Cognitive Biases in Criminal Case Evaluation: A Review of the Research

Vanessa Meterko¹ · Glinda Cooper¹

Accepted: 8 December 2020 / Published online: 23 June 2021
© The Author(s) 2021

Abstract

Psychological heuristics are an adaptive part of human cognition, helping us operate efficiently in a world full of complex stimuli. However, these mental shortcuts also have the potential to undermine the search for truth in a criminal investigation. We reviewed 30 social science research papers on cognitive biases in criminal case evaluations (i.e., integrating and drawing conclusions based on the totality of the evidence in a criminal case), 18 of which were based on police participants or an examination of police documents. Only two of these police participant studies were done in the USA, with the remainder conducted in various European countries. The studies provide supporting evidence that lay people and law enforcement professionals alike are vulnerable to confirmation bias, and there are other environmental, individual, and case-specific factors that may exacerbate this risk. Six studies described or evaluated the efficacy of intervention strategies, with varying evidence of success. Further research, particularly in the USA, is needed to evaluate different approaches to protect criminal investigations from cognitive biases.

Keywords Cognitive bias · Confirmation bias · Police · Investigation

Introduction

Decades of research in cognitive and social psychology have taught us that there are limitations to human attention and decision-making abilities (see, for example, Gilovich et al. 2002). We cannot process all the stimuli that surround us on a daily basis, so instead we have adapted for efficiency by attuning to patterns and developing mental shortcuts or rules of thumb to help us effectively navigate our complex world. While this tendency to rely on heuristics and biases can serve us well by allowing us to make quick decisions with little cognitive effort, it also has the potential to inadvertently undermine accuracy and thus the fair administration of justice.

Cognitive bias is an umbrella term that refers to a variety of inadvertent but predictable mental tendencies which can impact perception, memory, reasoning, and behavior. Cognitive biases include phenomena like confirmation bias (e.g., Nickerson 1998), anchoring (e.g., Tversky & Kahneman 1974), hindsight bias (e.g., Fischhoff 1975), the availability

heuristic (e.g., Tversky & Kahneman 1973), unconscious or implicit racial (or other identifying characteristics) bias (e.g., Greenwald et al. 1998; Staats et al. 2017), and others. In this context, the word “bias” does not imply an ethical issue (e.g., Dror 2020) but simply suggests a probable response pattern. Indeed, social scientists have demonstrated and discussed how even those who actively endorse egalitarian values harbor unconscious biases (e.g., Pearson et al. 2009; Richardson 2017) and how expertise, rather than insulating us from biases, can actually create them through learned selective attention or reliance on expectations based on past experiences (e.g., Dror 2020). Consequently, we recognize the potential for these human factors to negatively influence our criminal justice process.

In an effort to explore the role of cognitive biases in criminal investigations and prosecutions, we conducted a literature review to determine the scope of available research and strength of the findings. The questions guiding this exercise were as follows: (1) what topics have been researched so far and where are the gaps?; (2) what are the methodological strengths and limitations of this research?; and (3) what are the results, what do we know so far, and where should we go from here?

✉ Vanessa Meterko
vmeterko@innocenceproject.org

¹ Science & Research Department, Innocence Project,
New York, NY, USA

Methods

We searched PsycINFO for scholarly writing focused on cognitive biases in criminal investigations and prosecutions in December 2016 and again in January 2020.¹ We reviewed all results by title and then reviewed the subset of possibly-relevant titles by abstract, erring on the side of over-inclusivity. We repeated this process using the Social Sciences Full Text, PubMed, and Criminal Justice Abstracts with Full Text databases to identify additional papers. Finally, we manually reviewed the reference lists in the identified papers for any unique sources we may have missed in prior searches.

We sorted the articles into categories by the actor or action in the criminal investigation and prosecution process that they addressed, including physical evidence collection, witness evaluation, suspect evaluation, forensic analysis and testimony, police case evaluation (i.e., integrating and drawing conclusions based on the totality of the evidence), prosecutors, defense attorneys, judges, juries, and sentencing. Within each of these categories, we further sorted the articles into one of three types of sources: “primary data studies” describing experimental or observational studies that involved data collection or analysis, “intervention studies” that were solution-oriented and involved implementing some type of intervention or training to prevent or mitigate a phenomenon, and “secondary sources” (e.g., commentaries, letters, reviews, theoretical pieces, general book chapters) that discussed cognitive biases but did not present primary data.

To narrow the scope of this review, we did not include articles that focus solely on implicit racial bias or structural racial bias in the criminal legal system. The foundational and persistent problem of racial (particularly anti-Black) bias throughout our legal system—from policing to sentencing (e.g., Voigt et al. 2017; NYCLU 2011; Blair et al. 2004; Eberhardt et al. 2006)—has been clearly demonstrated in laboratory experiments and analyses of real-world data and is well-documented in an ever-growing body of academic publications and policy reports (e.g., Correll et al. 2002; Chanin et al. 2018; Owens et al. 2017; Staats et al. 2017).

¹ We used the following search terms and Boolean Operators: (criminal OR justice OR police OR investigat* OR forensic* OR jury OR juries OR judge* OR conviction* OR prosecut* OR defense OR defender* OR attorn*) in any field (e.g., text, title) AND (“cognitive bias” OR “cognitive dissonance” OR “tunnel vision” OR “confirmation bias” OR “interpretive bias” OR “belief perseverance” OR “asymmetrical skepticism”) in any field (e.g., text, title).

Results

Scope of Available Research and Methodology

Cognitive biases in forensic science have received the most attention from researchers to date (for a review of these forensic science studies, see Cooper & Meterko 2019). The second most substantial amount of scholarship focused on case evaluation (i.e., integrating and drawing conclusions based on the totality of the evidence in a case). Ultimately, we found 43 scholarly sources that addressed various issues related to the evaluation of the totality of evidence in criminal cases: 25 primary data (non-intervention) studies, five intervention studies, and one additional paper that presented both primary data *and* interventions, and 12 secondary sources. For the remainder of this article, we focus solely on the primary data and intervention studies. One of the primary data studies (Fahsing & Ask 2013) described the development of materials that were used in two subsequent studies included in this review (Fahsing & Ask 2016; 2017), and thus, this materials-development paper is not reviewed further here. Table 1 presents an overview of the research participants and focus of the other 30 primary data and intervention studies included in our review.

One challenge in synthesizing this collection of research is the fact that these studies address different but adjacent concepts using a variety of measures and—in some instances—report mixed results. The heterogeneity of this research reveals the complex nature of human factors in criminal case evaluations.

Eighteen of the 30 papers (13 primary data and three intervention) included participants who were criminal justice professionals (e.g., police, judges) or analyzed actual police documents. An appendix provides a detailed summary of the methods and results of the 18 criminal justice participant (or document) studies. Fifteen papers were based on or presented additional separate analyses with student or lay participants. Recruiting professionals to participate in research is commendable as it is notoriously challenging but allows us to identify any differences between those with training and experience versus the general public, and to be more confident that conclusions will generalize to real-world behavior. Of course, representativeness (or not) must still be considered when making generalizations about police investigations.

Reported sample sizes ranged from a dozen to several hundred participants and must be taken into account when interpreting individual study results. Comparison or control groups and manipulation checks are also essential to accurately interpreting results; some studies incorporated these components in their designs while others did not.

Table 1 Overview of 30 papers on cognitive biases in criminal case evaluation

Authors (Year)	Location	Participants ^a	Sample size	Research focus			
				Human nature	Environment and culture	Individual characteristics	Case-specific
Ask and Granhag (2005)	Sweden	Study 1: Police; Study 2: Students or community	Study 1: 50; Study 2: 68			Dispositional desire for closure	
Ask and Granhag (2007a)	Sweden	Police	61			Emotional state	
Ask and Granhag (2007b)	Sweden	Police	49	Asymmetrical skepticism	Time pressure		
Ask et al. (2008)	Sweden	Police	117	Asymmetrical skepticism			Type of evidence
Ask et al. (2011a)	Sweden	Police	104		Organizational norms of efficiency versus thoroughness		
Ask et al. (2011b)	Germany	Studies 1 and 2: Students or community	Study 1: 47; Study 2: 60	Cognitive dissonance; asymmetrical skepticism			Type of evidence
Charman et al. (2015)	USA	Students or community	382	Context effects; recency effects			
Charman et al. (2017)	USA	Police and students or community	89 (police) and 227 (students)	Confirmation bias			
Ditrich (2015)	Austria	Police	Not reported	Various cognitive fallacies			
Dando and Ormerod (2017)	United Kingdom	Police	60			Level of experience	Type of crime
Eerland et al. (2012)	Netherlands	Students or community	63	Feature positive effect			
Eerland and Rassin (2012)	Netherlands	Students or community	188	Confirmation bias; Feature positive effect			
Fahsing and Ask (2016) ^b	Norway and England	Police	124		Type of training	Level of experience	Tipping point
Fahsing and Ask (2017) ^b	Norway	Police	166			Reasoning abilities	Tipping point
Greenspan and Surich (2016)	USA	Studies 1 and 2: Students or community	Study 1: 119; Study 2: 127	Coherence-based reasoning			

Table 1 (continued)

Authors (Year)	Location	Participants ^a	Sample size	Research focus			
				Human nature	Environment and culture	Individual characteristics	Case-specific
Groenendaal and Helsloot (2015)	Netherlands	Police	60				Criminal Investigation Reinforcement Programme
Haas et al. (2015)	Switzerland	Students or community	127				Training and checklist tool
Jones et al. (2008)	United Kingdom	Police	12				Review tool
Kerstholt and Eikelboom (2007)	Netherlands	Police	38	Groupthink		Level of experience	
Marksteiner et al. (2010)	Sweden	Police	107	Asymmetrical skepticism			
O'Brien (2009)/O'Brien (2007) ^c	USA	Studies 1 and 2: Students or community	Study 1: 108; Study 2: 109	Confirmation bias			Crime severity
Price and Dahl (2014)	Canada	Students or community	179	Recency effects			Generating alternative hypotheses, challenging a hypothesis, accountability to process or outcome, need to persuade others
Rassin (2010)	Netherlands	Study 1: Police ^d , Study 2: Students or community	Study 1: 118; Study 2: 178	Study 1: Confirmation bias		Study 2: Dispositional desire for closure	
Rassin (2018a)	Netherlands	Study 2: Students or community ^e	71			Reasoning abilities	
Rassin (2018b)	Netherlands	Study 1: Students or community; Study 2: Police ^d	Study 1: 277; Study 2: 45				Pen and paper tool
Rassin et al. (2010)	Netherlands	Studies 1–3: Students or community	Study 1: 79; Study 2: 129; Study 3: 182	Confirmation bias			Crime severity
Salet and Terpstra (2014)	Netherlands	Police	26 (case files) and 47 (interviews)				Use of “contrarians” and “critical review” procedures
Simon et al. (2004)	USA	Studies 1 and 2: Students or community	Study 1: 154; Study 2: 344	Coherence-based reasoning			
Wallace (2015)	USA	Police	166	Confirmation bias		Level of experience	Crime severity

Table 1 (continued)

Authors (Year)	Location	Participants ^a	Sample size	Research focus			
				Human nature	Environment and culture	Individual characteristics	Case-specific
Wastell et al. (2012)	Australia	Students or community	Study 1: 50; Study 2: 44	Hypothesis confirmation			

^a“Police” participant studies were those that used police personnel (including police officers, criminal investigators, crime analysts, police trainees or recruits) or a review of police documents (including investigative case files or decision logs); “Students or community” participant studies were those that used undergraduate students, graduate students, law students, US citizens, or a general public/online sample

^bFahsing and Ask developed materials for these studies in 2013 by conducting semi-structured interviews to elicit factors that could disrupt optimal decision-making in homicide investigations; used content analysis to develop categories of tipping points (naming, arresting, or charging a suspect, choice of main hypotheses or lines of inquiry) and related situational (availability of information/evidence, external pressure/community impact, internal pressure/organizational issues, time pressure) and individual (detective experience, training and education, personal characteristics) factors

^cO’Brien’s 2009 publication includes a subset of studies from her 2007 dissertation. Consequently, the numbering of studies in these two papers is different: published “Study 1” = dissertation “Study 2” ($N = 108$) and published “Study 2” = dissertation “Study 3” ($N = 109$). The 2007 dissertation also reported on a Study 1 ($N = 50$), a Study 4 ($N = 123$), and a Study 5 ($N = 96$), all with student participants as well. The numbers in this table are taken from the published paper

^dIn addition to police officers, these studies included other criminal justice professional participants (district attorneys and judges)

^eRassin 2018a also included a Study 1, a descriptive analysis of logical reasoning skills in a sample of 208 students and 40 criminal trial judges. This study did not attempt to analyze reasoning skills in relation to aspects of case evaluation, and so this study is not included in the summary

Most studies used vignettes or case materials—both real and fictionalized—as stimuli. Some studies did not include enough information about stimulus or intervention materials to allow readers to critically interpret the results or replicate an intervention test. Future researchers would benefit from publishers making more detailed information available. Further, while the use of case vignettes is a practical way to study these complex scenarios, this approach may not completely mimic the pressures of a real criminal case, fully appreciate how the probative value of evidence can depend on context, or accurately reflect naturalistic decision-making.

Notably, only two of the criminal case evaluation studies using professional participants were conducted in the USA; all others were based in Europe (Austria, Netherlands, Norway, Sweden, and the UK). The differences between police training, operations, and the criminal justice systems writ large should be considered when applying lessons from these studies to the USA or elsewhere.

Finally, all of these papers were published relatively recently, within the past 15 years. This emerging body of research is clearly current, relevant, and has room to grow.

Research Findings

The primary data studies address a constellation of concepts that demonstrate how human factors can inadvertently undermine the seemingly objective and methodical process of a criminal investigation. To organize these concepts, we used a taxonomy originally developed to describe potential sources of bias in forensic science observations and conclusions as a guide (Dror 2017; Dror et al. 2017) and adapted it to this collection of case evaluation literature.² As in Dror’s taxonomy, the broad base of this organizing pyramid is “human nature,” and as the pyramid narrows to its peak, potential sources of bias become increasingly dependent on environmental, individual, and case-specific circumstances and characteristics (Fig. 1). Some authors in this collection address more than one of these research areas within the same paper through multiple manipulations or a series of studies (Table 1).

Human Nature

The “human nature” studies include those that demonstrate universal psychological phenomena and their underlying mechanisms in the context of a criminal case evaluation. Several studies focused on confirmation bias. Confirmation

² As Dror (2017) notes, the development of this taxonomy began in a paper in 2009 (Dror 2009) and was further developed in a 2014 paper (Stoel et al. 2014), with additional sources of bias added subsequently (in Dror 2015, and Zapf & Dror 2017).

Fig. 1 Organizational framework for case evaluation studies, adapted from Dror's (2017) taxonomy of different sources of potential bias that may cognitively contaminate forensic observations and conclu-

sions. The specific factors listed in this pyramid are those that were examined in the collection of studies in the present literature review

bias, sometimes colloquially referred to as “tunnel vision,” denotes selective seeking, recalling, weighting, and/or interpreting information in ways that support existing beliefs, expectations, or hypotheses, while simultaneously avoiding or minimizing inconsistent or contradictory information (Nickerson 1998; Findley 2012). Some authors in this collection of studies used other terms to describe this concept or elements of it, including “context effects,” the term used by Charman et al. (2015) to describe when “a preexisting belief affects the subsequent interpretation of evidence” (p. 214), and asymmetrical skepticism (Ask & Granhag 2007b; Marksteiner et al. 2010).

Eight studies with law enforcement personnel (Ask & Granhag 2007b; Ask et al. 2008; Charman et al. 2017; Ditrich 2015; Groenendaal & Helsloot 2015; Marksteiner et al. 2010; Rassin 2010; Wallace 2015) examined aspects of confirmation bias; one addressed the distinct but related phenomenon of groupthink (Kerstholt & Eikelboom 2007). The importance of this issue was demonstrated by a survey of an unspecified number of professional crime scene officers conducted by Ditrich (2015), asking for their opinions

about the relative frequency and severity of various cognitive errors that could potentially negatively affect a criminal investigation; based on their experiences, respondents highlighted confirmation bias (as well as overestimating the validity of partial information and shifting the burden of proof to the suspect). The other studies within this group used experimental designs to assess police officers' evaluation of evidence. Charman et al. (2017) reported that police officers' initial beliefs about the innocence or guilt of a suspect in a fictional criminal case predicted their evaluation of subsequent ambiguous evidence, which in turn predicted their final beliefs about the suspect's innocence or guilt. This is not the only study to demonstrate that, like the rest of us, police officers are susceptible to confirmation bias. Ask and colleagues (2008) found that police recruits discredited or supported the same exact evidence (“the viewing distance of 10 m makes the witness identification unreliable” versus “from 10 m one ought to see what a person looks like”) depending on whether it was consistent or inconsistent with their hypothesis of a suspect's guilt. Ask and Granhag (2007b) found that when experienced criminal investigators

read a vignette that implied a suspect's guilt (but left room for an alternative explanation), they rated subsequent guilt-consistent evidence as more credible and reliable than evidence that was inconsistent with their theory of guilt; similar results were seen in a study of police officers, district attorneys, and judges by Rassin (2010).

Marksteiner et al. (2010) investigated the motivational underpinnings of this type of asymmetrical skepticism among police trainees, asking whether it is driven by a desire to reconcile inconsistent information with prior beliefs or by the goal of case closure, and encountered mixed results. The group who initially hypothesized guilt reacted as expected, rating subsequent incriminating evidence as more reliable, but in the group whose initial hypothesis was innocence, there was no difference in the way that they rated additional consistent or inconsistent information. Wallace (2015) found that the order in which evidence was presented influenced guilt beliefs. When police officers encountered exculpatory evidence prior to inculpatory evidence, guilt belief scores decreased, suggesting their final decisions were influenced by their initial impressions. Kerstholt and Eikelboom (2007) describe how teams tend to converge on one interpretation, and once such an interpretation is adopted, individual members are less able to examine underlying assumptions critically. They asked independent crime analysts to evaluate a realistic criminal investigation with fresh eyes and found that they were demonstrably influenced when they were aware of the investigative team's existing working hypothesis.

Studies in student and general populations examining confirmation bias and other aspects of human cognition (Ask et al. 2011b; Charman et al. 2015; Eerland et al. 2012; Eerland & Rassin 2012; Greenspan & Surich 2016; O'Brien 2007; 2009; Price & Dahl 2014; Rassin et al. 2010; Simon et al. 2004; Wastell et al. 2012) reported similar patterns to those described above with police participants. O'Brien (2007; 2009) found that students who named a suspect early in a mock criminal investigation were biased towards confirming that person's guilt as the investigation continued. O'Brien measured memory for hypothesis-consistent versus hypothesis-inconsistent information, interpretation of ambiguous evidence, participants' decisions to select lines of inquiry into the suspect or an alternative, and ultimate opinions about guilt or innocence. In a novel virtual crime scene investigation, Wastell et al. (2012) found that all students (those who ultimately chose the predetermined "correct" suspect from the multiple available people of interest and those who chose incorrectly) sought more chosen-suspect-consistent information during the exercise. However, those who were ultimately unsuccessful (i.e., chose the wrong person) spent more time in a virtual workspace (a measure of the importance placed on potential evidence) after accessing confirmatory information. They also found that students who settled on a suspect early in the exercise—measured

by prompts throughout the virtual investigation—were comparatively unsuccessful.

Other psychological phenomena such as recency effects (i.e., our ease of recalling information presented at the end of a list relative to information presented at the beginning or middle) and the feature positive effect (i.e., our tendency to generally attune to presence more than absence) were also examined in studies with student or general population participants. Price and Dahl (2014) explored evidence presentation order and found that under certain circumstances, evidence presented later in an investigation had a greater impact on student participant decision-making in a mock criminal investigation. Charman and colleagues also found order of evidence presentation influenced ratings of strength of evidence and likelihood of guilt in their 2015 study of evidence integration with student participants. These results appear to provide evidence *against* the presence of confirmation bias, but recency effects still demonstrate the influence of human factors as, arguably, the order in which one learns about various pieces of evidence—whether first *or* last—should not impact interpretation. Several research teams found that a positive eyewitness identification is seen as more credible than a failure to identify someone (Price & Dahl 2014, p.147) and the presence of fingerprints—as opposed to a lack of fingerprints—is more readily remembered and used to make decisions about a criminal case (Eerland et al. 2012; Eerland & Rassin 2012), even though the absence of evidence can also be diagnostic. Other researchers highlighted our psychic discomfort with cognitive dissonance (Ask et al. 2011b) and our tendency to reconcile ambiguity and artificially impose consistency in a criminal case by engaging in "bidirectional coherence-based reasoning" (Simon et al. 2004; Greenspan & Surich 2016).

Environment and Culture

The three "environment and culture" studies with police personnel (Ask & Granhag 2007b; Ask et al. 2011a; Fahsing & Ask 2016) revealed the ways in which external factors can influence an investigation. For instance, type of training appears to impact the ability to generate a variety of relevant hypotheses and actions in an investigation. English and Norwegian investigators are trained and performed differently when faced with semi-fictitious crime vignettes (Fahsing & Ask 2016). Organizational culture can impact the integrity of an investigation as well. Ask and colleagues (2011a) concluded that a focus on efficiency—as opposed to thoroughness—produces more cursory processing among police participants, which could be detrimental to the accurate assessment of evidence found later in an investigation. Ask and Granhag (2007b) observed that induced time pressure influenced officers' decision-making, creating a higher

tendency to stick with initial beliefs and a lower tendency to be influenced by the evidence presented.

Individual Characteristics

Seven “individual characteristics” studies with police personnel (Ask & Granhag 2005; 2007a; Dando & Ormerod 2017; Fahsing & Ask 2016; 2017; Kerstholt & Eikelboom 2007; Wallace 2015) plus two studies with student populations (Rassin 2010, 2018a) examined ways in which personal attributes can influence an investigation. Varying amounts of professional experience may matter when it comes to assessments of potential criminal cases and assumptions about guilt. For instance, police recruits appear to have a strong tendency toward criminal—as opposed to non-criminal—explanations for an ambiguous situation like a person’s disappearance (Fahsing & Ask 2017) and less experienced recruits show more suspicion than seasoned investigators (Wallace 2015). In a departure from the typical mock crime vignette method, Dando and Ormerod (2017) reviewed police decision logs (used for recording and justifying decisions made during serious crime investigations) and found that senior officers generated more hypotheses early in an investigation, and switched between considering different hypotheses both early and late in an investigation (suggesting a willingness to entertain alternative theories) compared with inexperienced investigators. An experimental study, however, found that professional crime analyst experience level (mean 7 months versus 7 years) was not related to case evaluation decisions and did not protect against knowledge of prior interpretations of the evidence influencing conclusions (Kerstholt & Eikelboom 2007).

Two studies examined differences in reasoning skills in relation to the evaluation of evidence. Fahsing and Ask (2017) found that police recruits’ deductive and inductive reasoning skills were not associated with performance on an investigative reasoning task. In contrast, in a study with undergraduate students, accuracy of decision-making regarding guilt or innocence in two case scenarios was associated with differences in logical reasoning abilities as measured by a test adapted from the Wason Card Selection Test (Rassin 2018a).

Ask and Granhag (2005) found inconsistent results in a study of police officers’ dispositional need for cognitive closure and the effect on criminal investigations. Those with a high need for cognitive closure (measured with an established scale) were less likely to acknowledge inconsistencies in case materials when those materials contained a potential motive for the suspect, but were *more* likely to acknowledge inconsistencies when made aware of the possibility of an alternative perpetrator. In a replication study with undergraduate students, Ask & Granhag (2005) found that initial hypotheses significantly affected subsequent evidence

interpretation, but found no interaction with individual need for cognitive closure. Students who were aware of an alternative suspect (compared with those aware of a potential motive for the prime suspect) were simply less likely to evaluate subsequent information as evidence supporting guilt.

In another study, when Ask and Granhag (2007a) induced negative emotions in police officers and then asked them to make judgments about a criminal case, sad participants were better able to substantively process the consistency of evidence or lack thereof, whereas angry participants used heuristic processing.

Case-Specific

Four studies of police personnel (Ask et al. 2008; Fahsing & Ask 2016; 2017; Wallace 2015), one using police records (Dando & Omerod 2017), and three studies of student populations (Ask et al. 2011b; O’Brien 2007; 2009; Rassin et al. 2010) examined “case-specific” and evidence-specific factors. In a study of police officers, Ask and colleagues (2008) showed that the perceived reliability of some types of evidence (DNA versus photographs versus witnesses) is more malleable than others; similar results pertaining to DNA versus witness evidence were found in a study of law students (Ask et al. 2011b).

Fahsing and Ask (2016) found that police recruits who were presented with a scenario including a clear “tipping point” (an arrest) did not actually produce significantly fewer hypotheses than those who were not presented with a tipping point (though they acknowledge that the manipulation—one sentence embedded in a case file—may not have been an ecologically valid one). In a subsequent study with police recruits, the presence of a tipping point resulted in fewer generated hypotheses, but the difference was not statistically significant (Fahsing & Ask 2017).

Other studies using law students (Rassin et al. 2010) or undergraduate students (O’Brien 2007) examined the influence of crime severity on decision-making. Rassin et al. (2010) observed that the affinity for incriminating evidence increases with crime severity, but in one of O’Brien’s (2007) studies, crime severity did not have a demonstrable impact on confirmation bias.

Interventions

Taken together, this body of work demonstrates vulnerabilities in criminal investigations. Some researchers have suggested theoretically supported solutions to protect against these vulnerabilities, such as gathering facts rather than building a case (Wallace 2015) or institutionalizing the role of a “contrarian” in a criminal investigation (MacFarlane 2008). Few studies have tested and evaluated these potential remedies, however. Testing is an essential prerequisite to

any advocacy for policy changes because theoretically sound interventions may not, in fact, have the intended effect when applied (e.g., see below for a description of O'Brien's work testing multiple interventions with differing results).

Four studies have examined various intervention approaches with police departments or investigators (Groenendaal & Helsloot 2015; Jones et al. 2008; Rassin 2018b; Salet & Terpstra 2014). Jones et al. (2008) created a tool that helped an experimental group of investigators produce higher quality reviews of a closed murder case than those working without the aid of the review tool. Their article provides an [appendix](#) with “categories used in the review tool” (e.g., crime scene management, house-to-house enquiries, community involvement) but lacks a detailed description of the tool itself and the outcome measures. Importantly, the authors raise the possibility that a review tool like this may improve how officers think through a case because of the structure or content of the tool *or* it may succeed by simply slowing them down so they can think more critically and thoroughly. Another approach that shows promise in reducing tunnel vision is using a pen and paper tool to prompt investigators to consider how well the same evidence supports different hypotheses (Rassin 2018b). In a study of actual case files, supplemented with interviews, Salet and Terpstra (2014) explored “contrarians” and found that there are real-world challenges to the position's efficacy (e.g., personal desire to be a criminal investigator, desire for solidarity with colleagues) and considerable variability in the way contrarians approach their work, with some opting for closeness to an investigation and others opting for distance; individuals also embraced different roles (e.g., supervisor, devil's advocate, focus on procedure). The researchers concluded that, in practice, these contrarians appear to have exerted subtle influence on investigations but there is no evidence of a radical change in case trajectory. Similarly, members of criminal investigation teams in the Netherlands reported that, in practice, designated devil's advocates tend to provide sound advice but do not fundamentally change the course of investigations (Groenendaal & Helsloot 2015). Groenendaal and Helsloot describe the development and implementation of the Criminal Investigation Reinforcement Programme in the Netherlands, which was prompted by a national reckoning stemming from a widely publicized wrongful conviction. The program included new policies aimed at, among other things, reducing tunnel vision (including the use of devil's advocates, structured decision-making around “hypotheses and scenarios,” and professionalized, permanent “Command Core Teams” dedicated to major crimes). This deliberate intervention provided an opportunity for researchers to interview investigators who were directly impacted by the new policies. Groenendaal and Helsloot conclude that the main effect of this intervention was an increased awareness about the potential problem of tunnel vision, and they focus on an

unresolved a tension between “efficacy” (more convictions) and “precaution” (minimizing wrongful convictions). Their work underscores the importance of collecting criminal legal system data, as interviewees reported their experiences and impressions but could not report whether more correct convictions had been obtained or more wrongful convictions avoided.

Other studies have examined various intervention ideas with student populations (Haas et al. 2015; O'Brien 2007; 2009). Haas et al. (2015) found that using a checklist tool to evaluate evidence appears to improve students' abductive reasoning and reduce confirmation bias. O'Brien (2007; 2009) found that orienting participants to being accountable for good process versus outcome had no impact, and that when participants expected to have to persuade someone of their hypothesis, this anticipation actually worsened bias. More promisingly, she discovered that participants who were asked to name a suspect early in an investigation, but were then told to consider how their selected suspect could be innocent and then generate counter-arguments, displayed less confirmation bias across a variety of measures (they looked the same as those who did not name a suspect early). But another approach—asking participants to generate two additional alternative suspects—was not effective (these participants showed the same amount of bias as those who identified just one suspect).

Discussion

Zalman and Larson (2016) have observed “the failure of innocence movement advocates, activists, and scholars to view the entirety of police investigation as a potential source of wrongful convictions, as opposed to exploring arguably more discrete police processes (e.g., eyewitness identification, interrogation, handling informants)” (p.3). While the thorough examination of these discrete processes has led to a better understanding of risk factors and, ultimately, reforms in police practices (e.g., see the Department of Justice 2017 guidelines for best practices with eyewitnesses), a recent shift towards viewing wrongful convictions from a “sentinel events”³ perspective advances the conversation around these criminal justice system failures (Doyle 2012; 2014; Rossmo & Pollock 2019).

This literature review has identified a body of research that lends support to this holistic perspective. The studies reviewed here address a constellation of concepts that

³ According to the National Institute of Justice (2017), a sentinel event is a significant negative outcome that (1) signals underlying weaknesses in the system or process, (2) is likely the result of compound errors, and (3) may provide, if properly analyzed and addressed, important keys to strengthen the system and prevent future adverse outcomes.

demonstrate how the human element—including universal psychological tendencies, predictable responses to situational and organizational factors, personal factors, and characteristics of the crime itself—can unintentionally undermine truth-seeking in the complex evidence integration process. Some concepts are addressed by one study, some are addressed by several, and some studies explored multiple variables (e.g., demonstrating the existence of confirmation bias *and* measuring how level of professional experience plays a role).

Several contemporary studies have demonstrated the existence of confirmation bias in police officers within the context of criminal investigations. Other psychological phenomena have not been examined in police populations but have been examined in student or general populations using study materials designed to assess the interpretation of criminal case evidence and decision-making. This collection of studies also investigates the role of environmental factors that may be specific to a department or organization, characteristics of individual investigators, or of the specific case under review. At the environmental level, type of training and organizational customs were influential and are promising areas for further research as these factors are within the control of police departments and can be modified. With respect to individual characteristics, a better understanding of advantageous dispositional tendencies and what is gained by professional experience, as well as the unique risks of expertise, could lead to better recruitment and training methods. Case-specific factors are outside the control of investigators, but awareness of factors that pose a greater risk for bias could serve as an alert and future research could identify ways to use this information in practice (see also Rossmo & Pollock 2019 for an in-depth discussion of “risk recipes”).

Charman and colleagues (2017) present a particularly interesting illustration of the way in which a criminal case is not merely the sum of its parts. In this study, the researchers presented law enforcement officers with exonerating, incriminating, or neutral DNA or eyewitness evidence, collected initial beliefs about guilt, asked participants to evaluate a variety of other ambiguous evidence (alibi, composite sketch, handwriting comparison, and informant information that could be reasonably interpreted in different ways), and then provide a final rating of guilt. As hypothesized, the researchers found those who were primed with incriminating evidence at the beginning were more likely to believe the suspect guilty at the end. However, even those who initially received exonerating information and initially rated the likelihood of suspect guilt as relatively low ended up increasing their guilt rating after reviewing the other ambiguous evidence. It appears that the cumulative effect of ambiguous evidence tilted the scales towards guilt. This unexpected outcome underscores the value of understanding how the

totality of evidence in a criminal case is evaluated, and has implications for the legal doctrine of “harmless error” rooted in assumptions of evidentiary independence (e.g., Hasel & Kassin 2009).

Consistently incorporating control groups into future study designs and including complete stimulus materials in future publications could build on this foundation. This would help future researchers fully interpret and replicate study results and would assist in determining what elements of intervention strategies work. Since the majority of these studies were conducted in Europe, it would be worthwhile to explore whether or not these results can be replicated in the USA, given the similarities and differences in our criminal justice systems and the variety of approaches used to select and train detectives across police departments. Finally, valuable future research will move beyond the demonstration of these human vulnerabilities and will design and test strategies to mitigate them in the complex real world.⁴ Vignettes and mock-investigations are clever ways of studying criminal investigations, but it is worth remembering that these approaches cannot fully capture the dynamics of a real criminal investigation. Collaboration between academic researchers and criminal investigators could generate robust expansions of this work.

Evidence evaluation and synthesis in criminal investigations is, of course, just one part of a larger legal process. In addition to police, defense attorneys, prosecutors, and judges have powerful roles in determining case outcomes, especially in a system that is heavily reliant on plea bargaining. Critically addressing the potential influence of cognitive biases throughout this system, and promoting and implementing proven, practical protections against these tendencies will advance accuracy and justice.

⁴ As Snook and Cullen (2008) assert, “it is unrealistic to expect police officers to investigate *all* possible suspects, collect evidence on *all* of those suspects, explore *all* possible avenues concerning the circumstances surrounding a crime, search for disconfirming and confirming evidence of guilt for every suspect, and integrate all of this information” (p. 72). Dando and Ormerod (2017) illustrate this real-world complexity when they describe an investigation that was delayed because a call for tips led to a flood of false leads, suggesting that more information is not *always* better. Further, though it addresses procedural justice in street policing rather than evidence integration in a criminal investigation (and thus was not included in this review), Owens et al. (2018) provide an example of a field study, complete with published scripts. Recognizing the automated thinking and behavior that comes with job experience, these researchers tested an intervention to reduce the number of incidents resolved with arrests and use of force by implementing a training program aimed at encouraging beat officers to think more slowly and deliberately during routine encounters; they also assessed the cost of this intervention in the police department.

Appendix. Detailed Summary of 18 Studies with Police Participants or Source Materials

Reference, Participants, Research Focus	Methods and Measures	Results																		
Ask and Granhag (2005)	<ul style="list-style-type: none"> Materials: Homicide case^a description that included either information about a named female suspect's motive or information about an alternative male suspect. Explanatory Measures: Dispositional need for cognitive closure measured with Need For Closure Scale (NFCS); homicide case scenario version (random assignment). Primary Outcome Measures: Ratings of probable guilt of female suspect, strength of evidence, criteria for probable cause fulfilled, adequacy of evidence for prosecution (1 to 9 Likert scale); ratings of degree to which 10 particular case details supported innocence of the named female suspect (-4 to 4 Likert scale); assessment of willingness to acknowledge evidence inconsistent with guilt hypothesis (number of case details rated zero or lower, i.e., consistent with innocence). 	<p>Ratings of probable guilt, strength of evidence, criteria for probable cause, and adequacy of evidence were not associated with NFCS or case scenario</p> <p>All participants appeared to believe the case details pointed toward the female suspect's guilt whether they were presented with her motive or with an alternative suspect.</p> <p>Because of this apparent guilt bias, researchers also assessed participants' willingness to acknowledge evidence consistent with innocence. An interaction between NFCS and case scenario was seen in this assessment: $\beta = -0.03$, $p = 0.20$ for suspect motive condition and $\beta = 0.04$, $p = 0.18$ for alternative suspect condition; thus, high-NFC participants were less likely than their low-NFC counterparts to acknowledge inconsistencies when presented with a suspect motive but were more likely to do so when presented with an alternative perpetrator</p>																		
Ask and Granhag (2007a)	<ul style="list-style-type: none"> Materials: Assault case^b description suggesting the victim's father was guilty + two witness statements. Second witness statement was either consistent or inconsistent with the hypothesis that the victim's father was responsible. Explanatory Measures: Emotional state induced by instructions to recall and write about an event that caused anger or sadness (random assignment); witness statements version (random assignment). Primary Outcome Measures: Ratings of witness reliability, perceived trustworthiness, perceived favorableness of the witnessing conditions, and weight of the witness evidence; global ratings of probable guilt of father and strength of evidence (all 1 to 9 Likert scales). 	<p>Results indicate emotional state may affect how open-minded a criminal investigator is.</p> <p>“Anger” group did not rate guilt or evidence strength differently in the consistent and inconsistent witness conditions, while “sadness” group in the consistent condition rated higher probability of guilt and stronger evidence against the father:</p> <table> <tr> <th></th><th colspan="2">[Mean $\pm$ SD]</th></tr> <tr> <th>Rating of Guilt</th><th>Anger</th><th>Sadness</th></tr> <tr> <td>Consistent witnesses</td><td>6.71 $\pm$ 1.53</td><td>7.36 $\pm$ 1.08</td></tr> <tr> <td>Inconsistent witnesses</td><td>6.20 $\pm$ 1.15 ($p = 0.31$)</td><td>5.40 $\pm$ 1.68 ($p < 0.001$)</td></tr> </table> <p>Strength of evidence</p> <table> <tr> <td>Consistent witnesses</td><td>4.76 $\pm$ 2.22</td><td>5.36 $\pm$ 2.34</td></tr> <tr> <td>Inconsistent witnesses</td><td>5.20 $\pm$ 2.01 ($p < 1.0$)</td><td>3.60 $\pm$ 1.84 ($p < 0.05$)</td></tr> </table>		[Mean $\pm$ SD]		Rating of Guilt	Anger	Sadness	Consistent witnesses	6.71 $\pm$ 1.53	7.36 $\pm$ 1.08	Inconsistent witnesses	6.20 $\pm$ 1.15 ($p = 0.31$)	5.40 $\pm$ 1.68 ($p < 0.001$)	Consistent witnesses	4.76 $\pm$ 2.22	5.36 $\pm$ 2.34	Inconsistent witnesses	5.20 $\pm$ 2.01 ($p < 1.0$)	3.60 $\pm$ 1.84 ($p < 0.05$)
	[Mean $\pm$ SD]																			
Rating of Guilt	Anger	Sadness																		
Consistent witnesses	6.71 $\pm$ 1.53	7.36 $\pm$ 1.08																		
Inconsistent witnesses	6.20 $\pm$ 1.15 ($p = 0.31$)	5.40 $\pm$ 1.68 ($p < 0.001$)																		
Consistent witnesses	4.76 $\pm$ 2.22	5.36 $\pm$ 2.34																		
Inconsistent witnesses	5.20 $\pm$ 2.01 ($p < 1.0$)	3.60 $\pm$ 1.84 ($p < 0.05$)																		
Ask and Granhag (2007b)	<ul style="list-style-type: none"> Materials: Homicide case^a description that included information about a named female suspect's motive + witness statement that either supported theory of female suspect's guilt or contradicting theory of female suspect's guilt. 	<p>Results provide evidence for asymmetrical skepticism in criminal investigations.</p>																		

Reference, Participants, Research Focus	Methods and Measures	Results																					
<i>N</i> =49 criminal investigators, mean experience 10 years, Sweden	<ul style="list-style-type: none">Explanatory Measures: Need for cognitive closure induced by time pressure (assignment to one of two groups; unlimited time to complete task or limited to 20 minutes, which was less than the median time the participants in the comparison group took); case scenario version (random assignment not specified).Primary Outcome Measures: Pre- and post-eyewitness evidence ratings of guilt or innocence (dichotomous), confidence, strength of evidence, adequacy of evidence for prosecution; ratings of witness credibility, conditions for making reliable observations, impact of witness emotional reaction, effect of 7-day delay in witness report to police, weight of witness evidence in relation to other evidence, agreement of witness evidence with other evidence (all 9-point Likert scales).	Lower ratings for witness reliability, witnessing conditions, retention interval, and weight of witness evidence for witness who provided evidence contradicting (exonerating), as opposed to supporting, the suspect's guilt (incriminating): <table><tr><td></td><td colspan="2">[Mean ± SD]</td></tr><tr><td></td><td>Incriminating</td><td>Exonerating</td></tr><tr><td>Witness reliability</td><td>6.32 ± 1.53</td><td>5.40 ± 1.54</td></tr><tr><td>Witnessing conditions</td><td>6.21 ± 1.65</td><td>5.20 ± 1.88</td></tr><tr><td>Retention interval</td><td>5.42 ± 1.22</td><td>4.80 ± 1.70</td></tr><tr><td>Weight of evidence</td><td>5.00 ± 2.19</td><td>4.18 ± 1.63</td></tr></table> <p>No significant differences were seen in ratings of witness credibility or impact of witness emotions, though the patterns emerged in the same direction.</p> <p>Little difference in asymmetrical skepticism results when comparing low- and high-time pressure groups, though the greatest response to information conveyed in the witness statement was found in the low-time pressure condition.</p>		[Mean ± SD]			Incriminating	Exonerating	Witness reliability	6.32 ± 1.53	5.40 ± 1.54	Witnessing conditions	6.21 ± 1.65	5.20 ± 1.88	Retention interval	5.42 ± 1.22	4.80 ± 1.70	Weight of evidence	5.00 ± 2.19	4.18 ± 1.63			
	[Mean ± SD]																						
	Incriminating	Exonerating																					
Witness reliability	6.32 ± 1.53	5.40 ± 1.54																					
Witnessing conditions	6.21 ± 1.65	5.20 ± 1.88																					
Retention interval	5.42 ± 1.22	4.80 ± 1.70																					
Weight of evidence	5.00 ± 2.19	4.18 ± 1.63																					
Ask, Rebelius, and Granthag (2008)	<ul style="list-style-type: none">Materials: Homicide case^c description suggesting guilt of a suspect apprehended near the scene + one additional type of evidence (DNA, eyewitness, photo) that was either consistent with or inconsistent with a guilty conclusion.Explanatory Measures: Type of evidence (random assignment); consistent or inconsistent evidence version (random assignment).Primary Outcome Measures: Ratings of guilt probability, strength of evidence before and after review of additional evidence, ratings of reliability of the specific evidence provided and of the reliability of that evidence type in general (all 9-point Likert scales).	Results provide evidence for asymmetrical skepticism in criminal investigations. Higher rating of reliability for evidence consistent with guilt compared to evidence inconsistent with guilt across types of evidence in the case: <table><tr><td></td><td colspan="2">[Mean ± SD]</td></tr><tr><td></td><td>Consistent</td><td>Inconsistent</td></tr><tr><td>DNA</td><td>7.50</td><td>0.74</td></tr><tr><td>Photo</td><td>6.65 ± 1.07</td><td>5.60 ± 1.85</td></tr><tr><td>Witness</td><td>7.00 ± 1.03</td><td>4.90 ± 1.18</td></tr></table> <p>Type of Evidence: DNA evidence rated higher reliability than photo or witness (<i>p</i> < 0.01); no difference between photo and witness</p> <table><tr><td>DNA</td><td>7.41 ± 1.09</td></tr><tr><td>Photo</td><td>6.16 ± 1.56</td></tr><tr><td>Witness</td><td>5.81 ± 1.52</td></tr></table>		[Mean ± SD]			Consistent	Inconsistent	DNA	7.50	0.74	Photo	6.65 ± 1.07	5.60 ± 1.85	Witness	7.00 ± 1.03	4.90 ± 1.18	DNA	7.41 ± 1.09	Photo	6.16 ± 1.56	Witness	5.81 ± 1.52
	[Mean ± SD]																						
	Consistent	Inconsistent																					
DNA	7.50	0.74																					
Photo	6.65 ± 1.07	5.60 ± 1.85																					
Witness	7.00 ± 1.03	4.90 ± 1.18																					
DNA	7.41 ± 1.09																						
Photo	6.16 ± 1.56																						
Witness	5.81 ± 1.52																						
<i>N</i> =117 police trainees, Sweden																							
Human Nature; Case-Specific																							
Similar results were seen with ratings of evidence in general.																							

Reference, Participants, Research Focus	Methods and Measures	Results															
Ask, Granhag, and Rebelius (2011a)	<ul style="list-style-type: none"> Materials: Given description of a “good investigator” that emphasized efficiency, thoroughness, or neither; then given assault case description suggesting the victim’s father was guilty + two witness statements. Second witness statement was either consistent or inconsistent with the hypothesis that the victim’s father was responsible. Explanatory Measures: Goal activation through “good investigator” definition emphasizing efficiency, thoroughness, or neither/control (random assignment); witness statements version (random assignment). Primary Outcome Measures: Task completion speed (relative to other participants), ratings of guilt probability and strength of evidence (1 to 7 Likert scales), combined into a composite score, ratings of their own case processing with respect to judgement spontaneity, cognitive effort, difficulty deciding guilt, and confidence (1 to 7 Likert scales). 	Results provide evidence for salient social norms influencing information processing in criminal investigations.															
N=104 criminal investigators, mean experience 21 years, Sweden		Efficiency versus thoroughness associated with composite rating of guilt, with interaction by type of witness evidence; guilt ratings in the thoroughness group were most influenced:															
Environment and Culture		<p>[Mean ± SD]:</p> <table> <tr> <td></td><td>Incriminating</td><td>Exculpatory</td></tr> <tr> <td>Efficiency</td><td>4.81 ± 0.84</td><td>4.44 ± 1.04 (p = 0.30)</td></tr> <tr> <td>Thoroughness</td><td>5.32 ± 0.98</td><td>3.97 ± 1.20 (p < 0.001)</td></tr> <tr> <td>Control</td><td>5.19 ± 1.06</td><td>4.14 ± 1.10 (p < 0.05)</td></tr> </table>		Incriminating	Exculpatory	Efficiency	4.81 ± 0.84	4.44 ± 1.04 (p = 0.30)	Thoroughness	5.32 ± 0.98	3.97 ± 1.20 (p < 0.001)	Control	5.19 ± 1.06	4.14 ± 1.10 (p < 0.05)			
	Incriminating	Exculpatory															
Efficiency	4.81 ± 0.84	4.44 ± 1.04 (p = 0.30)															
Thoroughness	5.32 ± 0.98	3.97 ± 1.20 (p < 0.001)															
Control	5.19 ± 1.06	4.14 ± 1.10 (p < 0.05)															
Charman, Kavetski and Mueller (2017)	<ul style="list-style-type: none"> Materials: Computer-administered homicide case description with either DNA or eyewitness evidence that was either incriminating, exonerating or neutral. After making an initial rating of likelihood of guilt, all participants reviewed additional ambiguous alibi, facial composite, handwriting, and informant evidence in a randomized order. Explanatory Measures: Type of evidence (DNA or eyewitness) and incriminating, exonerating, or neutral nature of evidence (random assignment). 	Results provide evidence of confirmation bias in evaluation of ambiguous criminal evidence.															
N=89 police officers, mean experience 20 years, United States		Interaction between type of evidence and its interpretation – higher initial ratings of guilt with incriminating DNA evidence compared with incriminating eyewitness evidence; lower initial ratings of guilt with exonerating DNA compared with exonerating eyewitness evidence:															
Human Nature	<ul style="list-style-type: none"> Primary Outcome Measures: Initial ratings of probable guilt (1 to 100 scale) after reviewing case description; ratings of each additional type of evidence (1 to 7 Likert scale for strength of alibi; 1 to 100 scale for similarity of facial composite to suspect, similarity of handwriting to suspect, the extent to which informant evidence implicated suspect); final ratings of probable guilt (1 to 100 scale) after reviewing all of the additional evidence. 	<p>[Mean ± SD]</p> <table> <tr> <td>Initial Probable</td><td>DNA Evidence</td><td>Eyewitness</td></tr> <tr> <td>Guilt Rating for:</td><td></td><td></td></tr> <tr> <td>Incriminating</td><td>79.4 ± 26.8</td><td>52.3 ± 24.0</td></tr> <tr> <td>Neutral</td><td>31.7 ± 23.2</td><td>29.9 ± 19.5</td></tr> <tr> <td>Exonerating</td><td>12.2 ± 21.9</td><td>31.6 ± 21.4</td></tr> </table>	Initial Probable	DNA Evidence	Eyewitness	Guilt Rating for:			Incriminating	79.4 ± 26.8	52.3 ± 24.0	Neutral	31.7 ± 23.2	29.9 ± 19.5	Exonerating	12.2 ± 21.9	31.6 ± 21.4
Initial Probable	DNA Evidence	Eyewitness															
Guilt Rating for:																	
Incriminating	79.4 ± 26.8	52.3 ± 24.0															
Neutral	31.7 ± 23.2	29.9 ± 19.5															
Exonerating	12.2 ± 21.9	31.6 ± 21.4															
		Strength of initial guilt belief influenced subsequent evaluation of evidence, except for alibi strength:															
		Strength of initial guilt belief influenced subsequent evaluation of evidence, except for alibi strength:															
		Composite similarity $\beta = 0.637$, $SE = 0.075$ (p < 0.001)															
		Handwriting similarity $\beta = 0.604$, $SE = 0.075$ (p < 0.001)															
		Informant $\beta = 0.672$, $SE = 0.070$ (p < 0.001)															
		Alibi strength $\beta = 0.099$, $SE = 0.114$ (p = 0.39)															
		[β = relation between initial guilt belief and evidence evaluation]															
		[β = relation between initial guilt belief and evidence evaluation]															

Reference, Participants, Research Focus	Methods and Measures	Results																																													
		<p>The cumulative effect of ambiguous evidence was to increase perception of guilt: all final guilt ratings were higher than initial guilt ratings, except for those who began with incriminating DNA evidence, perhaps indicating a ceiling effect:</p> <table><tr><th>Final Probable</th><th colspan="2">[Mean ± SD]</th></tr><tr><td>Guilt Rating for:</td><td>DNA Evidence</td><td>Eyewitness</td></tr><tr><td>Incriminating</td><td>78.4 ± 14.9</td><td>62.0 ± 23.1</td></tr><tr><td>Neutral</td><td>53.7 ± 23.8</td><td>55.3 ± 21.8</td></tr><tr><td>Exonerating</td><td>39.0 ± 27.2</td><td>55.2 ± 25.3</td></tr></table>	Final Probable	[Mean ± SD]		Guilt Rating for:	DNA Evidence	Eyewitness	Incriminating	78.4 ± 14.9	62.0 ± 23.1	Neutral	53.7 ± 23.8	55.3 ± 21.8	Exonerating	39.0 ± 27.2	55.2 ± 25.3																														
Final Probable	[Mean ± SD]																																														
Guilt Rating for:	DNA Evidence	Eyewitness																																													
Incriminating	78.4 ± 14.9	62.0 ± 23.1																																													
Neutral	53.7 ± 23.8	55.3 ± 21.8																																													
Exonerating	39.0 ± 27.2	55.2 ± 25.3																																													
Dando and Ormerod (2017)	<ul style="list-style-type: none">Materials: Randomly selected decision logs from two police forces in which entries concerned a crime, detectives made clear a preference of possible action, and a reason was given to follow the course of action. Plotted case timelines noting when hypotheses were generated and tested.Explanatory Measures: Police officer experience level (mean 10 versus mean 2 years).	<p>Results suggest that use of decision logs varies by type of case and the officer involved, but this documentation reveals differences in investigative decision-making by professional experience level throughout an investigation.</p> <p>Similar number of log entries each week by experience level:</p> <table><tr><th colspan="2">[Mean ± SD]:</th></tr><tr><td>More experienced</td><td>8.19 ± 4.13</td></tr><tr><td>Less experienced</td><td>9.62 ± 3.30 [p = 0.45]</td></tr></table>	[Mean ± SD]:		More experienced	8.19 ± 4.13	Less experienced	9.62 ± 3.30 [p = 0.45]																																							
[Mean ± SD]:																																															
More experienced	8.19 ± 4.13																																														
Less experienced	9.62 ± 3.30 [p = 0.45]																																														
N=60 police decision logs, United Kingdom																																															
Individual Characteristics; Case-Specific	<ul style="list-style-type: none">Primary Outcome Measures: Number of log entries, length of investigation; analyzed number of hypotheses generated, number of evidence sources examined to support these hypotheses, and order in which hypotheses were generated; ratio of horizontal to vertical activity transitions (indicating number of hypotheses being examined: > 1 = multiple lines of inquiry; < 1 = “satisficing”, or focusing on a single line of inquiry).	<p>Number of hypotheses generated was highest in first quartile and higher for more experienced investigators:</p> <table><tr><th>Mean number of hypotheses</th><th>Q1</th><th>Q2</th><th>Q3</th><th>Q4</th></tr><tr><td>Experience > 5 years</td><td>2.84</td><td>1.17</td><td>0.79</td><td>0.55</td></tr><tr><td>Experience < 3 years</td><td>1.38</td><td>0.62</td><td>0.51</td><td>0.82</td></tr></table> <p>Mean number of evidence sources opened was highest in the beginning of the investigation, but there was little difference by experience level:</p> <table><tr><th>Number of evidence sources opened</th><th>Q1</th><th>Q2</th><th>Q3</th><th>Q4</th></tr><tr><td>Experience > 5 years</td><td>3.09</td><td>1.25</td><td>2.04</td><td>1.90</td></tr><tr><td>Experience < 3 years</td><td>3.96</td><td>2.35</td><td>1.76</td><td>1.20</td></tr></table> <p>Experienced investigators explored multiple hypotheses in the beginning and at the end of the investigation; less experienced investigators focused on a single hypothesis throughout the investigation:</p> <table><tr><th>Mean ratio of horizontal to vertical activity transitions</th><th>Q1</th><th>Q2</th><th>Q3</th><th>Q4</th></tr><tr><td>Experience > 5 years</td><td>1.29</td><td>0.86</td><td>0.77</td><td>1.32</td></tr><tr><td>Experience < 3 years</td><td>0.61</td><td>0.84</td><td>0.89</td><td>0.84</td></tr></table> <p>[> 1 = multiple lines of inquiry; < 1 = “satisficing”, or focusing on a single line of inquiry]</p>	Mean number of hypotheses	Q1	Q2	Q3	Q4	Experience > 5 years	2.84	1.17	0.79	0.55	Experience < 3 years	1.38	0.62	0.51	0.82	Number of evidence sources opened	Q1	Q2	Q3	Q4	Experience > 5 years	3.09	1.25	2.04	1.90	Experience < 3 years	3.96	2.35	1.76	1.20	Mean ratio of horizontal to vertical activity transitions	Q1	Q2	Q3	Q4	Experience > 5 years	1.29	0.86	0.77	1.32	Experience < 3 years	0.61	0.84	0.89	0.84
Mean number of hypotheses	Q1	Q2	Q3	Q4																																											
Experience > 5 years	2.84	1.17	0.79	0.55																																											
Experience < 3 years	1.38	0.62	0.51	0.82																																											
Number of evidence sources opened	Q1	Q2	Q3	Q4																																											
Experience > 5 years	3.09	1.25	2.04	1.90																																											
Experience < 3 years	3.96	2.35	1.76	1.20																																											
Mean ratio of horizontal to vertical activity transitions	Q1	Q2	Q3	Q4																																											
Experience > 5 years	1.29	0.86	0.77	1.32																																											
Experience < 3 years	0.61	0.84	0.89	0.84																																											

Reference, Participants, Research Focus	Methods and Measures	Results																											
Ditrich (2015)	<ul style="list-style-type: none">Materials: List of cognitive errors adapted from medical literature to the criminal justice context, presented verbally.	An “inside view” from experienced crime scene officers suggested that the most common and potentially detrimental cognitive errors included confirmation bias, anchoring, and shifting the burden of proof from the investigator to the suspect.																											
“small” number of experienced crime scene officers, Austria	<ul style="list-style-type: none">Explanatory Measures: Not applicable.																												
Human Nature	<ul style="list-style-type: none">Primary Outcome Measures: Opinions about frequency of appearance (5-point Likert scale from “never” to “very often”), as well as concepts selected as having the strongest adverse effect in practice.																												
Fahsing and Ask (2016)	<ul style="list-style-type: none">Materials: Two missing person case^d descriptions that did or did not contain a tipping point (decision to arrest a particular suspect), asked to write down investigative hypotheses and actions to be taken.Explanatory Measures: Tipping point, experience level (> 10 years as a detective or < 2 years as police officer with no detective work), type of training (based on country). Training in England: standardized four-step qualification system for detectives, comprehensive procedural guidelines, detailed routines for systematic reviews and knowledge-sharing, annual refreshment training. Training in Norway: on-the-job-learning, no nation-wide program or standardized requirements.	Presence of tipping point was not associated with generation of hypotheses or actions (“case-specific”), but researchers saw an interaction between experience level (“individual characteristics”) and type of training (“environment and culture”).																											
N=124 crime investigators, about 50% experienced and 50% novices, 63 from Norway, 61 from England																													
Environment and Culture; Individual Characteristics; Case-Specific	<ul style="list-style-type: none">Primary Outcome Measures: Proportion of “gold standard” set of hypotheses and actions generated (“gold standard” developed by expert panel of homicide investigators from Norway and England; see Fahsing and Ask, Journal of Investigative Psychology and Offender Profiling, 10, 155-165, 2013 for details).	Experienced English participants outperformed experienced and inexperienced Norwegian participants, who outperformed inexperienced English participants. Experience level was not associated with performance among Norwegian participants.																											
		[Mean ± SD]:																											
	Proportion of Generated Gold Standard																												
	<table><tr><th></th><th>Hypotheses</th><th>Actions</th></tr><tr><td>England</td><td></td><td></td></tr><tr><td>Experienced</td><td>0.72 ± 0.18</td><td>0.73 ± 0.10</td></tr><tr><td>Novices</td><td>0.28 ± 0.12</td><td>0.39 ± 0.11</td></tr><tr><td>(p-value)</td><td>(p < 0.001)</td><td>(p < 0.001)</td></tr><tr><td>Norway</td><td></td><td></td></tr><tr><td>Experienced</td><td>0.45 ± 0.16</td><td>0.64 ± 0.12</td></tr><tr><td>Novices</td><td>0.50 ± 0.20</td><td>0.61 ± 0.12</td></tr><tr><td>(p-value)</td><td>(p = 0.99)</td><td>(p = 0.99)</td></tr></table>		Hypotheses	Actions	England			Experienced	0.72 ± 0.18	0.73 ± 0.10	Novices	0.28 ± 0.12	0.39 ± 0.11	(p-value)	(p < 0.001)	(p < 0.001)	Norway			Experienced	0.45 ± 0.16	0.64 ± 0.12	Novices	0.50 ± 0.20	0.61 ± 0.12	(p-value)	(p = 0.99)	(p = 0.99)	
	Hypotheses	Actions																											
England																													
Experienced	0.72 ± 0.18	0.73 ± 0.10																											
Novices	0.28 ± 0.12	0.39 ± 0.11																											
(p-value)	(p < 0.001)	(p < 0.001)																											
Norway																													
Experienced	0.45 ± 0.16	0.64 ± 0.12																											
Novices	0.50 ± 0.20	0.61 ± 0.12																											
(p-value)	(p = 0.99)	(p = 0.99)																											
Fahsing and Ask (2017)	<ul style="list-style-type: none">Materials: Given cognitive aptitude test measuring inductive and deductive reasoning skills. Then given two missing person case descriptions that contained a tipping point (decision to arrest a particular suspect) either early or late in the vignette and asked to write down investigative hypotheses and actions to be taken.	Results suggest that inductive and deductive reasoning aptitude does not predict superior criminal investigation skills, and a case tipping point (arrest) does not necessarily lead to closed-mindedness.																											

Reference, Participants, Research Focus	Methods and Measures	Results
<p>N=166 police trainees, Norway</p> <p>Individual Characteristics; Case-Specific</p>	<ul style="list-style-type: none"> Explanatory Measures: Cognitive reasoning skills (inductive and deductive reasoning scores based on standardized test), location of tipping point. Primary Outcome Measures: Proportion of “gold standard” set of hypotheses and actions generated (“gold standard” developed by expert panel of homicide investigators from Norway and United Kingdom). 	<p>Overall, participants generated a higher proportion of criminal versus non-criminal explanations for the events (Mean $\pm$ SD: 46.7% $\pm$ 13.8 versus 19% $\pm$ 19.1).</p> <p>More hypotheses generated for Case A (Mean $\pm$ SD 49.2 $\pm$ 20.3) than for Case B (29.5 $\pm$ 10.8), $p > 0.001$. Because of this difference, other analyses were conducted separately for Case A and Case B.</p> <p>In regression analyses, no association was seen between the number of “gold standard” hypotheses generated in either variation of the cases and the inductive reasoning score or deductive reasoning score:</p> <p>Inductive reasoning score</p> <p>Case A $\beta = 0.04$, SE = 0.16, $p = 0.78$</p> <p>Case B $\beta = 0.04$, SE = 0.08, $p = 0.66$</p> <p>Deductive reasoning score</p> <p>Case A $\beta = 0.03$, SE = 0.15, $p = 0.85$</p> <p>Case B $\beta = 0.01$, SE = 0.08, $p = 0.872$</p> <p>Earlier “tipping point” (arrest) did not significantly decrease the number of “gold standard” hypotheses generated, but there were trends in that direction.</p>
<p>Groenendaal and Helsloot (2015)</p> <p>N=60 members of Command Core Teams, Netherlands</p> <p>Intervention</p>	<ul style="list-style-type: none"> Materials: Historical record of “the Schiedammer Park murder” in the Netherlands; semi-structured group interviews with leaders and coordinators of Command Core Teams from police forces across the country. Intervention: Criminal Investigation Reinforcement Programme, developed and implemented in response to a widely publicized wrongful conviction. Criminal Investigation Reinforcement Programme elements included the creation of Major Investigation Teams/Command Core Teams aimed at improving professionalization, communication, record-keeping around investigative decision-making, and critical reflection. Primary Outcome Measures: Investigators’ self-reported experiences with various elements of the new Programme, as well as “efficacy” (i.e., “number of solved crimes”) and “precaution” (i.e., “minimisation of the chance of wrongful conviction”). 	<p>Narrative description of policies produced in response to the Schiedammer Park murder and subsequent Posthumus Commission.</p> <p>Interviewees reported that the Major Investigation Team model improved investigations, but did not know whether more crimes were solved as a result. They reported both pros and cons to permanent positions, record-keeping about the hypotheses and scenarios decision-making process. They felt positively about the devil’s advocate system but found that it generally confirmed the direction already being pursued and did not identify flaws. They reported that, compared to previous years, the culture had become more open.</p> <p>The study authors’ identified tension between the concepts of efficacy and precaution, and concluded that the main result of the Programme was increased awareness about (but no measurable elimination of) tunnel vision.</p>

Reference, Participants, Research Focus	Methods and Measures	Results												
Jones et al. (2008)	<ul style="list-style-type: none"> Materials: Given first two-months-worth of material of a concluded murder investigation; asked to review the investigation and identify positive and negative aspects. Experimental group given Review Tool (developed based on a literature review and in accordance with the Association of Chief Police Officers (ACPO) 2006 <i>Murder Investigation Manual</i> (MIM)) providing best practices for 31 categories of investigative activities; control group received no instructions but could consult ACPO MIM or other materials. Intervention: Review Tool or control group (random assignment); officer experience level (experienced = senior investigating officers, inexperienced = little or no murder investigation experience). Primary Outcome Measures: Amount of information produced in each investigation review; content analysis, with each comment rated for usefulness (1 to 7 Likert scale), amount of time to complete the review. 	<p>Review Tool produced increased quantity (amount not specified) and quality (37% higher) of information, and took longer to complete (approximately 33% longer).</p> <p>Experienced officers in both the intervention and control groups produced higher quantity and quality work</p> <p>Results show that both experience and the tested Review Tool helped (though mechanism is unclear – Review Tool might help officers think more thoroughly and critically because of the <i>content and structure</i> of the tool, or might simply succeed by <i>slowing officers down</i> so they can think more thoroughly and critically).</p>												
N=12 police officers, 6 experienced and 6 inexperienced investigators, United Kingdom														
Intervention														
Kerstholt and Eikelboom, (2007)	<ul style="list-style-type: none"> Materials: Two realistic case scenarios (Case 1: possible sex trafficking; Case 2: disappearance of a young woman). Half of participants received a plausible, but not the most likely (based on pilot testing), prior interpretation for each case (Case 1: “Rodriguez” played a key role, Case 2: missing woman’s father played role in disappearance). Explanatory Measures: Knowledge of prior plausible but unlikely interpretation or not; experience level (mean 7 years versus 7 months). Primary Outcome Measures: For Case 1, description of role of each person and most likely scenario, ranking of sources of information and importance, noted missing information, and suggestions for further investigation. For Case 2, number and type of possible explanations generated, conclusion about most likely hypothesis, suggestions for further investigation. 	<p>Groupthink, confirmation bias demonstrated. Results show that analysts who are privy to an investigative team’s working hypothesis will suggest that hypothesis as the most likely at a higher rate than analysts who merely get access to the facts of the case but no prior interpretation. Novices and experts performed the same way.</p> <p>Case 1: Knowledge of prior interpretation was associated with identifying “Rodriguez” as a “key player,” evidence mentioned and suggestions for further investigation, but not with ranking of information sources, or reports of missing information. Experience level was not associated with decision-making.</p>												
N=38 crime analysts, Netherlands														
Human Nature: Individual Characteristics														
		<table> <tr> <th></th><th>Prior Interpretation</th><th>No Prior Interpretation</th></tr> <tr> <td>“Rodriguez = key player”</td><td>56%</td><td>25% (p = 0.06)</td></tr> <tr> <td>Mentioned information from prior interpretation</td><td>89%</td><td>35% (p = 0.002)</td></tr> <tr> <td>Rodriguez-further Investigation</td><td>78%</td><td>35% (p = 0.01)</td></tr> </table>		Prior Interpretation	No Prior Interpretation	“Rodriguez = key player”	56%	25% (p = 0.06)	Mentioned information from prior interpretation	89%	35% (p = 0.002)	Rodriguez-further Investigation	78%	35% (p = 0.01)
	Prior Interpretation	No Prior Interpretation												
“Rodriguez = key player”	56%	25% (p = 0.06)												
Mentioned information from prior interpretation	89%	35% (p = 0.002)												
Rodriguez-further Investigation	78%	35% (p = 0.01)												

Reference, Participants, Research Focus	Methods and Measures	Results																								
Marksteiner et al. (2010) N=107 police trainees, Sweden Human Nature	<ul style="list-style-type: none">Materials: Computer-administered homicide case^c description suggesting guilt or innocence of suspect, plus incriminating or exonerating eyewitness evidence.Explanatory Measures: Eyewitness evidence consistent or inconsistent with initial beliefs (random assignment).Primary Outcome Measures: Ratings of strength of evidence, probability of guilt (both 1 to 9 Likert scales), conviction decision (dichotomous) before and after review of witness information; ratings of reliability of the specific witness evidence provided and of the reliability of witness evidence in general (1 to 9 Likert scales) after review of witness information.	<p>Case 2: Knowledge of prior interpretation was associated with inclusion of father scenario as a possible explanation and as the most likely explanation, and as suggestion for further investigation, but not with number of explanations generated. Experience level was not associated with decision-making.</p> <table><thead><tr><th></th><th colspan="2">Prior</th></tr><tr><th></th><th>Interpretation</th><th>No Prior Interpretation</th></tr></thead><tbody><tr><td>Father-mentioned as explanation</td><td>94%</td><td>45% (p = 0.007)</td></tr><tr><td>Father-most likely explanation</td><td>39%</td><td>10% (p = 0.06)</td></tr><tr><td>Father-further Investigation</td><td>56%</td><td>20% (p = 0.03)</td></tr></tbody></table> <p>Asymmetrical skepticism among participants with a guilty hypothesis: Incriminating witness evidence (case-specific, and in general) viewed most favorably by those who initially considered the suspect guilty; the same pattern did not emerge for those who initially considered the suspect innocent.</p>		Prior			Interpretation	No Prior Interpretation	Father-mentioned as explanation	94%	45% (p = 0.007)	Father-most likely explanation	39%	10% (p = 0.06)	Father-further Investigation	56%	20% (p = 0.03)									
		Prior																								
		Interpretation	No Prior Interpretation																							
Father-mentioned as explanation	94%	45% (p = 0.007)																								
Father-most likely explanation	39%	10% (p = 0.06)																								
Father-further Investigation	56%	20% (p = 0.03)																								
Rassin (2010) Study 1: N=118 police officers, district attorneys, and judges, Netherlands Human Nature	<ul style="list-style-type: none">Materials: Homicide case^a description that included information about a named female suspect's motive and information about an alternative male suspect.Explanatory Measures: Case scenario version (random assignment).Primary Outcome Measures: Ratings of probability of guilt (0 to 100), and decision to convict (dichotomous); ratings of degree to which each of 20 case details supported innocence or guilt of the named suspect (-5 to 5 Likert scale), combined into a composite score.	<table><thead><tr><th></th><th colspan="2">[Mean $\pm$ SD]: Witness Evidence:</th></tr><tr><th>Reliability Rating for:</th><th>Incriminating</th><th>Exonerating</th></tr></thead><tbody><tr><td>Specific Case</td><td></td><td></td></tr><tr><td>Guilt</td><td>6.59 $\pm$ 1.67</td><td>4.04 $\pm$ 1.76 (p < 0.004)</td></tr><tr><td>Innocence</td><td>4.68 $\pm$ 2.26</td><td>4.43 $\pm$ 1.79 (p = 0.27)</td></tr><tr><td>In General</td><td></td><td></td></tr><tr><td>Guilt</td><td>6.22 $\pm$ 0.93</td><td>4.92 $\pm$ 1.38 (p < 0.004)</td></tr><tr><td>Innocence</td><td>4.96 $\pm$ 1.50</td><td>4.61 $\pm$ 1.32 (p = 0.12)</td></tr></tbody></table> <p>Results provided some evidence in support of and some evidence against the presence of confirmation bias.</p> <p>Case scenario version was associated with ratings of probable guilt (p = 0.03), but not with decision to convict (p = 0.86) or evaluation of case details (p = 0.15)</p>		[Mean $\pm$ SD]: Witness Evidence:		Reliability Rating for:	Incriminating	Exonerating	Specific Case			Guilt	6.59 $\pm$ 1.67	4.04 $\pm$ 1.76 (p < 0.004)	Innocence	4.68 $\pm$ 2.26	4.43 $\pm$ 1.79 (p = 0.27)	In General			Guilt	6.22 $\pm$ 0.93	4.92 $\pm$ 1.38 (p < 0.004)	Innocence	4.96 $\pm$ 1.50	4.61 $\pm$ 1.32 (p = 0.12)
	[Mean $\pm$ SD]: Witness Evidence:																									
Reliability Rating for:	Incriminating	Exonerating																								
Specific Case																										
Guilt	6.59 $\pm$ 1.67	4.04 $\pm$ 1.76 (p < 0.004)																								
Innocence	4.68 $\pm$ 2.26	4.43 $\pm$ 1.79 (p = 0.27)																								
In General																										
Guilt	6.22 $\pm$ 0.93	4.92 $\pm$ 1.38 (p < 0.004)																								
Innocence	4.96 $\pm$ 1.50	4.61 $\pm$ 1.32 (p = 0.12)																								
Rassin (2018b)	<ul style="list-style-type: none">Materials: Homicide case^a description that included either information about a named female suspect's motive or information about an alternative male suspect.	<p>Intervention was promising but no control group incorporated into study.</p>																								

Reference, Participants, Research Focus	Methods and Measures	Results
Study 2: <i>N</i> =45 police officers, district attorneys, and judges, Netherlands	<ul style="list-style-type: none"> Intervention: Pen and paper exercise rating degree to which each of 20 case details indicated the named suspect's guilt (10-point scale; 0 = exonerating, 10 = incriminating), probability of guilt (0 = definitely innocent to 100 = definitely guilty), and conviction decision (yes/no). Then told to imagine that the unnamed suspect had committed the crime and asked to rate the degree to which each of 20 case details indicated the alternative suspect's guilt (10-point scale; 0 = exonerating, 10 = incriminating). After this exercise, rated probability of named suspect's guilt a second time (0 = definitely innocent to 100 = definitely guilty), and made a final conviction decision (yes/no). Primary Outcome Measures: Guilt estimates and conviction rates before and after pen and paper and imagination exercise. 	<p>Similar ratings of the extent to which the evidence fit named suspect hypothesis and alternative suspect hypothesis [Mean $\pm$ SD]:</p> <p>Named suspect 4.91 $\pm$ 0.72 Alternative suspect 5.19 $\pm$ 1.16 [p = 0.21]</p> <p>Guilty estimate decreased with intervention:</p> <p>Before 45.67 $\pm$ 22.1 After 32.44 $\pm$ 20.13 [p < 0.001]</p> <p>Decision to convict decreased with intervention:</p> <p>Before 22% of participants After 7% of participants [p < 0.04]</p>
Salet and Terpstra (2014)	<ul style="list-style-type: none"> Materials: In response to wrongful conviction in a child sex abuse and murder case in the Netherlands, a national commission recommended use of a critical review protocol (including the use of contrarians) for complex criminal investigations. Researchers assessed the implementation and results of this critical review procedure through a review of real case files and review dossiers from five different police forces and interviews with lead investigators and contrarians. Intervention: Critical review procedures introduced into policework. 	<p>Critical reviews had concrete effects on criminal investigations but did not radically change the direction of any case.</p> <p>Most police forces were practicing critical reviews in 2011, but variability in the role of contrarians between forces. Some contrarians used a closeness strategy (actively involved in the investigation), while others used a distance strategy (reviewing decisions after they were made). Different contrarians played different roles:</p> <ul style="list-style-type: none"> - ritualists (superficial review) - proceduralists (concentrate of reviewing procedures) - criminal investigators (concentrate on reviewing substantive elements) - coach (concentrate on improving quality of leadership and social relations within the team) - devil's advocate (challenge with alternative interpretations and hypotheses)
<i>N</i> =26 case files and dossiers; interviews with 47 leaders of investigative teams and "contrarians," Netherlands	<ul style="list-style-type: none"> Primary Outcome Measures: Qualitative comparison of case files and review dossiers and information from lead investigators and contrarians. 	
Intervention		

Reference, Participants, Research Focus	Methods and Measures	Results
Wallace (2015)	<ul style="list-style-type: none"> Materials: Computer-administered sexual assault case vignette (with either a child or adult victim), including 10 items of evidence (presented either simultaneously, sequentially, or in reverse-sequential order). Rated confidence in guilt or innocence after every piece of evidence and provided an overall, final decision about guilt or innocence. 	Results did not support the hypothesis that extreme emotion did not impact guilt judgements (or that victim age is not an effective manipulation of extreme emotion) as the age of the sexual assault victim did not influence guilt judgments.
N=166 police officers (basic training recruits, patrol officers, and criminal investigators), United States	<ul style="list-style-type: none"> Explanatory Measures: Presumed emotional impact of case scenario version (random assignment), professional expertise (recruit, patrol, investigator), evidence presentation order – exculpatory evidence presented early or late (random assignment). 	The order of evidence presentation did matter, demonstrating confirmation bias. When exculpatory evidence was viewed prior to inculpatory evidence, guilt belief scores decreased significantly.
Human Nature, Individual Characteristics; Case-Specific	<ul style="list-style-type: none"> Primary Outcome Measures: Confidence in suspect's guilt or innocence (0 to 10 Likert scale), final decision about guilt or innocence (yes/no). 	Finally, confirmation bias was greater among police recruits (i.e., those with the least professional experience).

^aHomicide case vignette was the same as the others with this designation

^bAssault case vignette was the same as the others with this designation

^cHomicide case vignette was the same as the others with this designation

^dMissing person case vignettes were the same as others with this designation

^eThe second study reported in this article used undergraduate student participants

Acknowledgments Thank you to Dr. Karen Amendola (Police Foundation), Ms. Prahelika Gadtaula (Innocence Project), and Dr. Kim Rossmo (Texas State University) for their thoughtful reviews of earlier drafts.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical Statement This article does not contain any studies with human participants or animals performed by either of the authors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ask K, Granhag PA (2005) Motivational sources of confirmation bias in criminal investigations: the need for cognitive closure. *J Investig Psychol Offender Profiling* 2(1):43–63. <https://doi.org/10.1002/jip.19>
- Ask K, Granhag PA (2007a) Hot cognition in investigative judgments: the differential influence of anger and sadness. *Law Hum Behav* 31(6):537–551. <https://doi.org/10.1007/s10979-006-9075-3>
- Ask K, Granhag PA (2007b) Motivational bias in criminal investigators' judgments of witness reliability. *J Appl Soc Psychol* 37(3):561–591. <https://doi.org/10.1111/j.1559-1816.2007.00175.x>
- Ask K, Granhag PA, Rebelius A (2011a) Investigators under influence: how social norms activate goal-directed processing of criminal evidence. *Appl Cogn Psychol* 25(4):548–553. <https://doi.org/10.1002/acp.1724>
- Ask K, Rebelius A, Granhag PA (2008) The 'elasticity' of criminal evidence: a moderator of investigator bias. *Appl Cogn Psychol* 22(9):1245–1259. <https://doi.org/10.1002/acp.1432>
- Ask K, Reinhard M-A, Marksteiner T, Granhag PA (2011b) Elasticity in evaluations of criminal evidence: exploring the role of cognitive dissonance. *Leg Criminol Psychol* 16:289–306
- Blair IV, Judd CM, Chappleau KM (2004) The influence of Afrocentric facial features in criminal sentencing. *Psychol Sci* 15(10):674–679. <https://doi.org/10.1111/j.0956-7976.2004.00739.x>
- Chanin J, Welsh M, Nurge D (2018) Traffic enforcement through the lens of race: a sequential analysis of post-stop outcomes in San Diego. *California Criminal Justice Policy Review* 29(6–7):561–583. <https://doi.org/10.1177/0887403417740188>
- Charman SD, Carbone J, Kekessie S, Villalba DK (2015) Evidence evaluation and evidence integration in legal decision-making: order of evidence presentation as a moderator of context effects. *Appl Cogn Psychol* 30(2):214–225. <https://doi.org/10.1002/acp.3181>
- Charman SD, Kavetski M, Mueller DH (2017) Cognitive bias in the legal system: police officers evaluate ambiguous evidence in a belief-consistent manner. *J Appl Res Mem Cogn* 6(2):193–202. <https://doi.org/10.1016/j.jarmac.2017.02.001>
- Cooper GS, Meterko V (2019) Cognitive bias research in forensic science: a systematic review. *Forensic Sci Int* 297:35–46. <https://doi.org/10.1016/j.forsciint.2019.01.016>
- Correll J, Park B, Judd CM, Wittenbrink B (2002) The police officer's dilemma: using ethnicity to disambiguate potentially threatening individuals. *J Pers Soc Psychol* 83(6):1314–1329. <https://doi.org/10.1037/0022-3514.83.6.1314>
- Dando CJ, Ormerod TC (2017) Analyzing decision logs to understand decision making in serious crime investigations. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 59(8):1188–1203. <https://doi.org/10.1177/0018720817727899>
- Department of Justice (2017) Retrieved from: <https://www.justice.gov/file/923201/download>
- Ditrich H (2015) Cognitive fallacies and criminal investigations. *Sci Justice* 55:155–159
- Dror IE (2009) How can Francis Bacon help forensic science? The four idols of human biases. *Jurimetrics: The Journal of Law, Science, and Technology* 50(1):93–110
- Dror IE (2015) Cognitive neuroscience in forensic science: understanding and utilizing the human element. *Philos Trans R Soc B* 370(1674). <https://doi.org/10.1098/rstb.2014.0255>
- Dror IE (2017) Human expert performance in forensic decision making: seven different sources of bias. *Aust J Forensic Sci* 49(5):541–547. <https://doi.org/10.1080/00450618.2017.1281348>
- Dror IE (2020) Cognitive and human factors in expert decision making: six fallacies and the eight sources of bias. *Anal Chem* 92(12):7998–8004. <https://pubs.acs.org/doi/10.1021/acs.analchem.0c00704>
- Dror IE, Morgan RM, Rando C, Nakhaeizadeh S (2017) Letter to the Editor: The bias snowball and the bias cascade effects: two distinct biases that may impact forensic decision making. *J Forensic Sci* 62(3):832–833. <https://doi.org/10.1111/1556-4029.13496>
- Doyle JM (2012) Learning about learning from error. *Police Foundation* 14:1–16
- Doyle JM (2014) NIJ's sentinel events initiative: looking back to look forward. *National Institute of Justice Journal* 273:10–14. <https://www.ncjrs.gov/pdffiles1/nij/244144.pdf>
- Eberhardt JL, Davies PG, Purdie-Vaughns VJ, Johnson SL (2006) Looking deathworthy: perceived stereotypicality of Black defendants predicts capital-sentencing outcomes. *Psychol Sci* 17(5):383–386. <https://doi.org/10.1111/j.1467-9280.2006.01716.x>
- Eerland A, Post LS, Rassin E, Bouwmeester S, Zwaan RA (2012) Out of sight, out of mind: the presence of forensic evidence counts more than its absence. *Acta Physiol (Oxf)* 140(1):96–100. <https://doi.org/10.1016/j.actpsy.2012.02.006>
- Eerland A, Rassin E (2012) Biased evaluation of incriminating and exonerating (non)evidence. *Psychol Crime Law* 18(4):351–358. <https://doi.org/10.1080/1068316X.2010.493889>
- Fahsing I, Ask K (2013) Decision making and decisional tipping points in homicide investigations: an interview study of British and Norwegian detectives. *J Investig Psychol Offender Profiling* 10(2):155–165. <https://doi.org/10.1002/jip.1384>
- Fahsing I, Ask K (2016) The making of an expert detective: the role of experience in English and Norwegian police officers' investigative decision-making. *Psychol Crime Law* 22(3):203–223. <https://doi.org/10.1080/1068316X.2015.1077249>
- Fahsing IA, Ask K (2017) In search of indicators of detective aptitude: police recruits' logical reasoning and ability to generate investigative hypotheses. *J Police Crim Psychol* 33(1):21–34. <https://doi.org/10.1007/s11896-017-9231-3>
- Findley KA (2012) Tunnel vision. In B. L. Cutler (Ed.), *Conviction of the innocent: Lessons from psychological research* (pp. 303–323). American Psychological Association. <https://doi.org/10.1037/13085-014>

- Fischhoff B (1975) Hindsight does not equal foresight: the effect of outcome knowledge on judgment under uncertainty. *J Exp Psychol Hum Percept Perform* 1(3):288–299
- Gilovich T, Griffin DW, Kahneman D (Eds.) (2002) *Heuristics and biases: The psychology of intuitive judgment*. Cambridge, U.K.; New York: Cambridge University Press
- Greenspan R, Scurich N (2016) The interdependence of perceived confession voluntariness and case evidence. *Law Hum Behav* 40(6):650–659. <https://doi.org/10.1037/lhb0000200>
- Greenwald AG, McGhee DE, Schwartz JLK (1998) Measuring individual differences in implicit cognition: the implicit association test. *J Pers Soc Psychol* 74(6):1464–1480
- Groenendaal J, Helsloot I (2015) Tunnel vision on tunnel vision? A preliminary examination of the tension between precaution and efficacy in major criminal investigations in the Netherlands. *Police Pract Res* 16(3):224–238
- Haas HS, Pisarszewska Fuerst M, Tönz P, Gubser-Ernst J (2015) Analyzing the psychological and social contents of evidence-experimental comparison between guessing, naturalistic observation, and systematic analysis. *J Forensic Sci* 60(3):659–668. <https://doi.org/10.1111/1556-4029.12703>
- Hasel LE, Kassin SM (2009) On the presumption of evidentiary independence. *Psychol Sci* 20(1):122–126
- Jones D, Grieve J, Milne B (2008) Reviewing the reviewers: a tool to aid homicide reviews. *The Journal of Homicide and Major Incident Investigation* 4(2):59–70
- Kerstholt JH, Eikelboom AR (2007) Effects of prior interpretation on situation assessment in crime analysis. *J Behav Decis Mak* 20(5):455–465. <https://doi.org/10.1002/bdm.570>
- MacFarlane BA (2008) Wrongful convictions: the effect of tunnel vision and predisposing circumstances in the criminal justice system. Goudge Inquiry Research Paper. https://www.attorneygeneral.jus.gov.on.ca/inquiries/goudge/policy_research/pdf/Macfarlane_Wrongful-Convictions.pdf
- Marksteiner T, Ask K, Reinhard M-A, Granhag PA (2010) Asymmetrical scepticism towards criminal evidence: the role of goal- and belief-consistency. *Appl Cogn Psychol* 25(4):541–547. <https://doi.org/10.1002/acp.1719>
- National Institute of Justice (2017) Retrieved from: <https://nij.ojp.gov/topics/articles/sentinel-events-initiative>
- Nickerson RS (1998) Confirmation bias: a ubiquitous phenomenon in many guises. *Rev Gen Psychol* 2(2):175–220
- NYCLU (2011) NYCLU Stop-And-Frisk Report 2011. New York Civil Liberties Union
- O'Brien BM (2007) Confirmation bias in criminal investigations: an examination of the factors that aggravate and counteract bias. ProQuest Information & Learning, US. (2007–99016–280)
- O'Brien B (2009) Prime suspect: an examination of factors that aggravate and counteract confirmation bias in criminal investigations. *Psychol Public Policy Law* 15(4):315–334. <https://doi.org/10.1037/a0017881>
- Owens E, Kerrison EM, Da Silveira BS (2017) Examining racial disparities in criminal case outcomes among indigent defendants in San Francisco. Quattrone Center for the Fair Administration of Justice
- Owens E, Weisburd D, Amendola KL, Alpert GP (2018) Can you build a better cop?: Experimental evidence on supervision, training, and policing in the community. *Criminol Public Policy* 17(1):41–87. <http://doi.wiley.com/10.1111/1745-9133.12337>
- Pearson AR, Dovidio JF, Gaertner SL (2009) The nature of contemporary prejudice: insights from aversive racism. *Soc Pers Psychol Compass* 3(3):314–338. <https://doi.org/10.1111/j.1751-9004.2009.00183.x>
- Price HL, Dahl LC (2014) Order and strength matter for evaluation of alibi and eyewitness evidence: Recency effects. *Appl Cogn Psychol* 28(2):143–150. <https://doi.org/10.1002/acp.2983>
- Rassin E (2010) Blindness to alternative scenarios in evidence evaluation. *J Investig Psychol Offender Profiling* 7:153–163. <https://doi.org/10.1002/jip.116>
- Rassin E (2018a) Fundamental failure to think logically about scientific questions: an illustration of tunnel vision with the application of Wason's Card Selection Test to criminal evidence. *Appl Cogn Psychol* 32(4):506–511. <https://doi.org/10.1002/acp.3417>
- Rassin E (2018b) Reducing tunnel vision with a pen-and-paper tool for the weighting of criminal evidence. *J Investig Psychol Offender Profiling* 15(2):227–233. <https://doi.org/10.1002/jip.1504>
- Rassin E, Eerland A, Kuijpers I (2010) Let's find the evidence: an analogue study of confirmation bias in criminal investigations. *J Investig Psychol Offender Profiling* 7:231–246
- Richardson LS (2017) Systemic triage: Implicit racial bias in the criminal courtroom. *Yale Law J* 126(3):862–893
- Rossmo DK, Pollock JM (2019) Confirmation bias and other systemic causes of wrongful convictions: a sentinel events perspective. *Northeastern University Law Review* 11(2):790–835
- Salet R, Terpstra J (2014) Critical review in criminal investigation: evaluation of a measure to prevent tunnel vision. *Policing* 8(1):43–50. <https://doi.org/10.1093/police/pat039>
- Simon D, Snow CJ, Read SJ (2004) The redux of cognitive consistency theories: evidence judgments by constraint satisfaction. *J Pers Soc Psychol* 86(6):814–837. <https://doi.org/10.1037/0022-3514.86.6.814>
- Snook B, Cullen RM (2008) Bounded rationality and criminal investigations: has tunnel vision been wrongfully convicted? In D. K. Rossmo (Ed.), *Criminal Investigative Failures* (pp. 71–98). Taylor & Francis
- Staats C, Capatosto K, Tenney L, Mamo S (2017) State of the science: implicit bias review. The Kirwan Institute
- Stoel R, Berger C, Kerkhoff W, Mattijssen E, Dror I (2014) Minimizing contextual bias in forensic casework. In M. Hickman & K. Strom (Eds.), *Forensic science and the administration of justice: Critical issues and directions* (pp. 67–86). SAGE
- Tversky A, Kahneman D (1973) Availability: a heuristic for judging frequency and probability. *Cogn Psychol* 5:207–232
- Tversky A, Kahneman D (1974) Judgment under uncertainty: heuristics and biases. *Science* 185(4157):1124–1131
- Voigt R, Camp NP, Prabhakaran V, Hamilton WL, Hetey RC, Griffiths CM, Eberhardt JL (2017) Language from police body camera footage shows racial disparities in officer respect. *Proc Natl Acad Sci* 114(25):6521–6526. <https://doi.org/10.1073/pnas.1702413114>
- Wallace WA (2015) The effect of confirmation bias in criminal investigative decision making. ProQuest Information & Learning, US. (2016–17339–014)
- Wastell C, Weeks N, Wearing A, Duncan P (2012) Identifying hypothesis confirmation behaviors in a simulated murder investigation: implications for practice. *J Investig Psychol Offender Profiling* 9:184–198
- Zalman M, Larson M (2016) Elephants in the station house: Serial crimes, wrongful convictions, and expanding wrongful conviction analysis to include police investigation. Retrieved from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2716155
- Zapf PA, Dror IE (2017) Understanding and mitigating bias in forensic evaluation: lessons from forensic science. *Int J Forensic Ment Health* 16(3):227–238. <https://doi.org/10.1080/14999013.2017.1317302>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.