

From devil's advocate to crime fighter: confirmation bias and debiasing techniques in prosecutorial decision-making

Moa Lidén, Minna Gräns & Peter Juslin

To cite this article: Moa Lidén, Minna Gräns & Peter Juslin (2019) From devil's advocate to crime fighter: confirmation bias and debiasing techniques in prosecutorial decision-making, *Psychology, Crime & Law*, 25:5, 494-526, DOI: [10.1080/1068316X.2018.1538417](https://doi.org/10.1080/1068316X.2018.1538417)

To link to this article: <https://doi.org/10.1080/1068316X.2018.1538417>



© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 31 Oct 2018.



[Submit your article to this journal](#)



Article views: 4433



[View related articles](#)



[View Crossmark data](#)



Citing articles: 6 [View citing articles](#)

From devil's advocate to crime fighter: confirmation bias and debiasing techniques in prosecutorial decision-making

Moa Lidén, Minna Gräns and Peter Juslin

Faculty of Law, Uppsala University, Uppsala, Sweden Department of Psychology, Uppsala University, Uppsala, Sweden

ABSTRACT

This research examines the role of confirmation bias in prosecutorial decisions before, during and after the prosecution. It also evaluates whether confirmation bias is reduced by changing the decision maker between arrest and prosecution. In Experiment 1, Swedish prosecutors ($N=40$) assessed 8 scenarios where they either decided themselves or were informed about a colleague's decision to arrest or not arrest a suspect. Participants then rated how trustworthy the suspect's statement was as well as the strength of new ambiguous evidence and the total evidence. They also decided whether to prosecute and what additional investigative measures to undertake. In Experiment 2 the same method was used with Law and Psychology students ($N=60$). Overall, prosecutors' assessments before the prosecution indicated that they were able to act as their own devil's advocate. Also, their assessments while deciding about whether to prosecute were reasonably balanced. However, after pressing charges, they displayed a more guilt-confirming mindset, suggesting they then took on the role as crime fighters. This differed from the student sample in which higher levels of guilt confirmation was displayed in relation to arrested suspects consistently before, during and after a prosecution decision. The role of prosecutors' working experience is discussed.

ARTICLE HISTORY

Received 26 April 2018
Accepted 12 October 2018

KEYWORDS

Confirmation bias; arrest; prosecute; debiasing

Introduction

The prosecutor's multiple roles as not only the leader responsible for the criminal inquiry but also the defendant's counter party in Court, have been recognized by several researchers (Green & Roiphe, 2017; Heuman, 2004). Many have also pointed to the very high, or even untenable, demands that these multiple roles put on prosecutors' cognitive flexibility, as prosecutors are required to justify decisions to arrest, press charges, etc. and simultaneously, strictly adhere to objectivity demands (Jacobson, 2008; Lindberg, 1997; O'Brien, 2009). Although prosecutors' more specific roles vary in different jurisdictions, their involvement in criminal cases often begins already with the opening of a preliminary

CONTACT Moa Lidén  moa.liden@jur.uu.se

 Supplemental data for this article can be accessed at <http://dx.doi.org/10.1080/1068316X.2018.1538417>.

© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

investigation, during which they make several judgments regarding the suspect, before the suspect's guilt has been tried in Court. As implied by extensive research on confirmation bias, this poses a risk that prosecutors form a hypothesis and then, more or less subconsciously, selectively search for and evaluate information in ways that are partial to that hypothesis (Klayman & Ha, 1987; Nickerson, 1998; Wason & Johnson-Laird, 1972). Provided the considerable influence prosecutors' decisions have for the focus and frames of the criminal inquiry and proceedings (Englich & Mussweiler, 2001; Englich, Mussweiler, & Strack, 2005), such a confirmation bias greatly increases the risk of wrongful suspicions or even wrongful convictions. Thus, the present research aims to examine the prevalence of confirmation bias in prosecutors' decision making in three stages following the chronology of criminal cases, stage (1) before the prosecution, stage (2) the prosecution decision and stage (3) after the prosecution. The research also aims to test one potential debiasing technique, that is, to change decision maker between arrest and prosecution.

Apart from prosecutors' multiple and sometimes conflicting roles (Green & Roiphe, 2017; Heuman, 2004), research also highlights other general risk factors of confirmation bias such as prosecutors' wide discretion (Babcock & Loewenstein, 1997; Hobbs, 1950; Levinson & Young, 2010), emotions and empathy, i.e. the capacity to read someone else's emotions (Wettergren & Bergman Blix, 2016) as well as institutional and societal incentives to be tough on crime (Hobbs, 1950; O'Brien, 2009).

However, more research is needed regarding at what point in time and in which more specific situations confirmation bias might arise. Some studies suggest that it can arise at a very early stage, for instance when the prosecutor decides who should be investigated and through what tactics (Griffin, 2001) or with the identification and/or arrest of a suspect (Fahsing & Ask, 2013). Such a premature narrowing of the inquiry can make prosecutors downgrade exculpatory evidence, for instance by categorizing it as irrelevant (Burke, 2006; Jonakait, 1987; McCloskey, 1989; Wastell, Weeks, Wearing, & Duncan, 2012; Yaroshesky, 1999) and also influence their judgments of whether to press charges, what charges, what sentence to seek upon conviction, etc. (Davis, 2001; Green, 2003; Griffin, 2001).

Another view is that prosecutorial neutrality is at its peak prior to the charging decision (Burke, 2006), possibly because prosecutors use this time period to carefully evaluate whether there are sufficient reasons to prosecute and thereby act as their own devil's advocates. However, with the charging decision the guilt hypothesis is consolidated since prosecutors only press charges if they are sufficiently certain of the accused's guilt (Freedman, 1990; Gershman, 2001) and in this assessment they anticipate the prospects of a conviction (Ernberg, Tidefors, & Landström, 2016; Lievore, 2005; Wettergren & Bergman Blix, 2016). This suggests that the charging decision results in a psychological shift whereby prosecutors become crime fighters whose primary objective is to get defendants convicted.

Thus, once charges have been pressed, prosecutors may primarily strive to make the crime and perpetrator obvious to the Court, for instance by mainly or exclusively conducting an additional investigation that can confirm the defendant's guilt (Davis, 2001; Green, 2003; Griffin, 2001). Such a guilt-confirming mindset was found by Engel and Glöckner (2013) among participants who were assigned the role of prosecutors and who still wanted the defendant to be convicted and more often

downplayed conflicting evidence than participants in the defense role, even when they were no longer in their assigned roles and offered 100 euros if their guilt decisions were accurate. This is important since prosecutors, with the charging decision, take on the role as the defendant's counter party and are likely to act in accordance with what they perceive is mandated by that role (Haney, Banks, & Zimbardo, 1973; Thompson & Loewenstein, 1992). Also, prosecutors are likely to feel accountable for their charging decision which increases the risk of defensive bolstering, that is, tenaciously holding on to a previous position even in the face of contrary evidence (Lerner & Tetlock, 1999; O'Brien, 2009). Furthermore, at this stage, prosecutors have been required to construct a relatively well-shaped narrative to explain an event (Wagenaar, van Koppen, & Crombag, 1993). For instance, a prosecutor recognizes a killing as drug-related or stemming from a domestic dispute and these so-called pattern-matching models influence which information they attend to (Bilalić, McLeod, & Gobet, 2008; Epelboim & Suppes, 2001; Hecht & Proffitt, 1995).

Since the prosecutors' more specific roles, levels of discretion, etc. vary between different jurisdictions (Gilliéron, 2013; Griffin, 2001), the more specific situations in which confirmation bias might arise will also vary. As this study concerns the Swedish legal context, it has been designed to, as accurately as possible, reflect the roles of Swedish prosecutors. This entails being the leader of criminal inquiries, deciding about whether to press charges as well as to act as the defendant's counterparty in Court (Lindberg, 1997). Although these roles include a range of judgments, decisions, etc. and they cannot all be examined within the frames of this study, the design provides a methodology for evaluating whether and when, during the course of the criminal inquiry and proceedings, prosecutors, gradually or suddenly, come to believe that a suspect is guilty. This attempt to detect a change in prosecutors' mindset is not only of interest to the Swedish legal context but of much wider interest since there seems to be different notions of when prosecutors can or cannot be expected to remain objective in different jurisdictions (Dickerson Moore, 2000; Gilliéron, 2013).

In a European context, the historically important distinction between the investigative phase and the trial phase, where the investigation was often carried out by an examining magistrate whereas the indictment was issued and presented by a public prosecutor, is today maintained to different extents in different European countries (Gilliéron, 2013). In Sweden, prosecutors' involvement in the investigative phase is not generally considered to make them unsuited to also act in Court (The Swedish Government Bill, Prop. 2000/01:92, p. 20) even if many scholars have discussed this risk (Jacobson, 2008; Lindberg, 1997; Wettergren & Bergman Blix, 2016). However, in France, where prosecutors today run close to all inquiries (Gilliéron, 2013), it has been considered necessary to counterbalance the power vested in the prosecution and this is the role of the investigative judge, who for instance has a lot of influence when it comes to coercive measures (Gilliéron, 2013). Dutch prosecutors, who are also involved both during the investigative phase, during which they supervise the police and are formally responsible for all aspects of the criminal investigation, and the trial phase, have been described as possessing 'enormous powers' (Tak, 2012, p. 7). For more serious crimes, Dutch prosecutors may apply to an investigative judge for a preliminary judicial investigation, although the decision to institute criminal proceedings has been exclusively reserved for the prosecutors (Tak, 2012).

In the U.S., the prosecutors' roles are quite different not only because prosecutors are often generally elected rather than appointed (Jacoby, 1980; Ramsey, 2002), but also because the prosecutors do not generally play a role in the investigative phase (Gilliéron, 2013). Thus, decisions about what crimes are investigated and who is arrested rest under police discretion (Harris, 2011; Weigend, 1978). Depending on state jurisdiction, even the decision to charge a suspect, at least for misdemeanors, is sometimes left to the police (Abadinsky, 1998). Thus, in such situations, the prosecutor's office receives the case only after charges against the individual have already been filed. A potential drawback of this is that the police might lack sufficient knowledge of complexities in substantive and procedural law, even if such knowledge is of clear relevance for decisions made during criminal inquiries (Gilliéron, 2013; Weigend, 1978). Also, depending on the size of the US Attorney's office, cases are prosecuted either vertically, which means that a single prosecutor is responsible for a case from start to finish, or horizontally, where different sections of the attorney's office handle the same case at different stages (Abadinsky, 1998). Small and medium offices will more likely follow the horizontal prosecution, while larger offices consist of different specialized sections that handle specific types of crimes (Gramckow, 2008). Thus, across different legal contexts, the notion of when the prosecutor is to become involved in a case varies and there also seems to be different ideas about the need to include another decision maker (such as the investigative judge or another prosecutor). This makes the division of prosecutorial decision-making into different stages, and examination of the prevalence of bias in these different stages, relevant. Also, testing the potential importance of a second decision maker is of interest.

However, the question of whether and to what extent confirmation bias is at play in prosecutorial decision making is unlikely to be only a question of 'when' and 'who' but, most probably, it is also a question of 'what', that is, for what type of tasks (judgments, decisions, etc.), the fear of bias is justified. For instance, prosecutors who themselves search for, or direct searches for, criminal evidence, may be more prone to bias than prosecutors who simply evaluate the evidence presented to them by the police. The evaluation of evidence preceding a decision about prosecution is common to all the jurisdictions described above. Also, the decision to prosecute is of utmost importance for the focus and content of criminal trials. As such, any bias in this evaluation would be of clear practical importance. Therefore, this study examines the evaluative component in the stage before and during the prosecution decision. However, in real life criminal cases, prosecutors also search for or at least supervise the search for, criminal evidence. This search for evidence will only be tested at the stage after prosecutors have decided about prosecution, in order to see whether their decisions about prosecution are related to what type of evidence they search for. In the Swedish context, prosecutors are not to carry out investigation themselves as this is believed to endanger their objectivity (Lindberg, 1997). They do decide about which inquiries shall be carried out as well as in which order but how such inquiries are made is usually a question for the police (The Swedish Government Official Reports, SOU 1988:18 pp. 112–113 and p. 295).

A common situation in the Swedish setting is that the police has decided to apprehend a suspect (Lidén, Gräns, & Juslin, 2018) and the prosecutor then has to decide whether the deprivation of liberty shall be maintained, that is, whether the suspect shall be arrested. In this assessment, they have to evaluate the evidence available to decide whether the evidentiary requirements for an arrest are fulfilled. As such, often when prosecutors' first get

in touch with a criminal case, their primary task is that of evaluation. Also when deciding about whether to prosecute, they have to conduct careful evaluations of the available evidence. Although prosecutors may of course also have been involved in the search for evidence at such early stages (in the sense that they supervised this search), the described situation is common and therefore relevant and appropriate to study empirically. Subsequently, this study examines the prevalence of confirmation bias in the prosecutor's evaluation of evidence before and during the decision about whether to press charges. The study also examines the prevalence of confirmation bias at a later stage, when the decision about whether to press charges has been made, and prosecutors direct any search for additional evidence. Since it is also of relevance to study confirmation bias in searches conducted at earlier stages, the basic experimental paradigm employed here can be adjusted to test such factors as well, even if it has not been possible in this study. If prosecutors instead (or also) had been asked to search for criminal evidence, they most probably would have searched for different pieces of evidence, and any variation in their evaluations of these different pieces of evidence could be due to inherent differences in the indications of the evidence. As such, this would defeat the purpose of examining whether prosecutors evaluate the same pieces of evidence differently, depending on whether the suspect had been arrested and who made that decision.

As implied by the above, even if the experimental design is specifically intended to test bias among Swedish prosecutors, the methodology with clear divisions into different stages, different types of tasks as well as the testing of a change of decision maker, has potential importance for understanding prosecutorial decision making in a more general sense. Across the described jurisdictions, it varies when prosecutors get involved, to what extent they get involved (what types of tasks) and the usage of a second decision maker (another prosecutor or an investigative judge) is at play in many of the jurisdictions.

As such, this research relates and adds to the already existing research on so-called decisional tipping points, decisions that result in a mindset where criminal investigators become focused on verifying the guilt of a suspect (Fahsing & Ask, 2013). Here, decisional tipping points in prosecutors are addressed and directly related to different stages, tasks and decision makers.

Prosecutorial shortcomings such as failures to disclose exculpatory evidence are often conceptualized as conscious misconduct (Flowers, 1998; Gershman 1998; Henning, 1999; Hetherington, 2002). This is also reflected in that suggested remedies often involve improving prosecutorial values and/or increasing the frequency and severity of sanctions against unethical prosecutors (Davis, 2001; Heller, 1997), which is unlikely to be effective for confirmation bias that acts largely outside of conscious awareness. A more feasible approach is to consult the explanations of confirmation bias and convert them into testable debiasing techniques. These explanations, which are not mutually exclusive, are the cognitive explanations; that confirmation bias arise due to capacity limits in human attention, memory, etc. (Carrasco, 2011; Doherty & Mynatt, 1990; Mynatt, Doherty, & Dragan, 1993; Mynatt, Doherty, & Sullivan, 1991; Rajsic, Wilson, & Pratt, 2015) as well as the social and motivational explanations which are closely tied to each other and portray confirmation bias as a self-enhancement bias (Greenwald, 1980; Munro & Ditto, 1997; Munro & Stansbury, 2009). This implies that humans reason in a one-sided way in order to convince others that they are right (Mercier, 2016; Mercier & Sperber, 2011) and want to defend their ideas and behaviors in order to maintain control and self-esteem

(Pyszczynski & Greenberg, 1987). In line with this, research has found that confirmation bias seems to be stronger in relation to self-generated hypotheses than others' hypotheses, in both the legal context (Lidén et al., 2018) and other contexts (Dunbar & Klahr, 1989; Haverkamp, 1993; Klahr, Dunbar, & Fay, 1990; Schunn & Klahr, 1992, 1993). In this research, the notion of confirmation bias as a self-enhancement bias is tested by changing decision maker between the arrest and prosecution to see whether this is functional as a debiasing technique. Hence, in line with previous research it is expected that:

Hypothesis (1) Prosecutors will display a stronger guilt presumption (in the evaluation of evidence and decisions to prosecute) with regard to arrested than non-arrested suspects.

Hypothesis (2) Prosecutors will demonstrate a stronger guilt presumption in relation to own decisions about arrest than a colleague's decisions to arrest.

Hypothesis (3) After having decided to prosecute, prosecutors will be less inclined to undertake additional investigation and the investigation they do undertake will be more guilt presumptive in character, than after decisions to not prosecute.

Experiment 1

Method

Participants

Study participants were 40 Swedish prosecutors (22 women, 18 men) from Swedish urban and rural districts. They were approached through personal contact with representatives of the different districts and informed that the study was about prosecutorial decision-making, encompassing e.g. evaluations of evidence in connection to a decision about prosecution. In total 742 prosecutors who regularly make decisions about arrests and prosecutions were asked to participate and participation was voluntary. Prosecutor's ages varied between 30 and 64 years ($M = 46.66$, $SD = 10.52$) and their length of experience as prosecutors ranged from 1 year to 35 years ($M = 16.61$, $SD = 10.60$). As illustrated by the Prosecution Authority's Statistics (2017), the sample is representative of the overall population of prosecutors, of which 60.70% are female and 49.30% male, with an average age of 43 years, despite the attrition. The study (including both experiments) was approved by the Regional Ethical Review Board in Uppsala, Sweden, before the commencement of data collection.

Design

The design was different for the three examined stages of decision making. In Stage (1) before the prosecution, a 2 (decision: arrest vs. no arrest) x 2 (decision maker: self vs. colleague) within subjects design was used in relation to the following dependent measures: the suspect's trustworthiness, ratings of ambiguous evidence and total evidence. Then, in Stage (2) the decision (about arrest), the decision maker (who decided about the arrest), trustworthiness, ambiguous and total evidence were used to predict decisions about whether to prosecute or not. After this, in Stage 3, the decision about prosecution, the decision about the arrest, the decision maker (who made the arrest), trustworthiness, ambiguous and total evidence were used to predict decisions about whether the additional investigation was necessary as well as the level of guilt confirmation in the additional investigation.

Due to one of the questions addressed in Stage 1 – the effect of whether participants themselves or a colleague decided about the arrest – this independent variable is experimentally controlled only in the two colleague’s decision cells of the design, whereas this variable is controlled by the participant’s own decision in the two own decision cells. Thus, whereas the scenarios are perfectly counter-balanced with a Latin Square in the colleague’s decision cells, the design opens up for self-selection effects in the own decision cells (since some scenarios may become overrepresented in the category arrested relative to not arrested due to the participants’ own decisions). If the scenarios had inherently different capacity to elicit guilt presumption, a different response pattern would be expected for own decisions and the colleague’s decision (which were experimentally controlled). The potential problems arising from self-selection are addressed by an additional analysis in Appendix. This analysis speaks against an interpretation in terms of selection effects. As such, it also suggests that scenario variability did not significantly influence the results.

Material

The 8 scenarios were presented to participants using a web-based survey. To enable testing of the hypotheses in a range of different and, as far as possible, realistic settings, the scenarios were all inspired by real criminal cases and unique with regard to for instance the crime type, the suspect’s gender, ethnicity, age, criminal record, etc. To further promote external validity, crime and suspect statistics from the National Council for Crime Prevention (2015a, 2015b) was incorporated. Accordingly, 2 out of 8 scenarios (25.00%) concerned crimes of stealing and 1 out of 8 scenarios (12.50%) concerned crimes against a person, inflicting damage, fraud, narcotic drug offences, traffic crimes and other crimes respectively. Furthermore, the suspect was male in 6 out of 8 scenarios (75.00%) and female in 2 out of 8 scenarios (25.00%). Also, since statistics were lacking regarding the suspects’ ethnicity (Swedish/fully or partially foreign) and criminal record (yes/no), these factors were both approximated to 50% of the scenarios. The scenarios are available as Online Supplemental Material.

Procedure

Participants were informed they would read and make decisions in 8 scenarios and that these scenarios were inspired by authentic criminal cases but that the circumstances had been changed to preserve confidentiality. Table S1, that is available as Online Supplemental Material, displays how the condition and scenario orders were counter-balanced across trials as well as how the scenarios were aggregated in the analysis. As illustrated by Table S1, all participants read 4 scenarios in each decision maker condition (self/colleague) before the next condition was introduced. For own decisions, participants themselves decided when to arrest and not arrest. For the colleague’s decisions, 2 decisions were arrests and 2 decisions were non-arrests and the order of these decisions was systematically counterbalanced (either 2 arrests followed by 2 non arrests or every other arrest/non arrest). Also, the decision maker variable was systematically counterbalanced (participants started either by making own decisions or by reading about the colleague’s decisions). Furthermore, to avoid confounders (such as the suspect’s criminal record and ethnicity), the order of the scenarios were systematically counter-balanced across the conditions using a Latin square and appeared the same number of times in

each condition ($8 \times 5 = 40$ prosecutors), see Table S1. There were in total 4 condition orders (see the orders in Trials 1–4 that are then repeated for Trials 5–8) and 8 scenario orders (Trials 1–8).

By the beginning of each scenario, participants were informed that they had been assigned a case concerning a certain crime. Across all conditions, participants read the same case information consisting of a brief police report containing the suspect's name, age, ethnicity, a description of the suspected crime and a suspect interrogation (the police's summary). In the own decision condition participants were asked to decide whether the suspect should be arrested and if they answered yes, they also indicated the reasons for this whereas in the colleague condition they were simply informed about the colleague's decision and the reasons for it. After having decided whether to arrest the suspect (or been informed about a colleague's decision), a new piece of ambiguous evidence was introduced. Then, participants rated how trustworthy the suspect was, how strongly the ambiguous evidence, as well as the total evidence, indicated the suspect's guilt. They also decided whether to press charges and whether and what additional investigation was necessary.

Measures

In Stage (1) total mean scores were calculated for each individual in each of the 4 cells of the experimental design as regards the suspect's trustworthiness, the ambiguous evidence, and the total evidence. In Stage (2) the variables: decision about arrest, decision maker, trustworthiness, ambiguous and total evidence were used to predict decisions about prosecution and in Stage (3) the variables: decision about prosecution, decision about arrest, decision maker, trustworthiness, ambiguous and total evidence were used to predict decisions about additional investigation as well as the level of guilt confirmation in the additional investigation.

Trustworthiness ratings

The question about trustworthiness was directed at the suspect's statement (rather than the suspect) which was described in the summary of the interrogation provided in each scenario (*Step (1)*). Since, in the real cases, it was considerably more common that the suspect denied than confessed, the suspect denied in 7 scenarios and confessed in 1 scenario and these were always the same scenarios. The purpose of including both confessions and denials was not to compare trustworthiness ratings for each category but instead to avoid any effects of base rate expectations that probably would have occurred had all the scenarios concerned confessions or denials or had the number of confessions and denial been the same, since this probably does not reflect prosecutors' ideas of how common confessions and denials are. Participants rated the trustworthiness of the suspect's statement on a scale from 1 to 7 where 1 meant 'not at all trustworthy' and 7 meant 'completely trustworthy'.

Ambiguous evidence

Prior to data collection, the ambiguity of the evidence was tested by having four independent judges rate the evidence on a scale from 1 to 7 where 1 meant 'inconsistent with the suspect's guilt', 4 meant 'ambiguous' and 7 meant 'consistent with the suspect's guilt'. The mean ratings were for Scenario (1) $M = 3.50$, $SD = 1.21$, 95% CI [2.89–3.72], Scenario (2), $M = 4.25$, $SD = .89$, 95% CI [3.21– 5.15], Scenario (3), $M = 4.00$, $SD = .67$, 95% CI [3.12–

4.67], Scenario (4) $M = 4.75$, $SD = .74$, 95% CI [3.41–5.05], Scenario (5), $M = 3.50$, $SD = .54$, 95% CI [2.34–3.52], Scenario (6), $M = 2.75$, $SD = .32$, 95% CI [2.05–3.76], Scenario (7), $M = 4.50$, $SD = .54$, 95% CI [3.45–5.22] and Scenario (8), $M = 3.75$, $SD = .31$, 95% CI [3.24–4.34]. The inter-rater reliability for the four judges was $ICC = .88$, 95% CI [.72–.94]. This indicates that the evidence was perceived of as ambiguous and that there were no significant differences between the different pieces of evidence presented in the different scenarios. During data collection, participants were asked to indicate how strongly the evidence indicated the suspect's guilt on a scale from 1 to 7 where 1 meant 'very weakly' and 7 meant 'very strongly'.

Total evidence strength

Participants made an overall assessment of how strongly the total evidence indicated the suspect's guilt, using the same scale and labels as for the ambiguous evidence (1–7).

Prosecution

Participants also decided whether the evidence was sufficient for prosecuting the suspect. Since the Swedish standard of proof for prosecution is *sufficient reason*, this question was formulated: 'Are there sufficient reasons to prosecute the suspect?' and participants answered 'yes' or 'no'.

Additional investigation

Regardless of whether participants decided to prosecute or not, they indicated whether ('yes' or 'no') additional investigation was necessary. If they answered 'yes' to this question, they were also asked to indicate (in free text) which additional investigation was necessary, with a minimum of 1 and a maximum of 3 investigative measures.

To assess the character of the suggested additional investigation, three independent judges, as well as the first author, rated all of the prosecutors' answers on a scale from 1 to 7 where 1 meant 'completely aimed at confirming the suspect's innocence', 4 meant 'neither aimed at confirming the suspect's innocence nor guilt' and 7 meant 'completely aimed at confirming the suspect's guilt' and mean scores for all four judges' ratings were produced. The raters were all blind to the conditions in which the additional investigation had been suggested. The inter-rater reliability for the three judges was $ICC = .65$, 95% CI [.47–.76].

Social desirability

After having completed all 8 scenarios, participants answered 14 questions from a short version of the Marlowe-Crowne Social Desirability Scale (M-C SDS) (Crowne & Marlowe, 1960; Robinson, Shaver & Wrightman, 1991) in order to assess response bias, which in this specific context could be at hand if participants realized what the purpose of the study was and adjusted their ratings and decisions in order to appear as if they were not influenced by previous decisions (i.e. the arrest or prosecution). Participants 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 participants' ratings of the defendants' trustworthiness, $r = -.011$, $p = .950$, the ambiguous evidence,

$r = .144$, $p = .423$, the total evidence strength, $r = -.066$, $p = .716$ or the character of the suggested additional investigation, $r = -.087$, $p = .630$, indicating that participants' responses were not significantly influenced by response bias.

Assumed purpose of the study

Lastly, participants stated what they thought was the purpose of the study (free text). The mentioned purposes were divided into five categories based on their content. These were (1) evaluation of evidence in general (31.60%), (2) prejudice and/or moral flaws in prosecutors (21.90%), (3) the participant had no idea what the purpose was (21.90%), (4) whether different prosecutors assess the same cases differently (18.30%) and (5) whether prosecutors are influenced by the suspect's gender, ethnicity, etc. (6.30%). None of the prosecutors mentioned bias, suggesting that they did not realize (or did not say that they realized) what the purpose of the study was. However, most of the responses in this regard were relatively brief and it is, of course, possible that prosecutors who answered e.g. 'prejudice' in fact mean bias or similar issues even if these words (in Swedish), in a literal sense, mean different things. Regardless of this uncertainty, any realization of the purpose of the study, that prosecutors disclosed, is unlikely to have influenced prosecutors' responses, since the assumed purpose of the study did not significantly influence the participants' ratings of the defendants' trustworthiness, $F(5, 35) = .662$, $p = .655$, the ambiguous evidence, $F(5, 35) = .555$, $p = .733$, the total evidence strength, $F(5, 35) = .164$, $p = .973$ or the character of the suggested additional investigation, $F(5, 35) = .342$, $p = .882$. If the prosecutors did realize what the purpose of the study was and chose to not disclose it, but were still influenced by this realization, such an influence would probably be visible in differences between the results obtained using within-subject analysis, the main analysis employed here, and between subject analysis regarding only the first conditions, when subjects had not yet been exposed to the experimental manipulations. The possibility of carry over effects is addressed by additional analyses in Appendix. Overall, the between-subject analyses indicate similar trends as the trends found using the within-subject analysis, which speaks against an interpretation in terms of carry-over effects.

Results

Descriptive statistics (M, SD) and post hoc comparisons for significant effects are displayed in Table 1.

Background variables

Exploratory analyses were conducted in order to examine whether participants' background characteristics, that is, gender, age and years of experience, were related to any of the dependent variables, which was not the case, see Appendix.

Trustworthiness ratings

A 2 (decision: arrest vs. no arrest) \times 2 (decision maker: self vs. colleague) Repeated Measures ANOVA was conducted. There was a significant effect of the decision on the ratings of the suspect's trustworthiness, $F(1, 34) = 42.57$, $p < .001$. Overall, arrested suspects were rated as less trustworthy than non-arrested suspects and the size of this effect was

Table 1. Means, SD and *post hoc* comparisons for significant effects by dependent variable, decision and decision maker in Experiment 1.

DM	D						Post hoc	
	A		NA		Total			
	M	SD	M	SD	M	SD		
<i>Trustworthiness ratings (1–7)</i>								
S	2.59	.86	3.60	1.39	3.10	1.13	<i>p</i> < .001 A vs. NA	
C	2.85	.81	3.59	1.10	3.22	.95		
Total	2.72	.84	3.60	1.25	3.16	1.04		
<i>Ambiguous evidence (1–7)</i>								
S	3.87	1.04	4.09	1.68	3.98	1.36	<i>p</i> < .001 for NA vs. A***	
C	3.56	1.08	3.92	1.58	3.74	1.33		
Total	3.72	1.06	4.01	1.63	3.86	1.35		
<i>Total evidence strength (1–7)</i>								
S	4.59	.88	4.99	1.33	4.79	1.11	<i>p</i> < .001 for NA vs. A***	
C	4.24	1.16	5.02	.98	4.63	1.07		
Total	4.42	1.06	5.01	1.16	4.71	1.09		
<i>Prosecution (and no prosecution) rates %</i>								
S	34.10 (65.90)		50.70 (49.30)		42.40 (57.60)			
C	36.00 (64.00)		50.70 (49.30)		43.35 (56.65)			
Total	35.00 (65.00)		50.70 (49.30)		42.86 (57.14)			
<i>Additional (and no additional) investigation rates %</i>								
P		DP				NP		
61.70 (38.30)						78.70 (21.30)		
	A	NA	D	NA	Total			
S	64.70 (35.30)	65.50 (34.50)	57.60 (42.40)	89.30 (10.70)	69.28 (30.72)			
C	50.00 (50.00)	70.40 (29.60)	86.50 (13.50)	75.00 (25.00)	70.48 (29.52)			
Total	57.35 (42.65)	67.95 (32.05)	72.05 (27.95)	82.15 (17.85)	69.88 (30.12)			

(Continued)

Table 1. Continued

Guilt confirmation in additional investigation (1–7)

	P		DP						NP		Total	Total
	4.61 (1.86)		NA		A		NA		3.39 (1.65)			
	<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>		
S	5.66	1.44	2.91	1.35	3.16	1.67	2.69	1.14	3.61	1.40	4.40	1.70
C	4.65	1.72	5.21	1.64	4.13	1.78	3.59	1.63	4.40	1.70	4.00	1.60
Total	5.16	1.58	4.06	1.50	3.65	1.73	3.14	1.39	4.00	1.60		

Note: D = Decision; A = Arrest, NA = No arrest; DM = Decision maker; S = Self; C = Colleague; DP = Decision about Prosecution, P = Prosecution, NP = No prosecution.

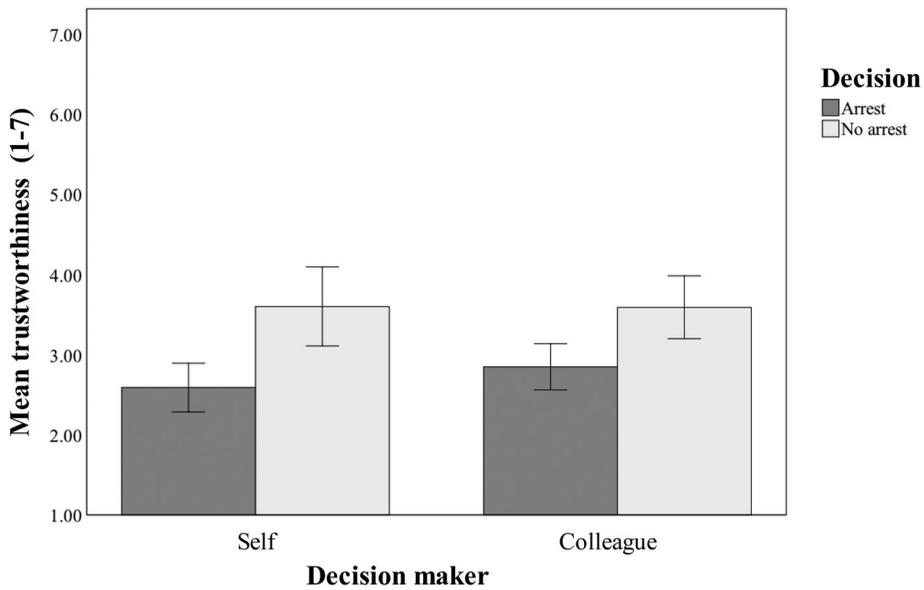


Figure 1. Mean trustworthiness ratings by decision and decision maker in Experiment 1. Error bars represent 95% CI.

$r = .74$. For more on r as an effect size measure see Field, 2012. The effect found for the trustworthiness ratings is illustrated in Figure 1. However, there was no significant effect of decision maker, $F(1,34) = .58, p = .451$ and no significant interaction between decision and decision maker, $F(1,34) = .39, p = .535$.

Ambiguous evidence

A 2 (decision: arrest vs. no arrest) \times 2 (decision maker: self vs. colleague) Repeated Measures ANOVA was conducted. No significant effects were found for decision, $F(1,34) = 1.36, p = .252$ or decision maker, $F(1,34) = 1.18, p = .286$ and there was no significant interaction, $F(1,34) = .107, p = .746$.

Total evidence

A 2 (decision: arrest vs. no arrest) \times 2 (decision maker: self vs. colleague) Repeated Measures ANOVA was conducted. There was a significant effect of decision, $F(1,34) = 10.93, p = .004$. As illustrated by Figure 2, the total evidence was rated as stronger when the suspect had not been arrested as compared to when the suspect had been arrested and the size of this effect was $r = .49$. However, there was no significant effect of decision maker, $F(1,34) = .73, p = .399$ and no significant interaction, $F(1,34) = .81, p = .375$.

Prosecution

A logistic regression was performed to examine whether decisions about the arrest, decision maker as well as ratings of trustworthiness, ambiguous and total evidence

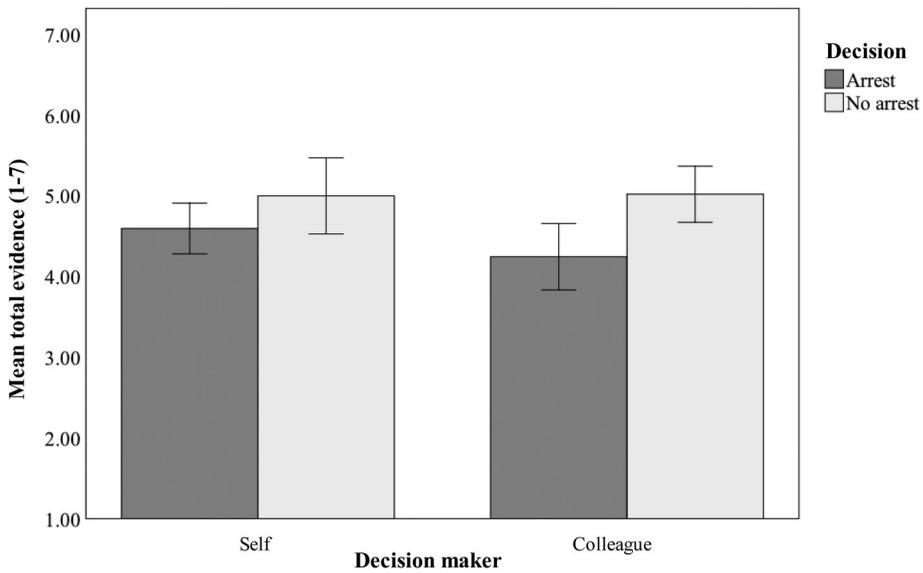


Figure 2. Mean ratings of total evidence strength by decision and decision maker in Experiment 1. Error bars represent 95% CI.

predicted decisions about prosecution. The categorical variables were dummy coded using the following codes, for decision, 1 = arrest, 0 = no arrest and for decision maker, 1 = self, 0 = colleague. Using the Enter method, non-significant predictors were excluded; see Table A1 in Appendix. When rerunning the model without the non-significant predictors, the results in Table 2 were obtained.

Table 2 suggests that the importance of the trustworthiness rating depended on whether the suspect had previously been arrested or not, because in interaction these variables significantly predicted decisions about prosecution. If the previous decision was to arrest (1) the suspect, the odds of a prosecution increased as the trustworthiness rating decreased and vice versa, the odds of a prosecution decreased as the trustworthiness rating increased. The odds ratio (and the negative value of the *b*-value) tells us that the odds of a prosecution was .52 times higher for each unit with which the trustworthiness rating (1–7) decreased. If instead, the previous decision was to not arrest (0) the suspect, the odds of a prosecution did not change as a result of the trustworthiness rating. However, the predictor most closely related to the decision about prosecution was the

Table 2. Predictors of decisions about prosecution in Experiment 1 using logistic regression analyses (enter method).

Variable	Decisions about prosecution			
	<i>b</i> 95% CI	95% CI for odds ratio		
		Lower	Odds	Upper
Constant	-10.82 *** [-14.59, -9.17]			
Arrest x Trustworthiness	-.65* [-1.53, -0.14]	.31	.52	.88
Total evidence	1.93*** [1.49, 2.74]	4.21	6.92	11.35

Note: Model $\chi^2(4) = 211.01, p < .001. R^2 = .69$ (Hosmer & Lemeshow).
p* < .05. *p* < .01. ****p* < .001.

rating of the total evidence strength. As the perceived evidence strength increased, the odds of a prosecution increased with 6.92 (per unit). Taken together, the decision about prosecution seems to be primarily related to prosecutors' ratings of the total evidence strength but a previous arrest (regardless of who made the arrest) also increased the odds of a prosecution, if combined with a low rating of trustworthiness.

Additional investigation

The additional investigation was analyzed in two steps, (1) whether the additional investigation was deemed necessary (yes/no), labeled additional investigation rate below, and (2) the level of guilt confirmation (1–7) in the suggested additional investigation.

Additional investigation rate

A logistic regression was performed to examine whether decisions about prosecution, decisions about the arrest, decision maker as well as ratings of trustworthiness, ambiguous and total evidence predicted decisions about whether the additional investigation was necessary. The categorical variables were dummy coded using the following codes, for decision about arrest, 1 = arrest, 0 = no arrest and for decision maker, 1 = self, 0 = colleague for decision about prosecution, 1 = prosecute, 0 = no prosecution. Using the Enter method, non-significant predictors were excluded; see Table A2 in Appendix. When rerunning the model without the non-significant predictors, the results in Table 3 were obtained.

This indicates that prosecutors' decisions about whether the additional investigation was necessary were significantly predicted by the decisions about arrest and prosecution. When the defendant had previously been arrested, the odds that prosecutors thought that additional investigation was necessary was 0.54 times lower. Furthermore, when the defendant had been prosecuted, the odds that prosecutors thought the additional investigation was necessary was 0.13 times lower. This suggests that the decisions to arrest and prosecute constituted a stop rule for acquiring more information.

Level of guilt confirmation

For the scenarios in which participants stated that additional investigation was necessary, a multiple regression analyses was performed to examine whether decisions about arrest,

Table 3. Predictors of decisions about the necessity of additional investigation in experiment 1 using logistic regression analyses (Enter method).

Variable	Decisions about additional investigation			
	<i>b</i> 95% CI	95% CI for odds ratio		
		Lower	Odds	Upper
Constant	-.44 [-1.49, 0.59]			
Arrest	-.62* [-1.18, -0.08]	0.31	0.54	0.93
Prosecute	-2.08** [-3.04, -1.31]	0.05	0.13	0.30

Note: Model $\chi^2(3) = 31.26, p < .001. R^2 = .43$ (Hosmer & Lemeshow).

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 4. Predictors of level of guilt confirmation in additional investigation in Experiment 1 using hierarchical multiple regression analyses.

Variable	Level of guilt confirmation (1–7)	
	ΔR^2	<i>b</i> 95% CI
Constant		2.08 *** [1.50, 2.67]
Step 1	.001	–.37 [–1.00, 0.27]
Arrest		
Step 2	–.002	–.19 [–.83, 0.44]
Decision maker		
Step 3	.74***	4.16 *** [3.82, 4.49]
Prosecution		
Step 4	.740	.12 [–0.033, 0.28]
Total evidence		
Step 5	.738	.014 [–0.11, 0.13]
Trustworthiness		

Note: For each step, only variables added in relation to the previous step are stated. $p < .05$. ** $p < .01$. *** $p < .001$. Model 3, $F = 198.03$ $\chi^2 p < .001$.

decision maker, decisions about prosecution, trustworthiness ratings and total evidence strength significantly predicted the level of guilt confirmation in additional investigation (1–7). The categorical variables were dummy coded using the following codes, for decision about arrest, 1 = arrest, 0 = no arrest and for decision maker, 1 = self, 0 = colleague for decision about prosecution, 1 = prosecute, 0 = no prosecution. Table 4 displays the results.

Table 4 suggests that the decision about prosecution was the only variable which significantly predicted the level of guilt confirmation in additional investigation. When prosecutors had decided to prosecute, this increased the level of guilt confirmation by 4.16 units. Thus, the overall analysis regarding additional investigation indicates that prosecutors more rarely thought that additional investigation was necessary after an arrest and a prosecution. Furthermore, the investigation they did suggest after a prosecution was to a larger extent aimed at confirming the defendant's guilt.

Discussion experiment 1

The findings in Experiment 1 were partly expected and partly unexpected. The unexpected findings were primarily found in Stage (1) namely the null findings for the ambiguous evidence and that the opposite of the predicted was found for the total evidence, since prosecutors rated the evidence as stronger in relation to non-arrested than arrested suspects, $r = .49$ (although the lower trustworthiness ratings of arrested than non-arrested suspects, $r = .74$ were expected). As indicated by the above analysis, it is unlikely that the unexpected findings were because prosecutors realized what the purpose of the study was or because of carry-over effects (see Appendix for these analyses). Another possibility we have evaluated (see Appendix) is that factors such as the suspect's gender, ethnicity or criminal record influenced the results. If this was the case, a different response pattern would be expected for own decisions and the colleague's decisions (which were experimentally controlled). As illustrated by Figure 2, the difference between the arrest and no arrest conditions was, in fact, larger (although not significantly larger) for the colleague's decisions. This speaks against an interpretation in terms of self-selection effects. A possible interpretation of the reversed effect in relation to the total evidence

could be because an arrest makes prosecutors more attentive to weaknesses (such as ambiguity) in the evidence and they therefore perceive of the total evidence as weaker. Possibly, they responded to the diagnosticity of the evidence and not the posterior odds of guilt, also referred to as the inverse fallacy (Villejoubert & Mandel, 2002). Although this is irrational using probabilistic definitions of rationality, it should be noted that primarily focusing on the implications of the evidence is what the law demands from prosecutors. Thus, it appears prosecutors were skeptical in relation to previous decisions and at this early stage they can therefore be described as devil's advocates. In fact, their skepticism was even stronger than what the law demands since if they were simply neutral neither arrests nor non-arrests would have an effect. This can indicate overcompensation, that is, after the arrest prosecutors soon have to decide whether there are sufficient reasons to prosecute and if they do prosecute, the evidence against the defendant has to be sound. Since the prosecutor carries the burden of proof, a strong skepticism towards the available evidence could be part of a pretrial strategy whereby prosecutors ensure that any charges brought before the Court are well founded. Such an interpretation is in line with the prosecutors' legal obligation to not initiate unjustified prosecutions.

However, in Stages (2) and (3) prosecutors responses were more in line with the expectations and, especially for Stage 3, the results signal a more closed mindset. In Stage (2) the odds of a prosecution were significantly higher if the suspect had been arrested and was rated as low in trustworthiness (whereas for non-arrests, the trustworthiness rating did not matter). Also, the odds of a prosecution increased as the perceived evidence strength increased. This would imply two things. Firstly, it seems that the primary link between the arrest and prosecution is the suspect's trustworthiness, not how the evidence is perceived (since the evidence was rated as weaker in relation to arrested suspects but the odds of a prosecution increased when the perceived evidence strength increased). Secondly, it seems like arrests have the opposite effects on prosecutors' evaluation of the evidence and their decision about prosecution. More specifically, if an arrested suspect's trustworthiness is low, the odds of a prosecution increase. However, the arrest does not make prosecutors perceive of the evidence as stronger (but in fact as weaker), whereas the odds of prosecution increase with higher perceived evidence strength. Taken together, these findings suggest that the decisions are partially linked but also partially disconnected. Possibly, this is a result of prosecutors' (consciously or subconsciously) adjusting their ratings of the evidence to not be biased by an arrest, but the bias instead seeps out in lower trustworthiness ratings of the defendant (which is associated with higher odds for prosecution). In Stage (3) prosecutors were less likely to deem additional investigation necessary after a prosecution, and even less so if there was a previous decision about the arrest. Furthermore, after initiating a prosecution the suggested additional investigation had higher levels of guilt confirmation. Hence, it seems like the primary trigger of confirmation bias was the prosecution (not the arrest). However, it should be noted that *after* a prosecution, a previous arrest of the suspect added to the more closed mindset (given the lower odds of initiating additional investigation). The onset of confirmation bias was therefore later than expected. Also opposing the hypotheses, no significant differences were found between own and the colleague's decisions to arrest in none of the stages.

Thus, overall it seems that in Stage (1) prosecutors acted as their own devil's advocate, possibly for strategical reasons (such as being able to justify an eventual prosecution), in

Stage (2) prosecutors were relatively balanced and rational (since the factor most strongly associated with the prosecution was the total evidence), whereas in Stage (3) they had more guilt-confirming mindsets, possibly also for strategic reasons (such as being able to prove the defendant's guilt in Court). If this interpretation is correct, the question is how prosecutors managed to postpone the onset of confirmation bias. One possibility is that they have acquired such a skill to reason independently of a previous hypothesis from their working experience. If this is true then it would be expected that decision makers without such training such as Law students would not have the same capacity. If the capacity is related to the legal education (rather than experience from working as a prosecutor), Law students would also display it, but Psychology students would not. To test this, Experiment 2 uses the same method with Law and Psychology students.

Experiment 2

Method

Participants

Study participants were 60 students (30 Law and 30 Psychology students) from the Faculty of Law and the Department of Psychology, Uppsala University. Of the students, 18 were male and 42 female and their mean age was 26.57 years ($SD = 8.31$). They responded to study announcements that were displayed on the Faculty/Department's premises.

Design

Experiment 2 used the same design as Experiment 1 for all three stages but with the addition of the Education variable. Thus, in Stage 1, Experiment 2 had a 2 (education: Law vs. Psychology) $\times 2$ (decision: arrest vs. no arrest) $\times 2$ (decision maker: self vs. colleague) Mixed subjects design. Education was a between-subjects factor whereas decision and decision maker were within-subjects factors. For Stages 2 and 3, Education was added as a predictor variable for decisions about prosecution, additional investigation and level of guilt confirmation in the additional investigation.

Material and procedure

The material and the procedure were the same as in Experiment 1.

Results

Descriptive statistics (M , SD) and post hoc comparisons for significant effects are displayed in Table 5.

Trustworthiness ratings

A $2 \times 2 \times 2$ mixed ANOVA with Education (Law vs. Psychology) as a between-subjects factor and decision (arrest vs. no arrest) as well as decision maker (self vs. colleague) as within-subject factors was conducted. There was a significant effect of decision on the trustworthiness ratings, $F(1,57) = 17.04$, $p < .001$. Overall, arrested suspects were rated as less trustworthy than non arrested suspects and the size of this effect was $r = .48$. There was

Table 5. Means, SD and *Post hoc* comparisons for significant effects by dependent variable, decision and decision maker in Experiment 2.

DM	A		D NA		Total		Post hoc		
	M	SD	M	SD	M	SD			
<i>Trustworthiness ratings (1–7)</i>									
S	3.29	1.15	4.41	0.92	3.85	1.04	$p < .001$ S/A vs. S/NA		
C	3.40	1.12	3.74	1.24	3.57	1.18	$p < .001$ S/NA vs. C/A		
Total	3.35	1.14	4.08	1.08	3.71	1.11			
<i>Ambiguous evidence (1–7)</i>									
S	4.14	1.69	3.28	1.70	3.71	1.70	$p = .012$ for S/A vs. S/NA		
C	3.90	1.57	3.83	1.66	3.87	1.09			
Total	4.02	1.63	3.56	1.68	3.79	1.53			
<i>Total evidence strength (1–7)</i>									
S	4.86	1.35	3.65	1.70	5.21	1.22	$p < .001$ for S/A vs. S/NA		
C	4.41	1.55	4.61	1.53	4.81	1.67	$p = .004$ for S/NA vs. C/A		
Total	4.64	1.45	4.13	1.62	4.70	1.49	$p < .001$ for S/NA vs. C/NA		
<i>Prosecution (and no prosecution) rates %</i>									
S	71.30 (28.70)		35.80 (64.20)		54.00 (46.00)				
C	58.10 (41.90)		67.70 (32.30)		43.45 (56.55)				
Total	64.70 (35.30)		51.75 (48.25)		53.48 (51.28)				
<i>Additional (and no additional) investigation rates %</i>									
			L 48.10 (51.90)		E 38.20 (61.80)				
			P 41.10 (58.90)		NP 45.70 (54.30)				
<i>Guilt confirmation in additional investigation (1–7)</i>									
			L 4.15 (2.39)		E 4.29 (2.37)				
			P 4.94 (1.37)		NP 3.68 (1.41)				
<i>Guilt confirmation in decision (1–7)</i>									
		A		NA		D		Total	
		M	SD	M	SD	M	SD	M	SD
S	5.55	1.31	4.42	1.09	3.92	1.47	3.24	1.13	4.28
C	4.23	1.63	5.56	.69	3.94	1.24	3.64	1.67	4.34
Total	4.89	1.47	4.99	.89	3.93	1.36	3.44	1.40	4.31

Note: D = Decision; A = Arrest, NA = No arrest; DM = Decision maker; S = Self; C = Colleague; DP = Decision about Prosecution, P = Prosecution, NP = No prosecution; E = Education, L = Law, P = Psychology.

also a significant interaction effect between decision and decision maker, $F(1,57) = 7.46$, $p = .008$, $r = .34$. Post hoc tests using Bonferroni revealed significant differences between when participants themselves had decided to arrest and not arrest, $p < .001$. As is illustrated by Figure 3, participants considered suspects they themselves had arrested as less trustworthy than non-arrested suspects. Such a difference was not present for the colleague's decisions, suggesting an increased bias in relation to own decisions. When comparing participants' own decisions to not arrest and the colleague's decision to arrest, a significant

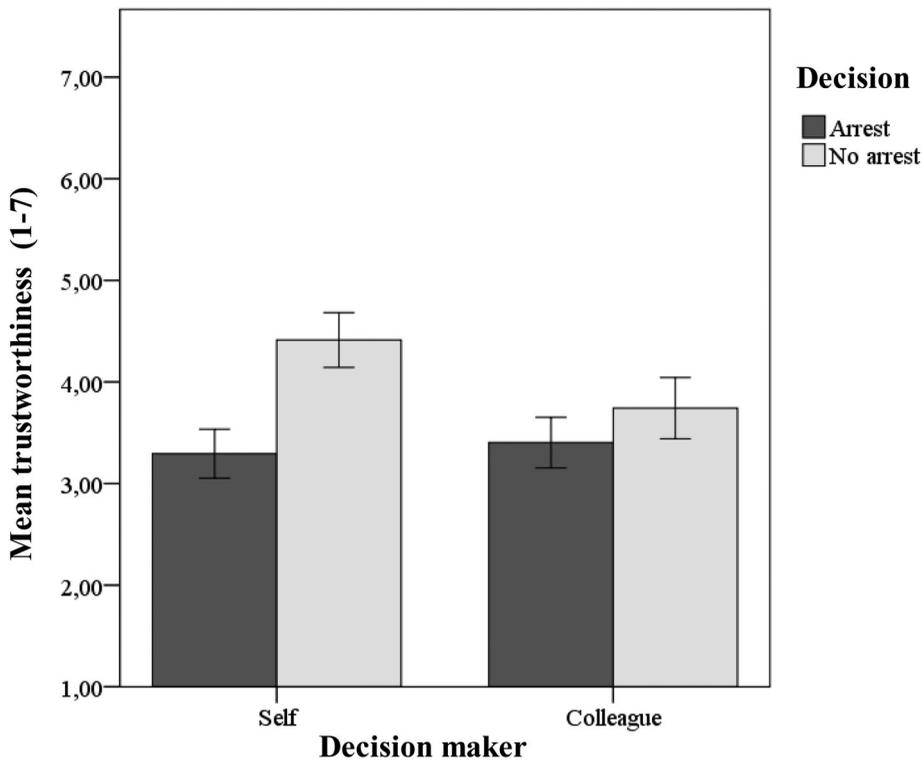


Figure 3. Mean trustworthiness ratings by decision and decision maker in Experiment 2. Error bars represent 95% CI.

difference was also found, $p < .001$, which suggests that the colleague's arrests also resulted in decision consistent ratings of trustworthiness and that there was at least some reliance on the colleague's decision.

There were no significant effects of education, $F(1,57) = 2.67$, $p = .423$ or decision maker, $F(1,57) = 2.38$, $p = .129$ and no significant interaction effects between decision and education, $F(1,57) = 0.498$, $p = .483$, decision maker and education, $F(1,57) = 0.13$, $p = .910$ or decision, decision maker and education, $F(1,57) = 0.299$, $p = .586$.

Ambiguous evidence

A $2 \times 2 \times 2$ mixed ANOVA with Education (Law vs. Psychology) as a between-subjects factor and decision (arrest vs. no arrest) as well as decision maker (self vs. colleague) as within-subject factors was conducted. There was a significant effect of decision on the ratings of the ambiguous evidence, $F(1,39) = 4.51$, $p < .001$. Overall, the ambiguous evidence was perceived as more indicative of guilt after an arrest and the size of this effect was $r = .32$. There was also a significant interaction effect between decision and decision maker, $F(1,39) = 4.79$, $p = .035$, $r = .33$. Post hoc tests using Bonferroni revealed significant differences between participants own arrests and non arrests, $p = .012$. As illustrated by [Figure 4](#), for own decisions, the perception of the ambiguous evidence was influenced by the arrest, with significantly higher perceived evidence strength for arrests than non-arrests, whereas no difference was observed for the colleague's decision.

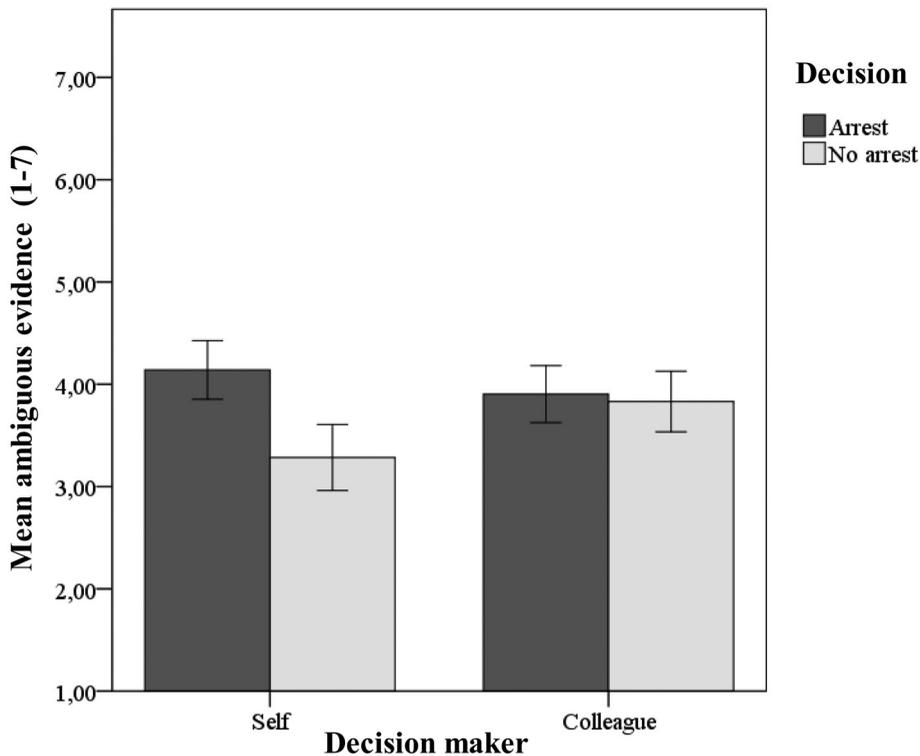


Figure 4. Mean ratings of ambiguous evidence by decision and decision maker in Experiment 2. Error bars represent 95% CI.

There was no significant effect of education, $F(1,39) = .31, p = .235$, or decision maker, $F(1,39) = 1.30, p = .262$ and no significant interaction effects between decision and education, $F(1,39) = 0.001, p = .977$, decision maker and education, $F(1,39) = 3.91, p = .055$ or decision, decision maker and student group, $F(1,39) = 0.321, p = .574$.

Total evidence

A $2 \times 2 \times 2$ mixed ANOVA with Education (Law vs. Psychology) as a between-subjects factor and decision (arrest vs. no arrest) as well as decision maker (self vs. colleague) as within-subject factors was conducted. There was a significant effect of decision, $F(1,38) = 9.13, p = .004$ and the size of this effect was $r = .44$ as well as decision maker, $F(1,38) = 4.31, p = .045, r = .32$ on the ratings of the total evidence strength. There was also a significant interaction effect between decision and decision maker, $F(1,38) = 20.98, p < .001, r = .60$. Post hoc tests using Bonferroni revealed significant differences between own arrests and own non arrests, $p < .001$. As Figure 5 illustrates, the total evidence was perceived as stronger when participants themselves had decided to arrest compared to when they decided to not arrest. No such difference was observed for the colleague condition. However, own non arrests resulted in significantly lower ratings of the total evidence strength compared to when a colleague had arrested the suspect, $p = .004$ as well as when the colleague did not arrest the suspect, $p < .001$.

There was no significant effect of education, $F(1,38) = 2.92, p = .332$ and no significant interaction effects between decision and student group, $F(1,38) = .101, p = .752$, decision

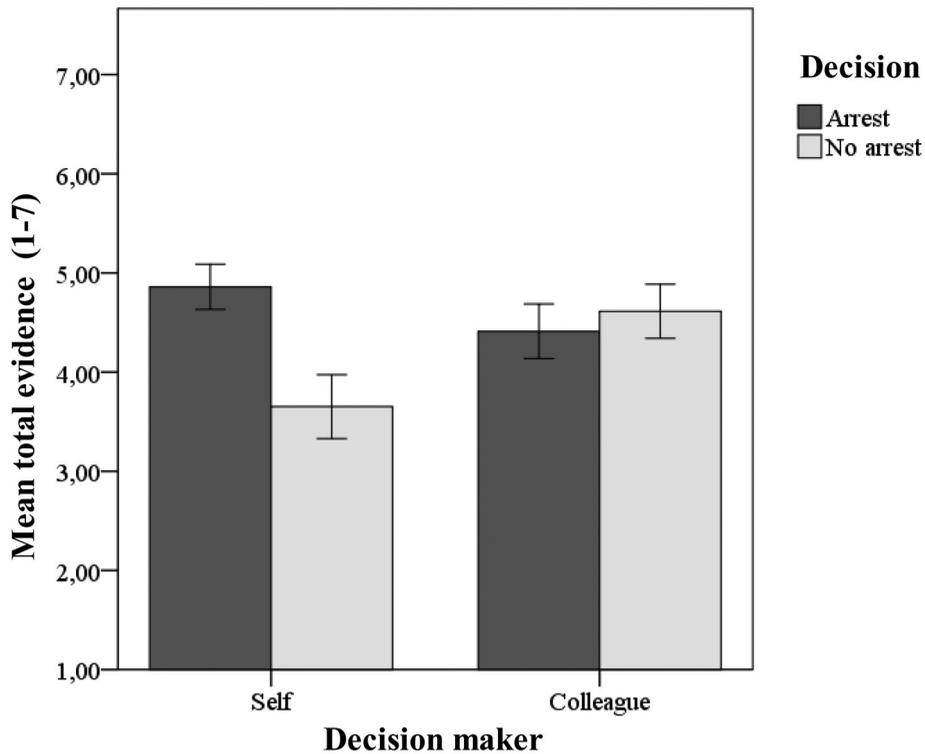


Figure 5. Mean ratings of total evidence by decision and decision maker in Experiment 2. Error bars represent 95% CI.

maker and student group, $F(1,38) = 3.27$, $p = .078$ or decision, decision maker and student group, $F(1,38) = 0.449$, $p = .507$.

Thus, overall both in the student and prosecutor samples, arrested suspects were rated as less trustworthy than non-arrested suspects ($r = .48$ for students and $r = .74$ for prosecutors). Apart from that, their responses in Stage 1 (before the prosecution) differed as students' answers for all measures were consistent with the hypotheses whereas prosecutors' answers were inconsistent with the hypotheses. More specifically, the interaction between decision and decision maker was displayed across all measures for students whereas it was absent in the prosecutors' responses. Also, for the ratings of the total evidence students and prosecutors show reversed effects, as students rated the total evidence as stronger in relation to arrested suspects whereas prosecutors rated it as stronger in relation to non-arrested suspects. The students' education did not significantly impact any of the ratings.

Prosecution

A logistic regression was performed to examine whether education, decisions about arrest, decision maker as well as ratings of trustworthiness, ambiguous and total evidence predicted decisions about prosecution. The categorical variables were dummy coded using the following codes, for education, 1 = Law, 0 = Psychology, for decision, 1 = arrest, 0 = no arrest and for decision maker, 1 = self, 0 = colleague. Using the Enter method, non-

Table 6. Predictors of decisions about prosecution in Experiment 2 using logistic regression analyses (enter method).

Variable	Decisions about prosecution			
	<i>b</i> 95% CI	Lower	Odds	Upper
Constant	0.83** [.28, 1.33]			
Arrest	0.73** [0.25, 1.29]	1.31	2.07	3.28
Arrest x Decision maker	-.61* [-1.18, -0.09]	0.32	0.55	0.92
Trustworthiness	-.18** [-0.30, -0.07]	0.74	0.83	0.94
Total evidence	1.37*** [1.13, 1.70]	0.90	1.10	1.34

Note: Model $\chi^2(4) = 274.75$, $p < .001$. $R^2 = .74$ (Hosmer & Lemeshow).

* $p < .05$. ** $p < .01$. *** $p < .001$.

significant predictors were excluded; see Table A3 in Appendix. When rerunning the model without the non-significant predictors, the results in Table 6 were obtained.

Both for prosecutors and students, the odds of a prosecution increased after a decision about arrest (although for prosecutors this was only in combination with low ratings of trustworthiness). In the student sample, the arrest was the predictor most closely associated with prosecution (Odds Ratio = 2.07) whereas for prosecutors the total evidence strength was the predictor most closely related to prosecution (Odds Ratio = 6.92), although both predictors were significant in both samples.

Additional investigation

Additional investigation rate

A logistic regression was performed to examine whether education, decisions about the arrest, decision maker, decisions about prosecution as well as ratings of trustworthiness, ambiguous and total evidence predicted decisions about whether the additional investigation was necessary. The categorical variables were dummy coded using the following codes, for education 1 = Law, 0 = Psychology, decision about arrest, 1 = arrest, 0 = no arrest and for decision maker, 1 = self, 0 = colleague for decision about prosecution, 1 = prosecute, 0 = no prosecution. Using the Enter method, non-significant predictors were excluded; see Table A4 in Appendix. When rerunning the model without the non-significant predictors, the results in Table 7 were obtained.

Similar to the prosecutors, students' decisions about whether the additional investigation was necessary were significantly predicted by the decision about prosecution. The odds that students thought the additional investigation was necessary was 1.84 times lower after a decision to prosecute (compare to 0.13 for prosecutors). Unlike the prosecutors, students' decisions were not significantly predicted by an arrest but instead an interaction between the prosecution decision and the decision maker variable was found. If students had themselves decided about the arrest (regardless of whether the decision was to arrest or not), they were somewhat less likely (0.44) to deem additional investigation necessary. Yet, in both experiments, the main predictor was the decision about prosecution.

Table 7. Predictors of decisions about the necessity of additional investigation in experiment 2 using logistic regression analyses (enter method).

Variable	Decisions about additional investigation			
	<i>b</i> 95% CI	Lower	Odds	Upper
Constant	-0.49** [-0.83, -.19]			
Prosecute	-0.52** [-0.14, -.78]	1.09	1.84	3.10
Prosecute x Decision maker	-0.81* [-1.61, -0.04]	0.21	0.44	0.92

Note: Model $\chi^2(3) = 5.85, p < .05. R^2 = .61$ (Hosmer & Lemeshow).

* $p < .05$. ** $p < .01$. *** $p < .001$.

Level of guilt confirmation

For the scenarios in which participants stated that additional investigation was necessary, a multiple regression analyses was performed to examine whether education, decisions about arrest, decision maker, decisions about prosecution as well as ratings of trustworthiness, ambiguous and total evidence significantly predicted the level of guilt confirmation in the suggested additional investigation (1–7). The variables were dummy coded using the following codes, for education 1 = Law, 0 = Psychology, decision about arrest, 1 = arrest, 0 = no arrest and for decision maker, 1 = self, 0 = colleague for decision about prosecution, 1 = prosecute, 0 = no prosecution. Table 8 displays the results.

Just like for the prosecutors, students' decisions to prosecute were associated with higher levels of guilt confirmation in the additional investigation. When students had decided to prosecute, the level of guilt confirmation increased by 2.67 units (compare to 4.16 for the prosecutors). Unlike with the prosecutors, students' level of guilt confirmation was also significantly higher when the suspect had previously been arrested (1.85) and if they themselves had decided to arrest (1.35). However, for both prosecutors and students, the decision to prosecute was the predictor most strongly associated with the level of guilt confirmation in additional investigation.

Table 8. Predictors of level of guilt confirmation in additional investigation in Experiment 2 using hierarchical multiple regression analyses.

Variable	ΔR^2	<i>b</i> 95% CI
Constant		3.92 *** [3.14, 4.70]
Step 1	.16***	1.85*** [1.29, 2.42]
Arrest		
Step 2	.24***	1.35*** [1.92, 0.79]
Decision maker		
Step 3	.54***	2.67 *** [2.22, 3.13]
Prosecution		
Step 4	.55	.26 [0.06, 0.46]
Total evidence		
Step 5	.55	.008 [-0.14, 0.15]
Trustworthiness		

Note: For each step, only variables added in relation to the previous step are stated. $p < .05$. *** $p < .01$.

*** $p < .001$. Model 3, $F = 82.24, p < .001$.

Overall, the analyses for decisions about prosecution as well as decisions relating to the additional investigation (after the decision about prosecution) indicates similar trends as for the prosecutors. Both students' and prosecutors' decisions about prosecution were significantly predicted by the arrest, the suspect's trustworthiness and the total evidence strength (separately or in interaction). Similarly, both the arrest and prosecution significantly predicted both students' and prosecutors' decisions about the additional investigation. The students' education did not significantly predict any of these decisions.

Smallest detectable effect sizes in Experiments 1 and 2

Although within-subjects designs such as the design employed here require far fewer participants than between-subjects designs (Thompson & Campbell, 2004), we conducted sensitivity power analyses indicating the smallest effect size that the design was powered to detect using G* Power. These analyses indicated that the smallest detectable effect size in Experiment 1 ($N=40$) was .15 and in Experiment 2 ($N=60$) the smallest detectable effect size was .12, when α was set to 0.05. This can be related to that 0.30 is often considered to be the smallest effect size of interest as any smaller effect sizes are unlikely to be of practical relevance (Lakens, 2014). Clearly, which more specific effect sizes are or are not of interest depend on what the effects concern. When it comes to biasing or debiasing factors in prosecutorial decision making, even relatively small effect sizes are of potential importance, although probably not as small effects as .15 or .12.

As regards the null effects for the decision maker variable in Experiment 1, the obtained smallest detectable effect size should also be related to the proportionally much larger effects of the independent variables, e.g. the main effect of the decision to arrest on prosecutors' trustworthiness ratings, which had the effect size $r=.74$. As such, even if the study would have failed to detect any effect sizes smaller than .15 for the intended debiasing technique, it is unlikely that a significant finding would have represented a practically relevant effect.

Furthermore, this analysis indicates that the smallest detectable effect sizes were quite similar for Experiments 1 and 2, despite that Experiment 2 had somewhat higher power due to the larger sample size. This, in turn, suggests that the partially different findings for Experiment 1 and 2 are not due to differences in power. As such, the null findings for both decision and decision-maker in relation to the ambiguous evidence as well as the null findings for the decision maker variable for the other measures, would probably have remained the same even if more prosecutors would have participated in Experiment 1. Furthermore, if the differences in sample size, despite this, explain the differences, any effects that we failed to detect in Experiment 1, both for the decision and decision maker variable are unlikely to be of any practical relevance as they would be smaller than .15.

General discussion

The results differed both between the three stages of the experiments and between the prosecutor and student samples. Before the prosecution (1) prosecutors' responses were largely inconsistent with the predictions whereas students' responded consistently in line with the predictions. The only variable in this stage regarding which prosecutors' responses were expected was the trustworthiness variable, since prosecutors rated arrested suspects as less trustworthy. It can be questioned whether this is really an irrational bias, since overall,

arrested suspects are less likely to be trustworthy (for instance because they lie in order to be set free) than non-arrested suspects. Thus, using probabilistic definitions of rationality this behavior appears to be rational but of course, that does not render it acceptable from a legal perspective. Since the arrest in itself has no evidentiary value the suspect should not be considered less trustworthy on that basis. For Stage (2) the prosecution and Stage (3) after the prosecution, prosecutors' and students' decisions were largely predicted by the same variables and were both in line with the predictions. A comparison of prosecutors' and students' responses in the different stages suggests that prosecutors but not students were able to delay the onset of confirmation bias until the prosecution. These findings suggest that previous arrests are associated with higher odds of prosecution. Furthermore, both the arrest and the prosecution lower the odds that additional investigation is deemed necessary and/or increase the level of guilt confirmation in the suggested additional investigation. This is more clearly related to confirmation bias (than the lower trustworthiness ratings in Stage (1), since it suggests that even when the evidence is kept constant, there is still an association between a previous guilt hypothesis (as manifested by the arrest) and the prosecution. Also, it suggests that the arrest and prosecution provided stop rules for the search for information, as expected in decision makers who are confined by their hypotheses. If the original hypothesis is still perceived as the only viable possibility, there is no need to search for additional information, unless of course, the additional information can confirm the original hypothesis. Since there were no significant differences between Law and Psychology students' decisions in neither of the stages, it is unlikely that legal education provides an explanation for the delayed onset of confirmation bias (which both groups of students displayed consistently in all stages). Yet, neither the Law nor the Psychology students have experience from working as prosecutors, which therefore could be an explanation for the null as well as contrary findings in the sample of prosecutors.

Limitations

Although the experimental method was the most appropriate method for the research questions of interest, one of the variables, namely the decisions about the arrest was (in part) not experimentally controlled. This was motivated by the interest in how formulating own decisions influence prosecutors' decision making but opened up for self-selection effects. However, the difference between the arrest and no arrest conditions was, in fact, larger (although not significantly larger) for the colleague's decisions (that were experimentally controlled), which speaks against an interpretation in terms of self-selection effects. Another potential confound that has also been evaluated and for which we found no support is carry-over effects (which could otherwise have explained the null or contrary findings in Experiment (1). Furthermore, although the sample of prosecutors was relatively small, the sensitivity power analysis suggests that the differences between Experiment 1 and 2 are not likely to be due to differences in power. A reasonable interpretation is therefore that the prosecutors were capable of acting as their own devil's advocates in this early stage. It is of course unknown whether there are systematic differences between the prosecutors that participated and those who refrained from doing so. Since the prosecutors were told that the study was about prosecutorial decision making, it is possible that prosecutors who perceive of themselves as particularly objective decision makers agreed to participate whereas other prosecutors did not. However, since

confirmation bias is a more or less subconscious process, also prosecutors who perceive of themselves as an objective may be biased. There is no apparent reason to believe that this subconscious bias would be stronger or weaker in participating than non-participating prosecutors, although the possibility should not be disregarded. In future research, the external validity could be increased by for instance having prosecutors interact with real parties and witnesses and also examine the importance of different kinds of relationships for whether and to what extent prosecutors are critical of their colleagues' assessments.

Implications and conclusion

This research has four primary implications. Firstly, prosecutors' ability to be skeptical of their previous arrests (as well as their colleagues') need to be further examined. If it is correct that prosecutors manage to be their own devil's advocate before the prosecution (Stage 1), it should be possible to replicate these findings and then try to better understand them. Secondly, provided that a devil's advocate mechanism is at play in prosecutors' decision making, it should be examined how such a mindset can be better incorporated into the decision about prosecution (Stage 2) as well as decisions and judgments made after the prosecution (Stage 3). In real life criminal cases, a skeptical mindset in relation to the evidence before a prosecution is clearly of limited value if there is still bias in the subsequent more influential decisions about whether to press charges and how to allocate investigative resources thereafter. The different results in these stages also suggest that care should be taken in increasing the connectedness between the evaluation of evidence before a prosecution and the prosecution decision, as the standard of proof requires that the latter decision is firmly anchored in the first assessment. However, the standard of proof does not provide any exact cut off point over or under which prosecutors must either prosecute or refrain from prosecuting. Coming to better understand what happens in the process when prosecutors translate the evidence into dichotomous decisions to either prosecute or not could be a first step in finding functioning debiasing techniques. Thirdly, since the prosecution triggered a relatively strong confirmation bias and a previous arrest added to this effect, it seems like the prosecution made prosecutors' shift from devil's advocates to crime fighters (although the arrest had a limited impact, moderated by the suspect's trustworthiness already when prosecutors decided about whether to press charges). Even though prosecutors become the defendant's counter party after having pressed charges, prosecutors are still expected by Swedish law to remain objective in relation to the question of the defendant's guilt, which is clearly not fulfilled if all or most investigation was undertaken after a prosecution aims to confirm a suspect's guilt. Furthermore, it increases the risk that the evidence presented at Court only represents a limited picture of reality, increasing the risk of a wrongful conviction. Fourthly, since the decision maker variable did not significantly influence any of the prosecutors' judgments, it seems like changing decision maker between decisions about arrest and prosecution is an insufficient debiasing technique. However, it should be noted that this debiasing technique was introduced in a stage where, contrary to the hypothesis, no evidence of confirmation bias was found. This implies that future research in this area should focus primarily on later stages (although the search component still needs further investigation also prior to the prosecution) as this, probably, is where the strongest bias is found and therefore also the stage in which debiasing efforts are likely

to be most needed. The null findings for the decision maker variable could be an indication that bias in prosecutors' does not primarily have social explanations, although this is at least partially contradicted by that the bias appeared with the prosecution decision, which is also when prosecutors assumed their social roles as parties in the proceedings against the defendant. Thus, the social explanations need further evaluation just like the role of for instance cognitive components (for instance when it comes to increasing connectedness between the evaluation of evidence and decisions about prosecution). Yet, it is important to note that a change in prosecutors' mindset may not only be a product of when (at what stage) but also what, that is, what task is conducted. In Stages 1 and 2 where no strong evidence of bias was found, prosecutors evaluated, but did not search for, evidence. This is different from Stage 3 in which the task was instead to decide whether it was necessary to search for more information and if yes, what information. Although evaluation is a pronounced element of prosecutorial decision making, not the least prior to a charging decision, it is possible that a bias prior to the charging decision had been found if the study instead had examined the search component in the early stages. This also means that a change of decision maker could be useful as a debiasing technique when it comes to searches for information, even if it was not for the evaluation of evidence. Thus, the social explanations need further evaluation just like the role of for instance cognitive components (for instance when it comes to increasing connectedness between the evaluation of evidence and decisions about prosecution). This ought to be a strong incentive for future research as better understanding and mitigating bias after a prosecution is directly connected to increasing the accuracy of the criminal proceedings and thereby also increasing the legitimacy of the legal system as a whole.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- Abadinsky, H. (1998). *Law and justice. An introduction to the American legal system* (4th ed.). Chicago: Nelson-Hall.
- Babcock, L., & Loewenstein, G. (1997). Explaining bargaining impasse: The role of self-serving biases. *The Journal of Economic Perspectives*, 11, 109–126. doi:10.1257/jep.11.1.109
- Bilalić, M., McLeod, P., & Gobet, F. (2008). Why good thoughts block better ones: The mechanism of the pernicious einstellung (set) effect. *Cognition*, 108, 652–661. doi:10.1016/j.cognition.2008.05.005
- Burke, A. S. (2006). Improving prosecutorial decision making: Some lessons of cognitive science. *William and Mary Law Review*, 47, 1588–1631. Retrieved from <http://scholarship.law.wm.edu/wmlr/vol47/iss5/3>
- Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, 51, 1484–1525. doi: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(4), 349–354. doi:10.1037/h0047358
- Davis, A. J. (2001). The American prosecutor: Independence, power, and the threat of tyranny. *Iowa Law Review*, 86, 393–465. Retrieved from http://digitalcommons.wcl.american.edu/facsch_lawrev/156/
- Dickerson Moore, S. A. (2000). Questioning the autonomy of prosecutorial charging decisions: Recognizing the need to exercise discretion – knowing there will be consequences for crossing the line. *Louisiana Law Review*, 60, 371–404. Retrieved from <https://digitalcommons.law.lsu.edu/lalrev/vol60/iss2/1>

- Doherty, M. E., & Mynatt, C. R. (1990). Inattention to P (H) and to P ($D \sim H$): A converging operation. *Acta Psychologica*, 75, 1–11. doi:10.1016/0001-6918(90)90063-L
- Dunbar, K., & Klahr, D. (1989). Developmental differences in scientific discovery processes. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of herbert A. Simon* (pp. 109–143). Hillsdale, NJ: Erlbaum.
- Engel, C., & Glöckner, A. (2013). Role-Induced bias in court: An experimental analysis. *Journal of Behavioral Decision Making*, 26, 272–284. doi:10.1002/bdm.1761
- Englich, B., & Mussweiler, T. (2001). Sentencing under uncertainty: Anchoring effects in the courtroom. *Journal of Applied Social Psychology*, 31, 1535–1551. doi:10.1111/j.1559-1816.2001.tb02687.x
- Englich, B., Mussweiler, T., & Strack, F. (2005). The last word in court: A hidden disadvantage for the defense. *Law and Human Behavior*, 29, 705–722. doi:10.1007/s10979-005-8380-7
- Epelboim, J., & Suppes, P. (2001). A model of eye movements and visual working memory during problem solving in geometry. *Vision Research*, 41, 1561–1574. doi:10.1016/S0042-6989(00)00256-X
- Ernberg, E., Tidefors, I., & Landström, S. (2016). Prosecutors' reflections on sexually abused preschoolers and their ability to stand trial. *Child Abuse & Neglect*, 57, 21–29. doi:10.1016/j.chiabu.2016.06.001
- 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, 155–165. doi:10.1002/jip.1384
- Field, A. (2012). *Discovering statistics using IBM SPSS statistics* (4th ed.). London: SAGE.
- Flowers, R. K. (1998). What you see is what you get: Applying the appearance of impropriety standard to prosecutors. *Missouri Law Review*, 63, 699–734. Retrieved from <http://scholarship.law.missouri.edu/mlr/vol63/iss3/3>
- Freedman, E. M. (1990). Innocence, federalism, and the capital jury: Two legislative proposals for evaluating post-trial evidence of innocence in death penalty cases. *New York University Review of Law & Social Change*, 18, 315–323. Retrieved from http://scholarlycommons.law.hofstra.edu/faculty_scholarship/216
- Gershman, B. L. (1998). Mental culpability and prosecutorial misconduct. *American Journal of Criminal Law*, 26, 122–164. Retrieved from <http://digitalcommons.pace.edu/lawfaculty/129/>
- Gershman, B. L. (2001). The prosecutor's duty to truth. *Georgetown Journal of Legal Ethics*, 14, 309–354. Retrieved from <http://digitalcommons.pace.edu/lawfaculty/128/>
- Gilliéron, G. (2013). *Public prosecutors in the United States and Europe – A comparative analysis with special focus on Switzerland, France, and Germany*. New York, NY: Springer.
- Gramckow, H. (2008). Prosecutor organization and operations in the United States. In *Comparative research. Open society institute Sofia* (pp. 386–429). Sofia: Open Society Institute.
- Green, B. A. (2003). Prosecutorial ethics as usual. *University of Illinois Law Review*, 5, 1573–1604. Retrieved from <https://www.illinoislawreview.org/wp-content/ilr-content/articles/2003/5/Green.pdf>
- Green, B. A., & Roiphe, R. (2017). Rethinking prosecutors' conflicts of interest. *Boston College Law Review*, 58, 464–538. Retrieved from <https://ssrn.com/abstract=2849658>
- Greenwald, A. (1980). Psychology attends to its self. *PsycCRITIQUES*, 25(12), 970–972. Retrieved from <https://faculty.washington.edu/agg/pdf/Greenwald.Self&Memory.1981.pdf>
- Griffin, L. C. (2001). The Prudent Prosecutor. Scholarly works, paper 728. Retrieved from <http://scholars.law.unlv.edu/facpub/728>
- Haney, C., Banks, C., & Zimbardo, P. (1973). Study of prisoners and guards in a simulated prison. *Naval Research Reviews*, 26, 1–17. doi:10.1037/h0040525
- Harris, D. A. (2011). *The interaction and relationship between prosecutors and police officers in the u.s., and how this affects police reform efforts*. University of Pittsburgh Legal Studies Research Paper Series (Working Paper No. 2011-19).
- Haverkamp, B. E. (1993). Confirmatory bias in hypothesis testing for client-identified and counsel self-generated hypotheses. *Journal of Counseling Psychology*, 40, 303–315. doi:10.1037/0022-0167.40.3.303
- Hecht, H., & Proffitt, D. R. (1995). The price of expertise: Effects of experience on the water-level task. *Psychological Science*, 6, 90–95. doi:10.1111/j.1467-9280.1995.tb00312.x
- Heller, R. (1997). Selective prosecution and the federalization of criminal law: The need for meaningful judicial review of prosecutorial discretion. *University of Pennsylvania Law Review*, 145,

- 1309–1358. Retrieved from http://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=3489&context=penn_law_review
- Henning, P. J. (1999). Prosecutorial misconduct in grand jury investigations. *South Carolina Law Review*, 51, 1–61. Retrieved from <https://ssrn.com/abstract=1131763>
- Hetherington, A. M. (2002). Prosecutorial misconduct. *The Georgetown Law Journal*, 90, 3–32. Retrieved from https://moe.fd.org/Docs/fjc2010_Prosecutorial_Misconduct.pdf
- Heuman, L. (2004). Domarens och åklagarens skyldighet att vara objektiva. *Juridisk Tidskrift*, 5, 42–52. Retrieved from <http://www.jt.se/author.asp?uid=16>
- Hobbs, S. E. (1950). Prosecutor's bias, an occupational disease. *Alabama Law Review*, 40. Retrieved from http://heinonline.org/HOL/Page?handle=hein.journals/bamalr2&div=9&g_sent=1&casa_token=&collection=journals
- Jacobson, K. (2008). "We can't just Do It Any which Way" – objectivity work among Swedish prosecutors. *Qualitative Sociology Review*, 4, 46–68. Retrieved from <http://ezproxy.its.uu.se/login?url=https://search.proquest.com/docview/1002331402?accountid=14715>
- Jacoby, J. E. (1980). *The American prosecutor. A search for identity*. Lexington: Lexington Books.
- Jonakait, R. N. (1987). The ethical prosecutor's misconduct. *Criminal Law Bulletin*, 23, 550–567. Retrieved from <https://www.ncjrs.gov/App/Publications/abstract.aspx?ID=108090>
- 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.
- Klayman, J., & Ha, Y. W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, 94, 211–228. doi:10.1037/0033-295X.94.2.211
- Lakens, D. (2014). Performing high-powered studies efficiently with sequential analyses: Sequential analyses. *European Journal of Social Psychology*, 44, 701–710. doi:10.1002/ejsp.2023
- Lerner, J., & Tetlock, P. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, 125, 255–275. doi:10.1037/0033-2909.125.2.255
- Levinson, J. D., & Young, D. (2010). Different shades of bias: Skin tone, implicit racial bias, and judgments of ambiguous evidence. *West Virginia Law Review*, 112, 307–350.
- Lidén, M., Gräns, M., & Juslin, P. (2018). The presumption of guilt in suspect interrogations: Apprehension as a trigger of confirmation bias and debiasing techniques. *Law and Human Behavior*, 42, 336–354. doi:10.1037/lhb0000287
- Lievore, D. (2005). Prosecutorial decisions in adult sexual assault cases. *Trends & Issues in Crime and Criminal Justice*, 291, 1–6. Retrieved from http://www.aic.gov.au/media_library/archive/publications-2000s/prosecutorial-decisions-in-adult-sexual-assault-cases.pdf
- Lindberg, G. (1997). Om åklagareetik. *Svensk Juristtidning*, 1997, 197–221. Retrieved from <http://svjt.se/svjt/1997/197>
- McCloskey, J. (1989). Convicting the innocent. *Criminal Justice Ethics*, 8, 140–141. doi:10.1177/001112876901500105
- Mercier, H. (2016). The argumentative theory: Predictions and empirical evidence. *Trends in Cognitive Science*, 9, 689–700. doi: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. doi:10.1017/S0140525X10000968
- Munro, G. D., & Ditto, P. H. (1997). Biased assimilation, attitude polarization, and affect in reactions to stereotyped-relevant scientific information. *Personality and Social Psychology Bulletin*, 23, 636–653. doi:10.1177/0146167297236007
- Munro, G. D., & Stansbury, J. A. (2009). The dark side of self-affirmation: Confirmation bias and illusory correlation response to threatening information. *Personality and Social Psychology Bulletin*, 35, 1143–1153. doi:10.1177/0146167209337163
- 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. doi: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. doi:10.1016/0001-6918(91)90023-5

- The National Council for Crime Prevention, Persons suspected of Crimes, Final Statistics. (2015a). Retrieved from <https://www.bra.se/bra-in-english/home/crime-and-statistics/crime-statistics/reported-offences.html>
- The National Council for Crime Prevention, Reported Crimes, Final Statistics for. (2015b). Retrieved from <https://www.bra.se/bra-in-english/home/crime-and-statistics/crime-statistics/reported-offences.html>
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2, 175–220. doi:10.1037/1089-2680.2.2.175
- O'Brien, B. (2009). A recipe for bias: An empirical look at the interplay between institutional incentives and bounded rationality in prosecutorial decision making. *Missouri Law Review*, 74, 1000–1049. Retrieved from <https://ssrn.com/abstract=1410118>
- Prosecution Authority's Statistics. (2017). Retrieved from <https://www.aklagare.se/karriar/att-jobba-hos-oss1/personalstatistik1/>
- Pyszczynski, T., & Greenberg, J. (1987). Toward an integration of cognitive and motivational perspectives on social inferences: A biased hypothesis-testing model. In I. L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 297–340). San Diego, CA: Academic Press.
- Rajic, J., Wilson, D. E., & Pratt, J. (2015). Confirmation bias in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 41(5), 1353–1364. doi:10.1037/xhp0000090
- Ramsey, C. B. (2002). The discretionary power of “public” prosecutors in historical perspective. *American Criminal Law Review*, 39, 1309–1393. Retrieved from <http://journals.sagepub.com/doi/10.1177/1057567708319204>
- Robinson, J. P., Shaver, P. R., & Wrightman, L. S. (1991). *Measures of personality and social psychological attitudes*. Gulf Professional Publishing.
- Schunn, C. D., & Klahr, D. (1992). Complexity management in a discovery task. In J. K. Kruschke (Ed.), *Proceedings of the fourteenth annual meeting of the cognitive science society* (pp. 177–182). Mahway, NJ: Erlbaum.
- Schunn, C. D., & Klahr, D. (1993). Other vs. Self-generated hypotheses in scientific discovery. In W. Kintsch (Ed.), *Proceedings of the fifteenth annual meeting of the cognitive science society* (p. 900). Mahway, NJ: Erlbaum.
- SOU. (1988). Rapport av Parlamentariska kommissionen med anledning av mordet på Olof Palme: 18. The Swedish Government Bill, Prop. 2000/01:92.
- Tak, P. J. The Dutch Prosecutor (2012). In Luna, E. & Wade, M. The Prosecutor in Transnational Perspective. Oxford University Press. doi:10.1093/acprof:osobl/9780199844807.003.0011
- Thompson, V. A., & Campbell, J. I. D. (2004). A power struggle: Between vs. Within-subjects designs in deductive reasoning research. *Psychologia*, 47, 277–296.
- Thompson, L., & Loewenstein, G. (1992). Egocentric interpretations of fairness and interpersonal conflict. *Organizational Behavior and Human Decision Processes*, 51, 176–197. doi:10.1016/0749-5978(92)90010-5
- Villejoubert, G., & Mandel, D. R. (2002). The inverse fallacy: An account of deviations from Bayes's theorem and the additivity principle. *Memory & Cognition*, 30, 171–178. doi:10.3758/BF03195278
- Wagenaar, W. A., van Koppen, P. J., & Crombag, H. F. M. (1993). *Anchored narratives. The psychology of criminal evidence. Harvester wheatsheaf*. Harvester Wheatsheaf: St Martin's Press.
- Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content*. London: Batsford.
- Wastell, C., Weeks, N., Wearing, A., & Duncan, P. (2012). Identifying hypothesis confirmation behaviors in a simulated murder investigation: Implications for practice. *Journal of Investigative Psychology and Offender Profiling*, 9, 184–198. doi:10.1002/jip.1362
- Weigend, T. (1978). *Anklagepflicht und ermessens. Die stellung der staatsanwalts zwischen legalitäts- u. Opportunitätsprinzip nach deutschem und amerikanischem recht*. Baden-Baden: Nomos.
- Wettergren, Å., & Bergman Blix, S. (2016). Empathy and objectivity in the legal procedure: The case of Swedish prosecutors. *Journal of Scandinavian Studies in Criminology and Crime Prevention*, 17, 19–35. doi:10.1080/14043858.2015.1136501
- Yaroshefsky, E. (1999). Cooperation with federal prosecutors: Experiences of truth telling and embellishment. *Fordham Law Review*, 68, 917–964. Retrieved from http://fordhamlawreview.org/wp-content/uploads/assets/pdfs/Vol_68/Yaroshefsky_December.pdf

Appendix

Non-significant predictors of prosecution and additional investigation rates

Tables A1–A4 display the non-significant predictors for decisions about prosecution and additional investigation in Experiments 1 and 2.

Background variables in Experiment 1

The exploratory analyses examining whether participants' gender, age and years of experience were related to any of the dependent variables showed no such effects. More specifically, there were no significant differences between males and females ratings of the defendant's trustworthiness, ($M = 3.10$, $SD = 1.23$, 95% CI [2.45–4.12], & $M = 3.33$, $SD = .76$, 95% CI [2.78–4.34] respectively), $t(35) = -2.254$, $p = .131$, the ambiguous evidence, ($M = 3.88$, $SD = 1.03$, 95% CI [2.98–4.35], & $M = 3.74$, $SD = .56$, 95% CI [2.78–4.12] respectively), $t(35) = .588$, $p = .560$, or the total evidence strength, ($M = 4.75$, $SD = .43$, 95% CI [3.74–5.12], & $M = 4.62$, $SD = .81$, 95% CI [3.62–4.98] respectively), $t(35) = .741$, $p = .463$. Neither did male and female participants' decisions regarding prosecution, $\chi^2(1) = 5.077$, $p = .424$ or the character of their suggested additional investigation, ($M = 3.88$, $SD = 1.34$, 95% CI [3.10–4.54] & $M = 3.51$, $SD = 1.32$, 95% CI [2.87–4.14] respectively), $t(35) = .841$, $p = .406$, vary significantly. Furthermore, participants' ages were not significantly correlated with their ratings of the defendant's trustworthiness, $r = -.198$, $p = .239$, the ambiguous evidence, $r = -.129$, $p = .446$, the total evidence strength, $r = -.099$, $p = .562$ or the character of the suggested additional investigation, $r = -.033$, $p = .848$. Similarly, the number of years of experience was not significantly related to participants' ratings of the defendant's trustworthiness, $r = -.217$, $p = .198$, the ambiguous evidence, $r = -.123$, $p = .468$, the total evidence strength, $r = -.007$, $p = .965$ or the character of the suggested additional investigation, $r = -.086$, $p = .612$.

Controlling for potential self-selection effects in experiment 1

Since the arrest decision was not experimentally controlled but instead controlled by the prosecutors' own decisions (in the own decisions condition), this opens up to potential self-selection effects. Such self-selection might be due to differences between the scenarios when it comes to for instance the suspect's personal characteristics or that the evidence more clearly indicated the suspect's guilt, etc. However, if the findings were due to self-selection a different response pattern would be expected for own decisions and the colleague's decisions (which were experimentally controlled). As is illustrated by Figure 2, the difference between the arrest and no arrest conditions was in fact larger (although not significantly larger) for the colleague's decisions. This speaks against an interpretation in terms of self-selection effects.

Controlling for potential carry over effects in experiment 1

For the ratings of the ambiguous evidence as well as the total evidence strength, the findings were unexpected. To test whether these findings were due to carry over effects, between subjects analyses examining only the data for the conditions that were presented first were carried out. This allowed an examination of whether the results would be different from when using a Repeated measures design.

Using a 2 (decision about arrest: arrest vs. no arrest) \times 2 (decision maker: self vs. colleague) Between Subjects ANOVA no significant effects were found for decision, $F(1,37) = .565$, $p = .453$ or decision maker, $F(1,37) = .711$, $p = .400$ and no significant interaction effect either, $F(1,37) = .767$, $p = .382$ on the ratings of the ambiguous evidence. For the total evidence strength, a 2 (decision about arrest: arrest vs. no arrest) \times 2 (decision maker: self vs. colleague) Between Subjects ANOVA was also used. A significant effect of decision was found, $F(1,37) = 7.08$, $p = .008$. Overall, the total evidence was rated as stronger in relation to non-arrested than arrested suspects, $r = .40$. However, there was no significant effect of decision maker, $F(1,37) = .535$, $p = .465$ and no significant interaction, $F(1,37) = 3.70$,

$p = .065$. Thus, overall, these between-subjects analyses indicate similar trends as with the within-subjects analyses, which speaks against an interpretation in terms of carry-over effects.

Table A1. Non-significant predictors for prosecutors' decisions about prosecution using logistic regression (Enter method) in Experiment 1.

Variable	<i>b</i> 95% CI	Wald χ^2	<i>p</i>
Arrest	1.35 [-0.53, 3.45]	1.50	.220
Decision maker	-.087 [-1.42, 1.07]	0.022	.881
Arrest x Decision maker	0.85 [-0.66, 2.69]	1.27	.260
Trustworthiness	-0.15 [-0.45, 0.11]	1.13	.859
Ambiguous evidence	0.24 [-0.05, 0.55]	3.12	.087

Table A2. Non-significant predictors for prosecutors' decisions about additional investigation using logistic regression (Enter method) in Experiment 1.

Variable	<i>b</i> 95% CI	Wald χ^2	<i>p</i>
Decision maker	0.23 [-.52, 1.03]	.35	.552
Arrest x Decision maker	-.38 [-1.47, 0.85]	.45	.686
Prosecute x Decision maker	-0.71 [-1.31, -.02]	.13	.789
trustworthiness	-0.09 [-.31, 0.12]	.89	.343
Ambiguous evidence	0.17 [-0.042, 0.41]	2.70	.100
Total evidence	0.46 [0.19, 0.83]	2.62	.921

Table A3. Non-significant predictors for students' decisions about prosecution using logistic regression (Enter method) in Experiment 2.

Variable	<i>b</i> 95% CI	Wald χ^2	<i>p</i>
Ambiguous evidence	0.66 [0.53, 0.84]	1.62	.202

Table A4. Non-significant predictors for students' decisions about additional investigation using logistic regression (Enter method) in Experiment 2.

Variable	<i>b</i> 95% CI	Wald χ^2	<i>p</i>
Arrest	0.15 [-0.33, 0.67]	0.35	.560
Arrest x Decision maker	0.057 [-0.67, 0.79]	0.024	.878
Trustworthiness	-0.033 [-0.15, 0.08]	.34	.56
Ambiguous evidence	-0.10 [-0.22, 0.016]	2.96	.085
Total evidence	0.055 [-0.10, 0.24]	0.46	.494