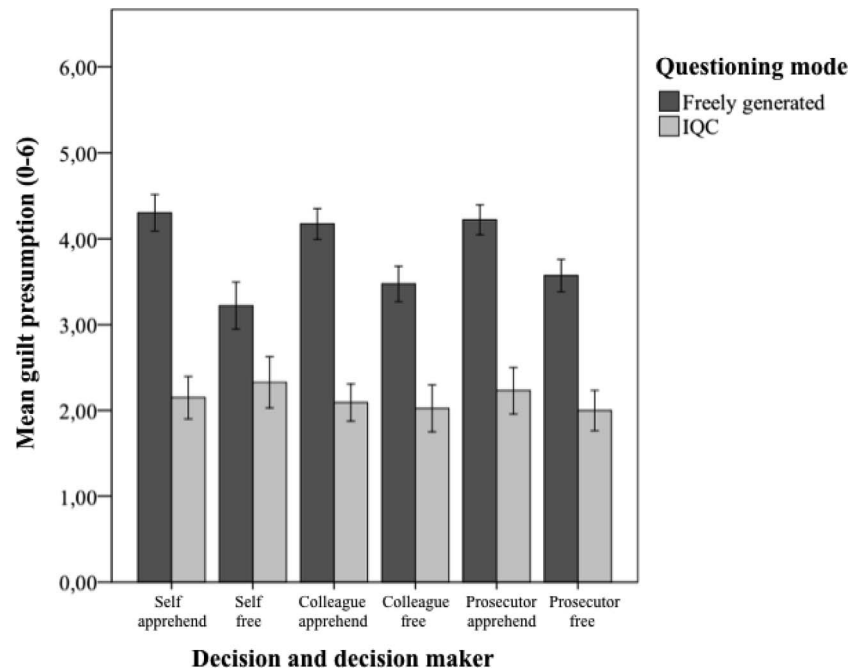


Figure 2. Mean guilt presumption by questioning mode, decision and decision maker in Experiment 1. Error bars represent 95% CI.

than not apprehended suspects (with exception for the prosecutor condition) and that the effect was qualified by an interaction suggesting that the effect was larger when police officers themselves had decided about apprehension.

**Controlling for potential self-selection of scenarios.** As noted in the Design section, the independent variable, *decision*, was experimentally controlled only in the four other (other police officer or prosecutor) decision cells of the design, whereas this variable is controlled by the participant's own decision in the two "self" cells. This opens for the possibility that the effects observed in these cells are not primarily driven by the decision to apprehend or not as such, but by a differential self-selection of scenarios that come to fall in these categories as a function of the participant's own decisions. For example, if some of the scenarios inherently elicit more guilt presumption, these scenarios may become over-represented in the apprehended decision category because of the decisions made by the participants (in the other cells this assignment is experimentally controlled). A different pattern in the condition where the decision maker is the police officer him- or herself might thus in principle arise because of a different and not counterbalanced set of scenarios in the categories apprehended versus not apprehended.

To evaluate selection effects, the distribution of scenarios across the decision categories was examined. There was a significant difference between the distribution of scenarios to the apprehend and no apprehend conditions,  $\chi^2(11) = 26.27, p = .006$ , when participants themselves decided. However, this difference does not appear to be the result of intrinsic differences in the scenarios' ability to generate guilt presumption. We computed a measure of each scenario's propensity to elicit guilt presumptive questions as the mean of the guilt assessments for this scenario across the four

cells where the scenario presentation was controlled and perfectly counterbalanced according to a Latin Square (i.e., thus excluding the data from the own decision situation). The scenarios in which subjects more often decided to apprehend ( $M = 4.35, SD = 1.12, 95\% \text{ CI } [3.78, 4.65]$ ) did not display a higher mean guilt presumption than the mean guilt presumption of the scenarios were subjects more often chose to not apprehend ( $M = 4.38, SD = 1.23, 95\% \text{ CI } [3.92, 4.83]$ ),  $t(50) = .795, p = .430$ . In the [Appendix](#), we report further extensive analyses involving stepwise-regression to demonstrate that the observed difference in guilt presumption between apprehended and not apprehended in the police him- or herself condition cannot be accounted for by a systematic difference in the distribution of scenarios assigned by the police to these two categories.

## Discussion

In Experiment 1, police officers demonstrated a higher guilt presumption in their freely generated questions in relation to apprehended than to non apprehended suspects. Also, the level of guilt presumption in the freely generated questions varied significantly across the decision makers, although this difference failed to reach statistical significance with the more conservative post hoc tests. These findings were largely in line with Hypotheses 1 and 2 and provide some support for the notion of confirmation bias as a self-enhancing bias. These hypotheses are further evaluated with a student sample in Experiment 2. An unexpected finding in Experiment 1 was that neither the decision nor the decision maker had any significant effects on the questions that participants choose from the IQC. Also, when comparing the questioning modes directly, a significantly lower

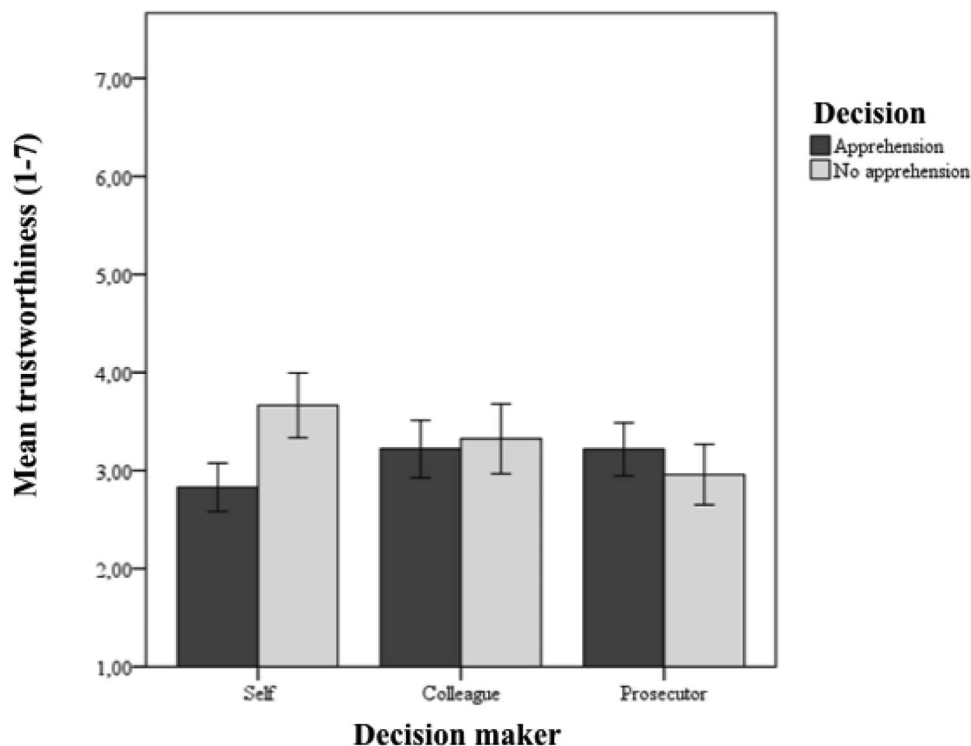


Figure 3. Mean trustworthiness ratings by decision and decision maker in Experiment 1. Error bars represent 95% CI.

level of guilt presumption was found when participants chose questions from the IQC as compared with when they generated the questions themselves. Although these results were obtained using a repeated measures design where the presentation order of questioning mode was not counterbalanced, they point to the possibility that the questioning mode influences the level of guilt presumption.

More specifically, it indicates that the level of guilt presumption in freely generated questions is influenced by the decision (and possibly the decision maker variable as well although not significantly in this study) whereas the level of guilt presumption in the questions chosen from the IQC is not. If guilt presumption, as expressed by the questions chosen from the IQC, does not vary as a result of that the suspect is apprehended or who has made that decision, then using the IQC could be a potential debiasing technique. This notion was tested systematically in Experiment 3 which directly compared the level of guilt presumption displayed by participants who either freely generated their own questions or chose questions from the IQC in a between-subjects design.

Given the current emphasis in the behavioral sciences on independent replication (for a review see Lilienfeld & Waldman, 2017), we decided to replicate Experiment 1 with an independent sample of participants. Because of the presumably universal cognitive mechanisms underlying confirmation bias, and because of the complication and special cost involved in using police officers, the replication was performed with law and psychology students.

## Experiment 2

### Method

**Participants.** Study participants were 60 students, 30 law and 30 psychology students (28 women, 32 men) from the Faculty of Law and the Department of Psychology, Uppsala University. Their mean age was 24.73 years ( $SD = 4.70$ ). They replied to study announcements that were displayed on the Faculty/Department's premises.

**Design.** The experiment had a 3 (decision maker: self vs. colleague vs. prosecutor)  $\times$  2 (decision: apprehension vs. no apprehension) within-subjects design.

**Materials.** The materials were the same as in Experiment 1.

**Procedure.** The procedure was the same as in Experiment 1 but with one exception, namely that participants generated three instead of six freely generated questions, in order to reduce the time required for participation. The freely generated questions were rated by two independent evaluators on a scale, ranging from 1 to 3 where 1 = *innocence presumptive*, 2 = *neutral*, and 3 = *guilt presumptive*. The degree of conformity in the evaluators' ratings was assessed for each question and the total interrater reliability for the two raters was  $1582/1970 = .803$  ( $\kappa = 1$ ).

### Results

Table 2 displays descriptive statistics (means and standard deviations) and post hoc comparisons for significant effects found for the main dependent variables.

Table 2

Means, Standard Deviations, and Post Hoc Comparisons for Significant Effects by Dependent Variable (DV), Decision, and Decision Maker in Experiment 2

DV/Decision maker	Decision						Post hoc	
	Apprehension		No apprehension		Total			
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Freely self-generated (FSG) questions								
Self (S)	2.53	.76	1.33	.56	1.93	.85	$p < .001$ for all A vs. all NA	
Colleague (C)	2.31	.88	1.55	.68	1.93	.54		
Prosecutor (P)	2.44	.78	1.50	.63	1.97	.66		
Total	2.43	.11	1.46	.12	1.94	.54		
Questioning mode								
FSG							$p < .001$ for all A/FSG vs. all A/IQC	
S	4.59	.64	.99	.27				$p < .001$ FSG:SA/SNA
C	3.93	1.32	1.65	.64				$p < .001$ FSG:CA/CNA
P	4.32	.82	1.50	.37			$p < .001$ FSG:PA/PNA	
Total	4.28	.93	1.38	.43				
IQC								
S	2.27	1.14	2.20	1.24				
C	2.36	1.12	2.31	1.13				
P	2.19	1.05	2.03	1.05				
Total	2.27	1.10	2.18	1.14				
Trustworthiness rating								
S	2.76	1.29	3.78	1.04	3.27	.72	$p = .016$ for S/A vs. S/NA	
C	3.26	1.39	3.52	1.44	3.39	.18		
P	2.90	1.37	3.39	1.52	3.15	.35	$p = .029$ for S/A vs. P/NA	
Total	2.97	.26	3.56	.20	3.27	.38	$p = .026$ for S/NA vs. P/A	

Note. IQC = Interrogation Questions Checklist.

**Freely generated questions.** A 3 (decision maker: self vs. colleague vs. prosecutor)  $\times$  2 (decision: apprehension vs. no apprehension) repeated-measures ANOVA was conducted.

There was a significant main effect of decision,  $F(1, 44) = 106.30$ ,  $p < .001$ , and the size of this effect was  $r = .84$  on the level of guilt presumption in the freely generated questions. The decision maker variable did not have a significant effect,  $F(2, 43) = 0.088$ ,  $p = .916$ . However, there was a significant interaction effect between decision and decision maker,  $F(2, 43) = 5.14$ ,  $p = .010$ ,  $r = .33$ . Post hoc tests using Bonferroni revealed significant differences in guilt presumption between all apprehension conditions and all no apprehension conditions ( $p < .001$ ; for all differences). These effects are illustrated in Figure 4. Also, there was a nearly significant difference in guilt presumption when the participants themselves had decided to apprehend and when colleagues had decided to apprehend ( $p = .088$ ). When the participants themselves had decided to apprehend the suspect, their questions were somewhat more guilt presumptive compared to when the colleague had decided to apprehend.

Note that the independent judges' mean ratings of the guilt presumptive questions in the IQC ( $M = 5.84$  and  $M = 5.12$ ) lies relatively close to the means for the self-generated questions in the apprehension conditions when these are transformed into the same scale.

Because the freely generated questions in Experiment 2 were rated on a scale ranging from 1 to 3 all ratings were transformed to allow relevant comparisons with the ratings of the IQC questions (1–7). The following transformation was used:  $Y_{\text{new}} = (Y_{\text{old}} - 1) \times 3 + 1$ , where 1 maps to 1 and 3 maps to 7. Using

this transformation, the overall mean for the freely generated questions in the apprehension conditions in Experiment 2 is  $(2.43 - 1) \times 3 + 1 = 5.29$  whereas the overall mean for the no apprehension conditions is  $(1.46 - 1) \times 3 + 1 = 2.38$ . Thus, mean guilt presumption in the apprehension conditions comes close to the means 5.84 and 5.12 for the IQC. This implies that the overall level of guilt presumption in the freely generated questions was quite high in the apprehension conditions.

**IQC.** A 3 (decision maker: self vs. colleague vs. prosecutor)  $\times$  2 (decision: apprehension vs. no apprehension) repeated-measures ANOVA was conducted. No significant main effects were found for decision,  $F(1, 59) = 1.99$ ,  $p = .165$ , or decision maker,  $F(2, 58) = 1.34$ ,  $p = .271$ , on the questions chosen from the IQC. Neither were there any significant interaction effects between decision and decision maker,  $F(2, 58) = .56$ ,  $p = .576$ .

**Comparing questioning modes.** Because guilt presumption was originally measured on different scales, ranging from 1 to 3 for the freely generated questions and 0 to 6 for the IQC, the responses were transformed to a unitary scale to allow relevant comparisons. This was done using the following transformation for all freely generated questions:  $Y_{\text{new}} = (Y_{\text{old}} - 1) \times 3$ , where 1 maps to 0 and 3 maps to 6.

Because Experiment 2 uses the same design as Experiment 1, and questioning mode therefore is confounded with order (the freely self-generated questions were always produced before the IQC), the effect of questioning mode on guilt presumption cannot be rigorously estimated in Experiment 2 either. Nonetheless, we tested whether the difference in guilt presumption was significantly higher with freely self-generated questions  $M = 2.83$ ,  $SD = 1.36$ , 95% CI [2.52, 3.17] as compared with the IQC ( $M = 2.23$ ,

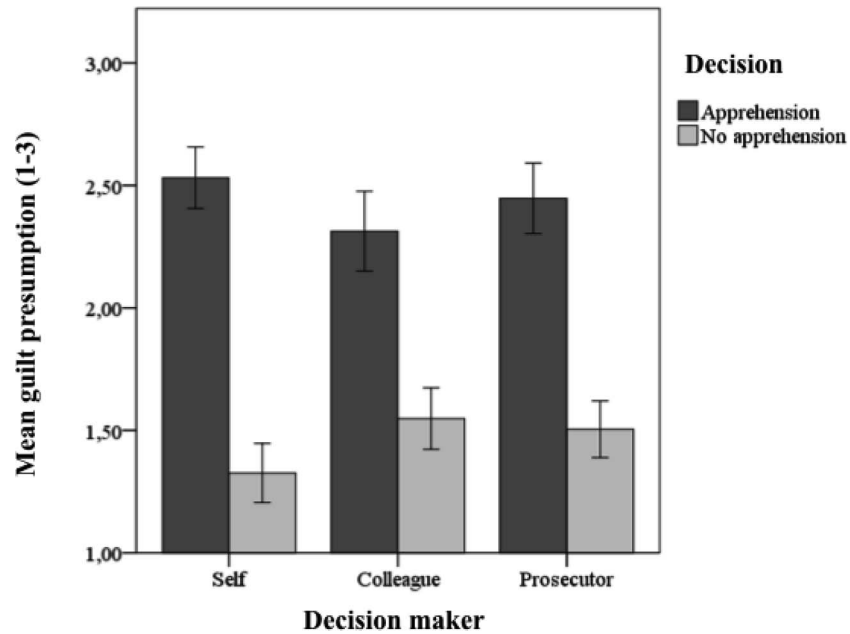


Figure 4. Mean guilt presumption in freely generated questions by decision and decision maker in Experiment 2. Error bars represent 95% CI.

$SD = 1.12$ , 95% CI [2.11, 2.36]), which proved to be the case,  $t(59) = 4.29$ ,  $p = .008$ , and the size of this difference was medium (Cohen's  $d = 0.48$ , 95% CI [0.32, 0.94]). Figure 5 displays the level of guilt presumption for freely self-generated questions and the IQC for all levels of the decision (apprehension vs. no

apprehension) and decision maker (self vs. colleague vs. prosecutor) variables. In Experiment 3, we used a more rigorous test of the effect of questioning mode in a direct between-subjects design that eliminates the confound with order in the current design.

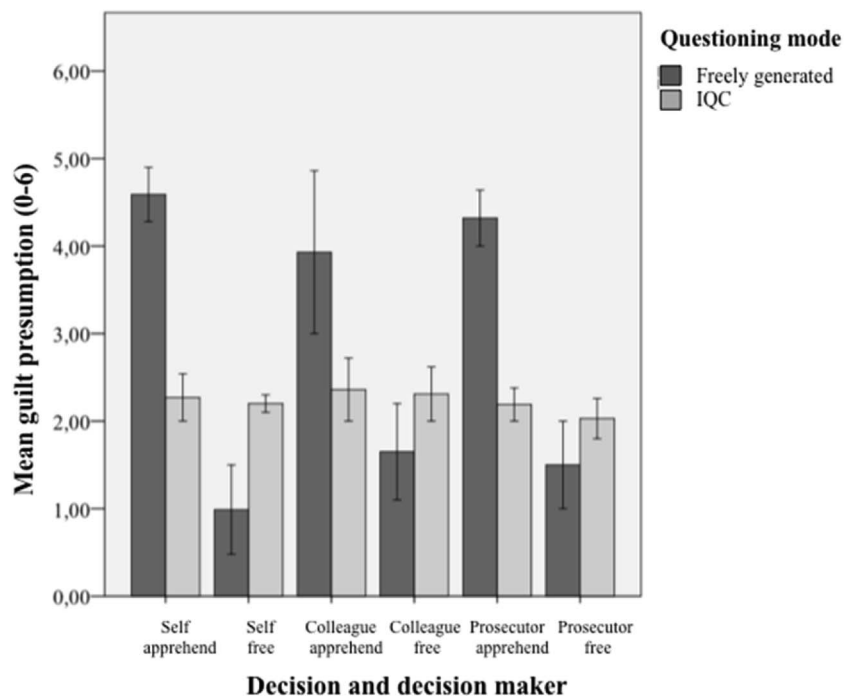


Figure 5. Mean guilt presumption by questioning mode, decision and decision maker in Experiment 2. Error bars represent 95% CI.



**Trustworthiness ratings.** A 3 (decision maker: self vs. colleague vs. prosecutor)  $\times$  2 (decision: apprehension vs. no apprehension) repeated-measures ANOVA was conducted. A significant main effect was found for decision,  $F(1, 46) = 23.14, p < .001$ , and the size of this effect was  $r = .58$ . Apprehended suspects were rated as less trustworthy than non apprehended suspects. There was also a significant effect of decision maker,  $F(2, 45) = 3.81, p = .029, r = .28$ , on the ratings of trustworthiness. The results of the Bonferroni post hoc tests are summarized in Table 3. As illustrated in Figure 6, and replicating the results of Experiment 1, there was a lower mean assessed trustworthiness after own decision to apprehend compared with own decisions to not apprehend ( $p = .010$ ; using Bonferroni post hoc tests).

## Discussion

The results in Experiment 2 largely replicate the results in Experiment 1. Both experiments in part support Hypothesis 1, which postulates that apprehensions trigger a confirmation bias as manifested by more guilt-presumptive freely self-generated questions ( $r = .59$  in Experiment 1 and  $r = .84$  in Experiment 2) and lower trustworthiness ratings ( $r = .28$  in Experiment 1 and  $r = .58$  in Experiment 2) in relation to apprehended versus non apprehended suspects. However, both experiments also contradict Hypothesis 1 because the questions chosen from the IQC were not significantly influenced by the apprehension.

Furthermore, in both experiments, overall effects of the decision maker variable on the freely self-generated questions were found (Hypothesis 2), but when using the more conservative post hoc tests, these differences failed to reach statistical significance in both experiments. Thus, the level of guilt presumption only differed marginally depending on who was the decision maker with a slightly higher level of decision-consistent questions in relation to own decisions than to the prosecutor's ( $p = .062$  in Experiment 1) or colleague's ( $p = .088$  in Experiment 2) decisions. For the trustworthiness ratings, the decision-maker variable had significant influences as participants perceived the suspects they themselves had apprehended as less trustworthy than non apprehended suspects, ( $r = .31$  in Experiment 1 and  $r = .28$  in Experiment 2). Such differences did not appear in relation to the colleague's or prosecutor's decisions. Thus in both Experiments 1 and 2, participants' displayed decision-consistent trustworthiness ratings when the decisions were their own. A small difference appeared in response to the prosecutor's decision because police officers perceived the suspects they themselves had decided to not apprehend as more

trustworthy than the suspects the prosecutor had decided to not apprehend ( $r = .22$ ; Experiment 1), whereas students trustworthiness ratings were similar both when they had made the decision and when the prosecutor had made the decision ( $p = .027$  for self to apprehend vs. prosecutor to not apprehend, and  $p = .026$  for self not to apprehend vs. prosecutor to apprehend). This could indicate a greater skepticism toward the prosecutor in the police sample than in the student sample but since this has not been tested systematically in this study, no reliable conclusions can be drawn. Overall, both experiments support the notion of apprehension as a trigger of confirmation bias but the potential of changing decision maker as a debiasing technique is more uncertain since the effect of the decision maker variable was not significant for all measures.

## Experiment 3

The results in both Experiment 1 and 2 suggest little or no confirmation bias when participants chose questions from a preset list (the IQC) compared with when they formulated the questions themselves. Because choosing questions from a preset list is a cognitively less demanding task than freely formulating own questions, this indicates that reducing cognitive load could reduce confirmation bias. However, in Experiment 1 and 2, participants always chose questions after having generated their own questions, which points to a potential confound between order and question format. To test whether the differences remained when participants only used one of the question formats in otherwise identical circumstances, Experiment 3 had a between-subjects design in which participants either chose or formulated questions.

## Method

**Participants.** Study participants were 60 students (39 women, 21 men) from the Department of Psychology, Uppsala University, with a mean age of 26.19 years ( $SD = 7.02$ ). They replied to study announcements that were displayed on the Department's premises.

**Design.** The experiment had a 2 (questioning mode: production and detection vs. detection)  $\times$  2 (decision: apprehension vs. no apprehension)  $\times$  3 (decision maker: self vs. colleague vs. prosecutor) mixed-subjects design. Questioning mode was varied between the groups, whereas decision and decision maker were within-subjects factors.

**Materials.** The materials were the same as in Experiment 1.

**Procedure.** The procedure used for Experiment 3 was the same as in Experiment 1, but with one exception. In preparation of

Table 3

Means, Standard Deviations, and Post Hoc Comparisons for Significant Effects on Guilt Presumption by Questioning Mode, Decision, and Decision Maker in Experiment 3

Decision maker	Apprehension				No apprehension				Total				Post hoc
	Production and detection		Detection		Production and detection		Detection		Production and detection		Detection		
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Self	4.53	.98	2.29	1.16	3.63	.96	2.20	.74	4.08	.64	2.25	.65	$p = .010$ SA vs. SNA
Colleague	4.46	.88	2.61	1.03	3.72	.98	2.34	.85	4.09	.52	2.48	.19	
Prosecutor	4.34	.85	2.30	.91	3.75	.83	2.31	.89	4.05	.42	2.31	.47	
Total	4.69	1.01	2.42	1.22	3.56	1.18	2.36	.99	4.07	.14	2.34	.41	

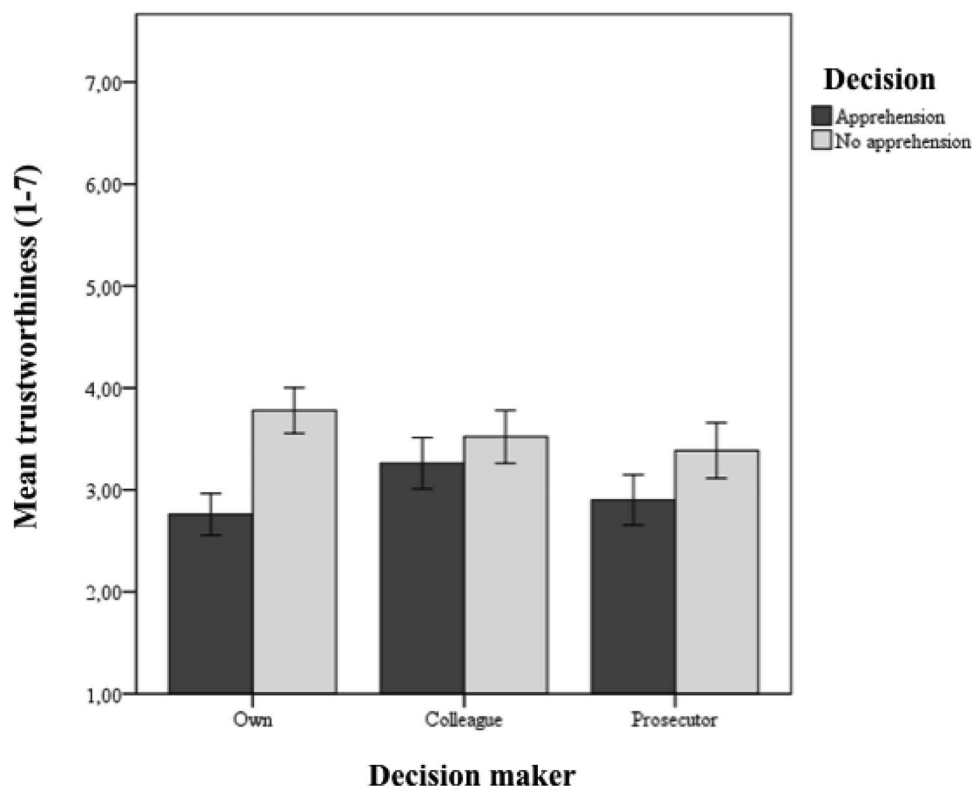


Figure 6. Mean trustworthiness ratings by decision and decision maker in Experiment 2. Error bars represent 95% CI.

the suspect interrogations, half of the participants freely self-generated six questions, whereas the other half chose six questions from the IQC. The freely self-generated questions were rated by the first author and two independent evaluators, and total mean scores from 0 to 6 were calculated for each condition. For the questions chosen from the IQC, mean scores from 0 to 6 were calculated for each condition. The interrater reliability for the three raters of the freely self-generated questions was  $\kappa = .71$ ,  $p < .001$ , 95% CI [.69, .99].

**Manipulation check.** A manipulation check for cognitive load was carried out by comparing the total time to complete the task for participants using the IQC and for those who freely self-generated questions. As expected provided that the IQC is a cognitively less demanding task, the total time to complete the task was significantly shorter in the IQC condition ( $M = 45$  min,  $SD = 24$  min, 95% CI [37 min, 54 min]) compared with in the freely self-generated condition ( $M = 1$  hr 49 min,  $SD = 38$  min, 95% CI [1 hr 35 min, 2 hr 3 min]),  $t = 7.74$ ,  $p < .001$ . The size of this difference was large (Cohen's  $d = 2.00$ , 95% CI [1.12, 3.71]).

## Results

Table 3 displays descriptive statistics (means and standard deviations) and post hoc comparisons for significant effects on the level of guilt presumption in interrogation questions.

**Questioning mode.** A  $2 \times 2 \times 3$  mixed ANOVA with questioning mode (production and detection vs. detection) as a between-subjects factor and decision (apprehension vs. no appre-

hension) as well as decision maker (self vs. colleague vs. prosecutor) as within-subject factors was conducted. There was a significant main effect of questioning mode on the level of guilt presumption in interrogation questions indicating a higher level of guilt presumption among participants who freely generated questions than among participants who chose questions from the IQC,  $F(1, 46) = 130.40$ ,  $p < .001$ . The size of this effect was  $r = .86$ .

There was also a significant interaction effect between the questioning mode and the decision,  $F(1, 46) = 5.24$ ,  $p = .027$ . Post hoc tests using Bonferroni revealed a significantly stronger effect after an apprehension as compared to non apprehension ( $p = .010$ ,  $r = .32$ ). Figure 7 shows the large main effect between the production and detection condition and the detection condition with much less guilt presumption in the detection condition, qualified by an interaction suggesting that for production and detection, there was higher guilt presumption in the apprehension than in the no apprehension condition, whereas for detection, the level of guilt presumption was approximately the same in the apprehension and no apprehension condition. There was no significant effect of decision maker,  $F(2, 92) = .72$ ,  $p = .491$ .

## General Discussion

The aim of the present experiments was to test whether decisions to apprehend a suspect triggers confirmation bias expressed by the questioning style in the subsequent suspect interrogation. Furthermore, the aim was to examine whether any such effects

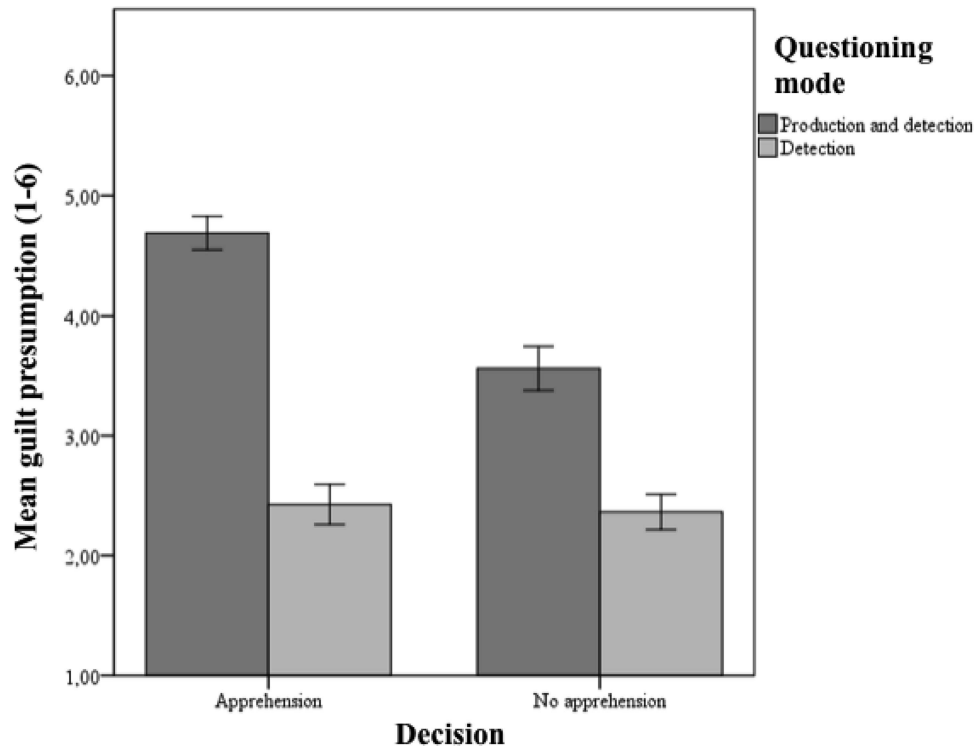


Figure 7. Mean guilt presumption by questioning mode and decision in Experiment 3. Error bars represent 95% CI.

would be moderated by who made the decision about apprehension as well as by reducing cognitive load through the questioning mode. In line with prior research, it was expected that suspects who had been apprehended would be interrogated in a more guilt presumptive manner than non apprehended suspects. These differences were expected to be larger when police officers themselves had decided about apprehension compared to when a colleague or a prosecutor had decided. Also, following the results of Experiments 1 and 2, it was hypothesized that guilt presumption would decrease with a cognitively less demanding questioning mode, which was tested in Experiment 3.

The results largely support the hypothesis that apprehensions trigger confirmation bias but the potential of changing decision maker is more uncertain provided that the differences between the decision makers were (close to) but not significant ( $p = .062$  in Experiment 1 and  $p = .088$  in Experiment 2). However, the results indicate a greater potential of reducing cognitive load as a debiasing technique, since this effect was significant and the effect size was large ( $r = .86$  in Experiment 3). Police officers' freely generated questions to apprehended suspects were more guilt presumptive and these findings were replicated with a larger effect size in the student sample in Experiment 2. Similarly, police officers' perceived that apprehended suspects were less trustworthy, which was also found in Experiment 2. Furthermore, in both experiments, there were significant differences in trustworthiness ratings between own decisions to apprehend and not apprehend, whereas no such differences were found for the other decision makers. However, police officers displayed skepticism toward the prosecutor's decision that was not present in the student sample.

Apart from this, the findings were consistent in both experiments, that is, the apprehension was associated with a more guilt presumptive questioning style and for own decisions, both police officers and students perceived of apprehended suspects as less trustworthy than non apprehended suspects.

By contrast, neither in Experiment 1 nor Experiment 2, did the decision or the decision maker significantly influence the questions participants chose from the IQC. This unexpected finding motivated Experiment 3 which compared the level of guilt presumption displayed by participants who either freely generated their own questions (production and detection) or chose questions from the IQC (detection). The results point to a generally lower level of guilt presumption when participants chose questions from the IQC and this effect was large (and emphasized by the null findings in Experiment 1 and 2 with two different populations).

As indicated by the independent judges' ratings of the questions in the IQC, the checklists contained both guilt presumptive and neutral questions, suggesting that the lower level of guilt presumption for the questions chosen from the IQC is unlikely to be due to restriction of range. Also, the effect was stronger in the apprehension condition than in the non apprehension condition. This indicates that the difference in levels of guilt presumption, as a result of the reduced cognitive load, is largest when the suspect is apprehended. As regards the tested debiasing techniques, the significant effect and large effect size ( $r = .86$ ), for reducing cognitive load (by changing questioning mode) suggests that reducing cognitive load has better potential as a debiasing technique than

what changing decision maker does, since these effects were non-significant.

### Limitations

The major reason for employing experimental methods when studying this topic was that it enabled control of the variables of interest as well as separation between them and other variables unrelated to the study. Since decisions to apprehend suspects and who has made the decision are variables naturally entangled with lots of other variables that are typically at play in criminal investigations, studying real cases would have made reliable measures of the variables' effects as well as effect sizes impossible. Furthermore, the experimental method allowed studying effects of variables across a range of scenarios, in which all other variables were kept constant.

However, some potential methodological limitations relating to sample size and representativeness should also be discussed. The sample of police officers ( $N = 60$ ) is not a random sample of the entire Swedish police force but includes participants from distinct geographical areas, of varying ages (25 to 61 years), gender and length of experience (1 to 39 years). Clearly, the student samples may differ from police officers but they were considered appropriate for the purpose of replication (Experiment 2) and examining what potentially could be an effective debiasing technique (Experiment 3) to test with police officers in future research. Provided that the total sample size is 180 participants and that both Experiment 1 ( $N = 60$ ) and 2 ( $N = 60$ ) use full within-subjects designs and the same goes for Experiment 3 ( $N = 60$ ), except for the question format, the total sample affords fairly good statistical power.

Another potential limitation is that participants' own decisions could not be manipulated (since then they would not be their own decisions). This means that two of the cells of the Decision variable (own decision to apprehend and not apprehend) were not experimentally controlled, whereas the remaining four cells were. As a countermeasure, extensive analysis regarding selection effects was carried out and contravened such effects. Note also that such selection effects, even in principle—if they were operative—cannot explain the large observed main effects of apprehension or not and of questioning mode.

Furthermore, there were also some issues related to the study's realism, for instance that the participants did not interact with the suspect or the colleague, that police officers do not use checklists like the IQC during real interrogations and real case material is usually more complex, diverse and presented in a sequential manner. However, research suggests that factors such as interaction with a suspect (Kassin et al., 2003), rich ambiguous information with many loose ends, sequential information retrieval and memory reliance (Ask & Granhag, 2007; Jonas, Schulz-Hardt, Frey, & Thelen, 2001) make real life investigations more vulnerable to bias.

### Implications and Conclusions

The contribution of these experiments to the already existing research on how confirmation bias might influence criminal investigations is twofold. To begin with, it provides a complementary way of thinking about confirmation bias as not only present in

criminal proceedings but also as a result of criminal procedural law. Since police officers apprehend suspects on a regular basis and with the support of procedural law, the significant effects of the apprehension point to a trigger of confirmation bias that is inherent in criminal investigations. As such, the study provides an example of when guilt expectations are formed naturally in criminal cases.

The finding that the apprehension did not have significant effects on the questions that participants chose from the IQC but on the freely generated questions (and that the level of guilt presumption was much higher for the freely generated questions in Experiment 3) can be a result of the different cognitive demands associated with these tasks. This explanation is in line with the research suggesting that confirmation bias is at least in part due to cognitive load (Doherty & Mynatt, 1990; Evans, 2006; Mynatt et al., 1993). It also ties back to dual-process theories and the distinction between System 1, that is, spontaneous, rapid, not entirely controllable and subconscious cognitive processes and System 2, that is, slower, deliberate and conscious cognitive processes (Evans, 2003; Stanovich, 1999; Stanovich & West, 2000). Whereas biases are typically categorized as System 1 processes, researchers increasingly emphasize the inhibitory role of System 2 in suppressing default knowledge and belief-based responses (Gathercole, 2003; Goel & Dolan, 2003; Kahneman & Frederick, 2002). However, System 2 requires working memory, a capacity known to vary between individuals. In fact, working memory and reasoning ability (including hypothetical thinking) are known to be highly correlated (Capon, Handley, & Dennis, 2003; Kyllonen & Christal, 1990; Markovitz, Doyon, & Simoneau, 2002). As such, the findings suggest that when decision makers are under high cognitive load, they may be more susceptible to confirmation bias. However, since the role of cognitive load still needs to be evaluated in more realistic settings, care should be taken in drawing conclusions from these results. In the present research a preset list of questions was used but such a rigid manual-like interrogation method is often inappropriate in real suspect interrogations. Another practically more feasible way to reduce cognitive load is to use interrogation techniques that do not constantly require police officers to come up with new ways of asking open-ended questions but rather use instructions or encouragements such as "Please tell me everything you remember about. . ." (Yarbrough, Hervé, & Harms, 2013, p. 87) or similar phrases that do not convey a belief that the suspect is guilty (Gudjonsson & Pearse, 2011; Meissner, Hartwig, & Russano, 2010). Apart from that such phrases are inherently less guilt presumptive, they are also less cognitively demanding for the police officer (but more cognitively demanding for the suspect). Furthermore, it is possible that continuous education and training in asking open-ended questions reduces cognitive load in the applied setting, which therefore should be evaluated systematically.

Provided the larger effect size of changing questioning mode than changing decision maker (and that the decision maker only had a significant impact on the trustworthiness ratings), the results primarily point to cognitive explanations of confirmation bias. However, since both of the tested techniques (changing questioning mode as well as changing decision maker) were associated with less guilt presumption it indicates that confirmation bias is best understood using both cognitive and social/motivational explanations (e.g., self-enhancing bias). Changing decision maker



between apprehension and interrogation seems to be both a practically feasible and a potentially successful debiasing technique. Whether there are other decision makers available after an apprehension may of course vary with the work situation but these results suggest that a change is appropriate and should be made whenever possible. However, more research is needed regarding to what extent the change of decision maker is successful in real criminal investigations as for instance the influence of different kinds of relationships to a colleague or a prosecutor still has to be examined.

In conclusion, the present study indicates that even if apprehended suspects are supposed to be presumed innocent until they have been convicted, they are not. Furthermore, provided how common apprehensions are, the results suggest that confirmation bias is not necessarily the consequence of some specific and relatively unusual circumstance such as a false confession (Government Offices of Sweden, 2015) but far more common. It is therefore crucial that techniques for mitigating confirmation bias after an apprehension, for example changing decision maker and reducing cognitive load, are further examined and adjusted to the conditions in real life criminal cases.

## References

- Ask, K., & Granhag, P. A. (2005). Motivational sources of confirmation bias in criminal investigations: The need for cognitive closure. *Journal of Investigative Psychology and Offender Profiling*, 2, 43–63. <http://dx.doi.org/10.1002/jip.19>
- Ask, K., & Granhag, P. A. (2007). Motivational bias in criminal investigators' judgments of witness trustworthiness. *Journal of Applied Social Psychology*, 37, 561–591. <http://dx.doi.org/10.1111/j.1559-1816.2007.00175.x>
- Blasi, G., & Jost, J. T. (2006). System justification theory and research: Implications for law, legal advocacy, and social justice. *California Law Review*, 94, 1119–1168. <http://dx.doi.org/10.15779/Z38ZH8P>
- Capon, A., Handley, S. J., & Dennis, I. (2003). Working memory and reasoning: An individual differences perspective. *Thinking & Reasoning*, 9, 203–244. <http://dx.doi.org/10.1080/13546781343000222>
- Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, 51, 1484–1525. <http://dx.doi.org/10.1016/j.visres.2011.04.012>
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24, 349–354. <http://dx.doi.org/10.1037/h0047358>
- Darley, J. M., & Latané, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, 8, 377–383. <http://dx.doi.org/10.1037/h0025589>
- Davis, D., & Leo, R. A. (2006). Strategies for preventing false confessions and their consequences. In M. Kebbell & G. Davies (Eds.), *Practical psychology for forensic investigations and prosecutions* (pp. 121–149). New York, NY: Wiley. <http://onlinelibrary.wiley.com/doi/10.1002/9780470713389.ch7/summary>. <http://dx.doi.org/10.1002/9780470713389.ch7>
- Doherty, M. E., & Mynatt, C. R. (1990). Inattention to P (H) and to P (D ~ H): A converging operation. *Acta Psychologica*, 75, 1–11. [http://dx.doi.org/10.1016/0001-6918\(90\)90063-L](http://dx.doi.org/10.1016/0001-6918(90)90063-L)
- Doyle, A. C. (2001). *The adventures of Sherlock Holmes*. London, UK: The Electric Book Company.
- Dror, I. E., & Cole, S. A. (2010). The vision in “blind” justice: Expert perception, judgment, and visual cognition in forensic pattern recognition. *Psychonomic Bulletin & Review*, 17, 161–167. <http://dx.doi.org/10.3758/PBR.17.2.161>
- Evans, J. S. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences*, 7, 454–459. <http://dx.doi.org/10.1016/j.tics.2003.08.012>
- Evans, J. S. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review*, 13, 378–395. <http://dx.doi.org/10.3758/BF03193858>
- Fahsing, I., & Ask, K. (2013). Decision making and decisional tipping points in homicide investigations: An interview study of British and Norwegian detectives. *Journal of Investigative Psychology and Offender Profiling*, 10, 55–165. <http://dx.doi.org/10.1002/jip.1384>
- Fugelsang, J. A., Stein, C. B., Green, A. E., & Dunbar, K. N. (2004). Theory and data interactions of the scientific mind: Evidence from the molecular and the cognitive laboratory. *Canadian Journal of Experimental Psychology*, 58, 86–95. <http://dx.doi.org/10.1037/h0085799>
- Gathercole, S. (2003). *Short-term and working memory*. London, UK: Taylor and Francis.
- Goel, V., & Dolan, R. J. (2003). Explaining modulation of reasoning by belief. *Cognition*, 87, 11–22. [http://dx.doi.org/10.1016/S0010-0277\(02\)00185-3](http://dx.doi.org/10.1016/S0010-0277(02)00185-3)
- Government Offices of Sweden. (2015). *Rapport från Bergwallkommissionen*. <http://www.regeringen.se/rattsdokument/statens-offentliga-utredningar/2015/06/sou-2015-52/>
- Granhag, P. A., Strömwall, L. A., & Cancino Montecinos, S. (2013). Polisens förhör med misstänkta. Svensk utbildning i internationell belysning. *Rikspolisstyrelsens utvärderingsfunktion, Rapport 2013*, 7, 1–50. [https://polisen.se/Global/www%20och%20Intrapolis/Utv%C3%A4rderingar/2013%20utv%C3%A4rderingsrapporter/Forhor\\_med\\_misstankta\\_2013\\_7\\_.pdf](https://polisen.se/Global/www%20och%20Intrapolis/Utv%C3%A4rderingar/2013%20utv%C3%A4rderingsrapporter/Forhor_med_misstankta_2013_7_.pdf)
- Gudjonsson, G. H., & Pearce, J. (2011). Suspect interviews and false confessions. *Current Directions in Psychological Science*, 20, 33–37. <http://dx.doi.org/10.1177/0963721410396824>
- Hartwig, M., Granhag, P. A., Strömwall, L. A., & Vrij, A. (2005). Detecting deception via strategic disclosure of evidence. *Law and Human Behavior*, 29, 469–484. <http://dx.doi.org/10.1007/s10979-005-5521-x>
- Haverkamp, B. E. (1993). Confirmatory bias in hypothesis testing for client-identified and counsel self-generated hypotheses. *Journal of Counseling Psychology*, 40, 303–315. <http://dx.doi.org/10.1037/0022-0167.40.3.303>
- Hill, C., Memon, A., & McGeorge, P. (2008). The role of confirmation bias in suspect interviews: A systematic evaluation. *Legal and Criminological Psychology*, 13, 357–371. <http://dx.doi.org/10.1348/135532507X238682>
- Isenberg, D. J. (1986). Group polarization: A critical review and meta-analysis. *Journal of Personality and Social Psychology*, 50, 1141–1151. <http://dx.doi.org/10.1037/0022-3514.50.6.1141>
- Jacoby, L. L., Toth, J. P., & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology*, 122, 139–154. <http://dx.doi.org/10.1037/0096-3445.122.2.139>
- Jacoby, L. L., Toth, J. P., Yonelinas, A. P., & Debnar, J. A. (1993). The relationship between conscious and unconscious influences: Independence or redundancy? *Journal of Experimental Psychology*, 123, 216–219. <http://dx.doi.org/10.1037/0096-3445.123.2.216>
- Janis, I. (1982). *Groupthink: Psychological studies of policy decisions and fiascoes*. New York, NY: Houghton Mifflin.
- 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. <http://dx.doi.org/10.1037/0022-3514.80.4.557>
- Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). New York, NY: Cambridge University Press.

- Kassin, S. M. (2005). On the psychology of confessions: Does innocence put innocents at risk? *American Psychologist*, *60*, 215–228. <http://dx.doi.org/10.1037/0003-066X.60.3.215>
- Kassin, S., Dror, I., & Kukucka, J. (2013). The forensic confirmation bias: Problems, perspectives, and proposed solution. *Journal of Applied Research in Memory & Cognition*, *2*, 42–52. <http://dx.doi.org/10.1016/j.jarmac.2013.01.001>
- Kassin, S. M., Goldstein, C. C., & Savitsky, K. (2003). Behavioral confirmation in the interrogation room: On the dangers of presuming guilt. *Law and Human Behavior*, *27*, 187–203. <http://dx.doi.org/10.1023/A:1022599230598>
- Kjäller, H. (2016). *En svensk tiger: Vittnesmål från poliser som vågat ryta ifrån* [A Swedish tiger: testimony from policemen who dared to shake off]. Stockholm, Sweden: Fri Tanke Förlag.
- Klahr, D., Dunbar, K., & Fay, A. L. (1990). Designing good experiments to test “bad” hypotheses. In J. Shrager & P. Langler (Eds.), *Computational models of discovery and theory formation* (pp. 355–402). San Mateo, CA: Morgan-Kaufman.
- Kronkvist, O. (2013). *If truth be known. Police interviews with suspects of serious crime*. Växjö, Sweden: Linnaeus University Press.
- Kruglanski, A. W. (1989). *Lay epistemics and human knowledge: Cognitive and motivational bases*. New York, NY: Plenum Press. <http://dx.doi.org/10.1007/978-1-4899-0924-4>
- Kruglanski, A. W., & Webster, D. M. (1996). Motivated closing of the mind: “seizing” and “freezing.” *Psychological Review*, *103*, 263–283. <http://dx.doi.org/10.1037/0033-295X.103.2.263>
- Kyllonen, P., & Christal, R. E. (1990). Reasoning ability is (little more than) working memory capacity? *Intelligence*, *14*, 389–433. [http://dx.doi.org/10.1016/S0160-2896\(05\)80012-1](http://dx.doi.org/10.1016/S0160-2896(05)80012-1)
- Leo, R. A., & Drizin, S. A. (2010). The three errors: Pathways to false confession and wrongful conviction. In G. D. Lassiter & C. A. Meissner (Eds.), *Police interrogations and false confessions, current research, practice, and policy recommendations* (pp. 9–30). Washington, DC: American Psychological Association.
- Lilienfeld, S. O., & Waldman, I. D. (2017). *Psychological science under scrutiny: Recent challenges and proposed solutions*. New York, NY: Wiley. <http://dx.doi.org/10.1002/9781119095910>
- Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*, 252–271. <http://dx.doi.org/10.1037/0033-295X.87.3.252>
- Markovitz, H., Doyon, C., & Simoneau, M. (2002). Individual differences in working memory capacity and conditional reasoning with concrete and abstract content. *Thinking & Reasoning*, *8*, 97–107. <http://dx.doi.org/10.1080/13546780143000143>
- Meissner, C. A., Hartwig, M., & Russano, M. B. (2010). The need for a positive psychological approach and collaborative effort for improving practice in the interrogation room. *Law and Human Behavior*, *34*, 43–45. <http://dx.doi.org/10.1007/s10979-009-9205-9>
- Mercier, H. (2016). The argumentative theory: Predictions and empirical evidence. *Trends in Cognitive Sciences*, *20*, 689–700. <http://dx.doi.org/10.1016/j.tics.2016.07.001>
- Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*, 57–74. <http://dx.doi.org/10.1017/S0140525X10000968>
- Mill, J. S. (1848). *Principles of political economy*. London, UK: Parker.
- Mortimer, A., & Shepherd, E. (1999). Frames of mind: Schema guiding cognition and conduct in the interviewing of suspected offender. In A. Memon & R. Bull (Eds.), *Handbook of the psychology of interviewing* (pp. 293–315). Chichester, UK: Wiley.
- Myers, D. G., & Lamm, H. (1975). The polarizing effect of group discussion. *American Scientist*, *63*, 297–303. <http://www.davidmyers.org/davidmyers/assets/Polarizing.Eff.Group.Disc.pdf>
- Mynatt, C. R., Doherty, M. E., & Dragan, W. (1993). Information relevance, working memory, and the consideration of alternatives. *The Quarterly Journal of Experimental Psychology*, *46*, 759–778. <http://dx.doi.org/10.1080/14640749308401038>
- Mynatt, C. R., Doherty, M. E., & Sullivan, J. A. (1991). Data selection in a minimal hypothesis testing task. *Acta Psychologica*, *76*, 293–305. [http://dx.doi.org/10.1016/0001-6918\(91\)90023-S](http://dx.doi.org/10.1016/0001-6918(91)90023-S)
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*, 175–220. <http://dx.doi.org/10.1037/1089-2680.2.2.175>
- Pines, J. M. (2006). Profiles in patient safety: Confirmation bias in emergency medicine. *Academic Emergency Medicine*, *13*, 90–94. <http://dx.doi.org/10.1111/j.1553-2712.2006.tb00990.x>
- Pyszczynski, T., & Greenberg, J. (1987). Toward an integration of cognitive and motivational perspectives on social inferences: A biased hypothesis-testing model. I. L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 297–340). San Diego, CA: Academic Press.
- Rajsic, J., Wilson, D. E., & Pratt, J. (2015). Confirmation bias in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *41*, 1353–1364. <http://dx.doi.org/10.1037/xhp0000090>
- Robinson, J. P., Shaver, P. R., & Wrightman, L. S. (1991). *Measures of personality and social psychological attitudes*. New York, NY: Academic Press.
- Schulz-Hardt, S., Jochims, M., & Frey, D. (2002). Productive conflict in group decision making: Genuine and contrived dissent as strategies to counteract biased information seeking. *Organizational Behavior and Human Decision Processes*, *88*, 563–586. [http://dx.doi.org/10.1016/S0749-5978\(02\)00001-8](http://dx.doi.org/10.1016/S0749-5978(02)00001-8)
- Schunn, C. D., & Klahr, D. (1993). Other vs. self-generated hypotheses in scientific discovery. In W. Kintsch (Ed.), *Proceedings of 15th Annual Meeting of the Cognitive Science Society* (p. 900). Mahwah, NJ: Lawrence Erlbaum.
- Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah, NJ: Erlbaum.
- Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, *23*, 645–665. <http://dx.doi.org/10.1017/S0140525X00003435>
- Taylor, S. E., Peplau, L. A., & Sears, D. O. (2005). *Social Psychology*. New Jersey, NJ: Pearson.
- The National Council for Crime Prevention. (2014a). *Persons suspected of offences, final statistics for 2014*. <https://www.bra.se/bra/brott-och-statistik/statistik.html>
- The National Council for Crime Prevention. (2014b). *Reported crimes, final statistic for 2014*. <https://www.bra.se/bra/brott-och-statistik/statistik.html>
- Tulving, E. (2000). *The Oxford handbook of memory*. New York, NY: Oxford University Press.
- Valacich, J. S., & Schwenk, C. (1995). Devil’s advocacy and dialectical inquiry effects on face-to-face and computer-mediated group decision making. *Organizational Behavior and Human Decision Processes*, *2*, 158–173. <http://dx.doi.org/10.1006/obhd.1995.1070>
- van den Eeden, C. A. J., de Poot, C. J., & van Koppen, P. J. (2016). Forensic expectations: Investigating a crime scene with prior information. *Science & Justice*, *56*, 475–481. <http://dx.doi.org/10.1016/j.scijus.2016.08.003>
- Wagenaar, W. A., van Koppen, P. J., & Crombag, H. F. M. (1993). *Anchored narratives: The psychology of criminal evidence*. New York, NY: St. Martin’s Press.
- Wason, P. C. (1966). Reasoning. In B. Foss (Ed.), *New horizons in psychology* (pp. 135–151). Harmondsworth, UK: Penguin Books.
- Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology*, *67*, 1049–1062. <https://www.ncbi.nlm.nih.gov/pubmed/7815301>. <http://dx.doi.org/10.1037/0022-3514.67.6.1049>
- Yarbrough, J., Hervé, H. F., & Harms, R. (2013). The sins of interviewing: Errors made by investigative interviewers and suggestions for redress. In

- B. S. Cooper, D. Griesel, & M. Ternes (Eds.), *Applied issues in investigative interviewing, eyewitness memory, and credibility assessment* (pp. 59–95). New York, NY: Springer. [http://dx.doi.org/10.1007/978-1-4614-5547-9\\_4](http://dx.doi.org/10.1007/978-1-4614-5547-9_4)
- Zenker, F., & Dahlman, C. (2016). Debiasing and rule of law. In E. Feteris, H. Kloosterhuis, J. Plug, & C. Smith (Eds.), *Legal argumentation and the rule of law* (pp. 217–219). The Hague, the Netherlands: Eleven International Publishing.
- Zenker, F., Dahlman, C., Bååth, R., & Sarvar, F. (2016). Giving Reason-spro et contra as a debiasing technique in legal decision making. In D. Mohammed & M. Lewinski (Eds.), *Studies in logic* (pp. 809–823). London, UK: College Publications.

## Appendix

### Evaluation of Selection Effects in the Own Decision Condition

To evaluate possible selection effects of the type of scenario on participants' responses in the condition in which police officers decided about apprehension themselves multiple regressions were carried out. First, a multiple regression was conducted to predict the level of guilt presumption in freely generated questions based on (1) the decision (apprehension vs. no apprehension) and (2) the scenario (1 through 12). If selection effects were present, the scenario variable would significantly predict the level of guilt presumption.

For the decision, a significant regression equation was found,  $F(1, 169) = 5.46, p < .001, R^2 = .22, R^2_{\text{Adjusted}} = .21$ . The analysis showed that the decision significantly predicted guilt presumption in freely generated questions ( $\beta = .95, t = 5.53, p < .001$ ). When adding the scenarios, a significant regression equation was also found,  $F(11, 159) = 4.49, p < .001, R^2 = .22, R^2_{\text{Adjusted}} = .21$ . The decision still significantly predicted guilt presumption in freely generated questions ( $\beta = .45, t = 5.53, p < .001$ ). However, none of the scenarios significantly contributed to the model (Scenario 1:  $\beta = .24, t = .93, p = .352$ ; Scenario 2:  $\beta = .12, t = .39, p = .694$ ; Scenario 3:  $\beta = -.14, t = -0.49, p = .625$ ; Scenario 4:  $\beta = .44, t = 1.49, p = .138$ ; Scenario 5:  $\beta = -.059, t = -.19, p = .850$ ; Scenario 6:  $\beta = .033, t = .081, p = .936$ ; Scenario 7:  $\beta = .004, t = 0.011, p = .991$ ; Scenario 8:  $\beta = -.29, t = -1.037, p = .301$ ; Scenario 9:  $\beta = .15, t = 0.41, p = .686$ ; Scenario 10:  $\beta = .064, t = .23, p = .817$ ; Scenario 11:  $\beta = -.26, t = -.84, p = .404$ ; Scenario 12:  $\beta = -.022, t = -.086, p = .931$ ). Then, the order in

which the predictors were entered was reversed, beginning with the scenario (1 through 12) and then the decision (apprehension vs. no apprehension). When only the scenarios were entered, no significant regression equation was found,  $F(11, 159) = 1.35, p = .246, R^2 = .008, R^2_{\text{Adjusted}} = .002$ . None of the scenarios significantly contributed to the model (Scenario 1:  $\beta = -.003, t = -.006, p = .995$ ; Scenario 2:  $\beta = .12, t = .39, p = .694$ ; Scenario 3:  $\beta = -.14, t = -.49, p = .625$ ; Scenario 4:  $\beta = .44, t = 1.49, p = .138$ ; Scenario 5:  $\beta = -.059, t = -.19, p = .850$ ; Scenario 6:  $\beta = .033, t = .081, p = .936$ ; Scenario 7:  $\beta = .004, t = .011, p = .991$ ; Scenario 8:  $\beta = -.291, t = -1.037, p = .301$ ; Scenario 9:  $\beta = -.18, t = -.56, p = .574$ ; Scenario 10:  $\beta = -.19, t = -.58, p = .565$ ; Scenario 11:  $\beta = -.261, t = -.84, p = .404$ ; Scenario 12:  $\beta = -.12, t = -.38, p = .705$ ). However, when adding the decision, a significant regression equation was found,  $F(11, 159) = 4.49, p < .001, R^2 = .20, R^2_{\text{Adjusted}} = .16$ . The analysis showed that the decision significantly predicted guilt presumption in freely generated questions ( $\beta = .95, t = 5.53, p < .001$ ). These analyses contradict self-selection effects since the scenarios did not significantly predict the level of guilt presumption.

Received October 30, 2017

Revision received March 23, 2018

Accepted March 29, 2018 ■