Eyewitness Identification Decisions and Metacognition:

Familiarity and Recollection in Simultaneous Lineups

Nicola Guerin, Grad Dip Psych, BA (Hons)

School of Psychology

Faculty of Social and Behavioural Sciences

A thesis submitted to Flinders University for the degree of Doctor of Philosophy,

February, 2017

TABLE OF CONTENTS

CHAPTER 6

CHAPTER 7

APPENDIX 1

APPENDIX 2

APPENDIX 3

APPENDIX 4

List of Figures

List of Tables

SUMMARY

I proposed a *best-match-first dual-process framework* for the analysis of simultaneous lineup decision-making by eyewitnesses. The framework combines 1) a best-match-first account of the component recognition decisions that contribute to a simultaneous identification response; and 2) a continuous dual-process account of memory. Eyewitnesses make simultaneous lineup decisions in two steps: first, selecting the lineup-member who best matches memory for the offender; before, second, deciding whether or not the best-match is the offender. This decision order encourages over-reliance on the global familiarity of lineup members, and neglect of recollected detail of the offender, leaving eyewitnesses susceptible to falsely identifying an innocent suspect when the offender is not present. Hence, I developed and tested a novel *presence-recollection lineup* to provide structural and metacognitive cues for greater recollection to increase the reliability of eyewitness identification decisions.

Eyewitness decision and memory processes were investigated in both a face recognition mini-lineup paradigm, and an eyewitness identification paradigm. In two experiments (Chapter 3, Experiments 1 & 2), patterns of response latency in the mini-lineup paradigm matched framework predictions for a best-match-first decision order. I tested the efficacy of reversing the best-match-first decision order for improving recollection with a modified *presence-first* procedure. Participants first indicated 1) whether or not the offender was in the lineup, before 2) identifying the offender, if present. In the modified procedure, participants better differentiated guilty from innocent suspects than in a standard simultaneous procedure. However, this improvement was only evident when participants in the presence-first procedure did not know that they would be asked to make the second offender-selection decisions.

In real situations, eyewitnesses will anticipate being asked to identify a guilty offender, if present. Therefore, in four mini-lineup experiments (Chapter 4, Experiments 1-4), I tested the efficacy of metacognitive instructions about the phenomenological qualities of memory that are most diagnostic of accuracy (i.e., detail, vividness and clarity) for encouraging recollection. Instructions were effective only when combined with a presence-first procedure (the presence-recollection lineup), that structurally disrupted best-match-first decision-making.

In two further mini-lineup experiments, more direct evidence for the memorial basis of simultaneous lineup decisions was found. Participants used recollection, when available, for standard simultaneous lineups, but were also willing to make (less accurate) familiarity-based positive identifications when objectively less recollection was available (Chapter 5). A similar pattern was evident for the presence-recollection procedure (Chapter 6), but with a tendency toward more appropriate weighting of familiarity and recollection than in the standard procedure. A lower overall likelihood of false identification was evident, with enhanced metacognitive monitoring and use of recollection.

Finally, in a more realistic identification paradigm (Chapter 7), the presence-recollection lineup did not demonstrate greater discriminability than either standard simultaneous or sequential procedures. If scant recollection was available, recall-to-reject, rather than post-retrieval memory editing mechanisms, is likely to have produced the presence-recollection advantage observed with mini-lineups. However, if eyewitnesses' willingness to choose was resistant to instructions, more radical structural change might be needed to improve the memorial basis of identification decisions in real criminal investigations. Overall, results demonstrate the value of extending basic theory to the understanding of problems with complex applied identification tasks.

DECLARATION

I certify that this thesis does not contain any material that has been accepted for the award of any other degree or diploma; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person, except where due reference is made in the text of the thesis or notes.

Nicola Guerin, Grad Dip (Psych), BA (Hons)

ACKNOWLEDGEMENTS

Thank you, Nathan Weber and Ruth Horry, for your practical and intellectual guidance and supervision.

Thank you, Neil Brewer, for your advice, assistance, and access to the Flinders Eyewitness Memory Lab. And thank you to the hundreds of people who took part in my studies. Without these facilities, and the time and interest of participants, the work completed for this thesis would not have been possible.[1]

Thank you, Paul Williamson, Jason McCarley, Melanie Takarangi, and Julie Mattiske for your advice and encouragement.

Thank you, everyone I have worked with in the School of Psychology at Flinders University.

Thank you, to the PhD students who have shared this time and work with me. Special thanks to Sarah Fletcher, Rosemary Walsh, Stacey Taylor-Aldridge, Caitlin Hitchcock, Gemma Sharp, Jennifer Mills, Jane Cooper, Jordan McLean, and Kate Coden, for sharing the everyday stuff with me along the way. A more hard-working and lovely bunch would be difficult to find.

Thank you, family and friends, for your love and support, and being the finer things in life.

An enormous and special thank you to my daughter, Ruby. Thank you for being by my side, and for everything, always. You're the best. Love you.

[1] This research was supported by funding from the School of Psychology, Flinders University, and Australian Research Council Discovery Grant DP150101905 to N. Brewer, N. Weber, & G. L. Wells.

There is nothing more deceptive than an obvious fact.

- Sir Arthur Conan Doyle

CHAPTER 1

General Introduction

Richard Alexander. Nathan Brown. Anthony Capozzi. Joseph Fears Jr.. Joe Jones. Each spent between five and 25 years in prison, wrongfully convicted on the basis of eyewitness misidentification, before being exonerated by DNA evidence (Innocence Project, 2016). Just five cases out of 337 exonerations in the U.S. since 1992. In 238 of these cases, an innocent suspect was misidentified as the culprit, sometimes by multiple witnesses. In many cases, eyewitness misidentification was the only evidence for conviction and was argued to outweigh strong evidence against guilt. The underlying incidence of wrongful convictions in which misidentification has played a role is unknown. DNA exoneration is only possible when biological evidence was collected, and this usually applies only in the most serious of offences. Further exonerations in the U.S. as a result of appeals based on new non-DNA evidence from serious crimes total 1410 since 1989 (National Registry of Exonerations, 2016). But police also make routine use of identification tests to guide the course of criminal investigation and prosecution for less serious offences (Wogalter, Malpass, & McQuiston, 2004). For cases when potentially exonerating evidence is not available, it is even more difficult to estimate the numbers of wrongful convictions to which eyewitness misidentification has contributed (e.g., Dioso-Villa, 2015).

While an accused person is protected by the presumption of innocence while on trial, this is reversed following a conviction. Convicted defendants are presumed guilty, and the right to appeal is usually hard won. If an appeal is lost, the prospects of a further appeal are radically diminished, even when new evidence emerges. In many jurisdictions, appeal is only possible for the most serious of offences, and when stringent criteria have been met. Further, the burden of appeal is on the defendant, who will frequently lack the resources necessary to pursue it (e.g., Australian Human Rights Commission, 2011; Sangha & Moles, 2012). Therefore, many cases of wrongful conviction based on misidentification might never be challenged. In England, the establishment of an independent Criminal Cases Review Commission in 1997 with powers to investigate possible miscarriages of justice and approve appeals, saw a dramatic increase in the number of convictions found to be unsafe (Criminal Cases Review Commission, 2016a). Each year, the Commission elevates substantial numbers of cases to appeals courts, with 70% of these resulting in quashed convictions, including cases of eyewitness misidentification (Criminal Cases Review Commission, 2016b). In jurisdictions where the right of appeal is more limited, evidence of potential misidentification and wrongful convictions is rarely revisited. Thus, the incidence of risk to innocent suspects and the system variables that influence it are hidden.

While the fallibility of eyewitness memory is well known, high numbers of identification procedures continue to be carried out each year in many jurisdictions, with the evidence often central to a prosecution case. The probable base rate of suspect guilt in police lineups will vary with offence type, jurisdiction, and time, and is not well tracked or analysed (e.g., Wells, Yang, & Smalarz, 2015). However, eyewitness responses to police lineups include surprisingly high rates of identification of known-innocent filler lineup members (e.g., Horry, Memon, Wright, & Milne, 2011; Wells, Steblay, & Dysart, 2015). Similarly, rates of false identification of innocent suspects in field and lab studies of mock crimes are substantial across varying conditions (e.g., Cutler & Penrod, 1995; Fitzgerald, Price, Oriet, & Charman, 2013; Steblay, Dysart, & Wells, 2011). The risk of misidentification for an innocent suspect who finds themselves in a police lineup, then, should not be underestimated. Further, we know that juries find identification evidence highly persuasive, even when informed of its weaknesses (e.g., Lindsay, Wells, & O'Connor, 1989; Semmler, Brewer, & Douglas, 2011).

Since its inception, psychological research into eyewitness memory has informed recommendations to criminal justice systems for the appropriate evaluation and weighting of identification evidence, and aimed to increase knowledge about useful indices of identification accuracy (e.g., Judicial Commission of NSW, 2012; National Research Council, 2014; Smalarz & Wells, 2012; Wells et al., 1998). Applied researchers have aimed not only to document the shortcomings of identification evidence to clarify the bounds of its usefulness and reliability in this way, but also to devise improved tests that can better diagnose suspect guilt or innocence from imperfect eyewitness memory for an offender (e.g., Brewer, Weber, Wootton, & Lindsay, 2012; Lindsay & Wells, 1985; Pozzulo & Lindsay, 1999; Pryke, Lindsay, Dysart, & Dupuis, 2004; Sauer, Brewer, & Weber, 2008; Weber & Perfect, 2012; Zajac & Karageorge, 2009). However, the relatively undeveloped theoretical base of much identification research has limited the scope and strength of conclusions about the comparative effectiveness of existing identification procedures (e.g., Brewer & Wells, 2011; Clark, 2012; Clark & Gronlund, 2015; Gronlund, Mickes, Clark, & Wixted, 2015). It has similarly limited the conceptualisation of novel modifications to identification test formats and procedures that might effectively improve the accuracy of identification evidence (cf. Brewer et al., 2012; Weber & Perfect, 2012). In this project, I extended established theoretical understandings from basic recognition memory and decision-making research that are well-supported by empirical evidence, to propose a novel framework for the analysis of eyewitnesses' decision-making about lineups. Guided by the framework, I developed a lineup procedure with a novel decision structure and metacognitive instructions. These modifications were designed to enhance the quality of memorial information retrieved to inform eyewitnesses' identification decisions, and increase the effectiveness of their metacognitive monitoring and control to improve the accuracy of their decisions.

Theory and Application in this Project: A Brief Overview

A novel analytical framework. In the best-match-first dual-process analytical framework proposed in this project, I extended continuous dual-process theories of recognition memory (e.g., Wixted, 2007; Wixted & Mickes, 2010), combined with a best-match-first account of simultaneous lineup decision-making (see also Clark, 2003, 2008; Wells, 1984, 1993), to explain patterns of identification responses for simultaneous lineups, and identify ways to improve it. The simultaneous lineup has a long history of use by police. In the typical procedure, the witness is asked to evaluate several lineup members presented at once and identify the culprit from the witnessed crime, if present, or reject the lineup as not including the culprit. In brief, dual-process models of memory propose that two memory processes contribute to recognition: decontextualized familiarity, and more specific recollection (Malmberg, 2008; Yonelinas, 2002). Best-match-first decision-making describes eyewitnesses' predisposition to choose from a lineup and thus focus on finding the best-match to memory for the offender when shown a simultaneous lineup, before deciding whether the best-match is, indeed, the offender. In turn, this best-match-first focus, facilitated by the structure of the simultaneous lineup task, encourages eyewitnesses to base their identification decisions predominantly on familiarity, and to underweight the importance of recollection. As a result, eyewitnesses are likely to falsely identify innocent suspects who seem very familiar due to their close resemblance to the offender, even without the confirmation that would be provided by recollective detail. Thus, my first major research question focused on whether the best-match-first dual-process analytical framework could explain observed patterns of response latency and identification responses for standard and modified simultaneous lineup procedures.

In two initial experiments (Experiments 1 & 2, Chapter 3), I investigated patterns of response latency as an index of the temporal order of the component best-match and offender-presence recognition judgments that contribute to simultaneous lineup decision-making. Findings supported framework predictions of spontaneous best-match-first decision-making for a standard simultaneous lineup task. In these and further experiments, I evaluated differences between standard and modified simultaneous procedures in patterns of identification responses to target-present and target-absent lineups to estimate the relative contributions of familiarity and recollection to these decisions. The framework predicted that differences in these patterns between lineup types would correspond to differences in the extent that eyewitnesses' relied on error-prone familiarity alone when following best-match-first decision processes, or made greater use of more accurate recollection. Overall, results were consistent with eyewitnesses adopting familiarity-based decision-making in standard and modified procedures that did not disrupt the best-match-first order of decision-making. In contrast, participants following a modified procedure that effectively disrupted this strategy showed greater discriminability, driven by lower rates of false identification, in a pattern typical of greater use of available recollection.

Novel lineup modifications. Building on these early findings, I developed and tested two key modifications to the simultaneous lineup procedure (Chapters 4, 6, & 7). These modifications were designed to target error-prone best-match-first decision-making strategies, and the neglect of recollection conceptualised in the framework, to increase the quality of the evidence-base of eyewitnesses' decisions. Thus, my second major research question was whether these modifications would effectively improve the reliability of identification decisions based on often fragile eyewitness memory, even if memory quality were poor. Specifically, I investigated the independent and combined effects of a novel, presence-first decision structure and metacognitive instructions. The presence-first decision structure was a two-step decision process, in place of the single identification decision required with a standard procedure. In the presence-first procedure, eyewitnesses first indicate only whether or not they think the offender is present in the lineup shown to them. Second, if the offender was indicated to be present, eyewitnesses point out the offender. This deconstructed approach to the lineup questions asked of eyewitnesses was designed to disrupt error-prone best-match-first decision-making. In the procedure, the best-match choice was explicitly separated from the offender-presence decision, and the natural decision order evident in standard simultaneous lineup decision-making was reversed. A limiting factor, though, on the efficacy of this change to decision structure, was eyewitnesses' robust predisposition to anticipate that the lineup will include a guilty suspect.

Metacognitive instructions were designed to counter eyewitnesses' pre-existing ideas about the purpose of the lineup task, and how to best approach its performance. I aimed to enhance mock eyewitnesses' awareness of the memory requirements of lineup tasks, and the diagnostic qualities of accurate memories. In this way, it was anticipated that eyewitnesses' expectations of the task would be more closely aligned to its demands, and thus improve performance.

Experimental approach. To pursue both major research questions, I used two experimental paradigms. First, in the majority of experiments (Chapters 3-6), I used a face recognition mini-lineup paradigm in which participants studied a series of photos of faces, and made identification decisions for each target. This approach allowed the collection of a large amount of data to better evaluate regularities in participants' responses vis-a-vis the predictions of the novel framework across a variety of stimuli. Second, to investigate the immediate applied implications of early findings, I also tested predictions for the novel lineup procedure in an eyewitness identification paradigm in which participants viewed several videos of mock crime scenarios, and made lineup decisions for an offender from each

(Chapter 7). The identification paradigm thus more closely approximated the conditions under which real eyewitnesses are asked to make identification decisions for police lineups. This was particularly important to test the effect of eyewitness preconceptions about lineup tasks when these were likely to be highly salient. I further discuss the benefits of each approach below.

The Thesis

In this chapter, I outline the form of the thesis, and the scope of work conducted in the two interwoven strands of the project: 1) testing the best-match-first dual-process framework for the analysis of simultaneous lineup decision-making; and 2) testing the efficacy of novel lineup modifications designed to improve recollection, and thus discriminability, in simultaneous procedures. I consider relevant areas of research into lineup procedures, recognition memory, eyewitness metacognition, and decision-making, and indicate where this research has been critically reviewed in the papers that comprise the main body of the thesis (Chapters 3-7). Finally, I provide a brief overview of the experimental work conducted in the course of the project.

Structure and scope. This thesis is presented as a collection of papers that have been submitted for publication to peer-reviewed journals for research in the psychological sciences (Chapters 3-7). As noted, the papers follow the course of conceptual development of both the original best-match-first dual-process analytical framework, and the novel presence-recollection lineup whose design was informed by the framework. I conducted all experimental work reported in these papers as part of this project.[1] Co-authors named on the papers submitted for publication were my supervisor, Nathan Weber[2], and my adjunct

---

[1] Except for Experiment 1 reported in Chapter 5 that I submitted toward examination for the award of BA (Hons.) (Psychology), Flinders University, 2011. This experiment does not form part of the material to be examined for award of the PhD.
[2] Flinders University, South Australia, Australia

supervisor, Ruth Horry.[3] Their contribution was limited to providing supervisory advice on the design of reported experiments (Nathan Weber), the analysis and interpretation of results (Nathan Weber, Ruth Horry), and manuscript editing (Nathan Weber, Ruth Horry). As first author, I designed and conducted all experiments, analysed and interpreted the results, and wrote the manuscripts. In addition to chapters reporting experimental work, Chapter 2 provides an overview of the analytical approach I took in all papers. Finally, Chapter 8 provides a general discussion of the results of this research in terms of the guiding research questions. I consider overall limitations, as well as theoretical and applied implications for future eyewitness identification research and practice, and the recognition memory literature more generally.

Eyewitness Identification Lineups

The two broad motivations guiding this project were: 1) the development of a model of memory and decision processes in lineup procedures that could explain the patterns of identification errors evident in a large body of empirical evidence; and 2) the development of theoretically-grounded interventions or modifications to existing lineup procedures that would effectively improve eyewitnesses use of memory to make more accurate identification decisions. These research aims emerged in response to the relative lack of progress in identification research in both providing conclusive evaluations of comparative identification performance for different lineup procedures commonly used by police, and in developing procedures to improve the diagnosticity of identification decisions. Progress has flagged due to the lack of development in theorising eyewitness identification performance (Brewer & Wells, 2011; Clark & Gronlund, 2015; Gronlund, Mickes, Clark, & Wixted, 2015; Lane & Meissner, 2008), despite a proliferation of empirical studies focusing on procedural comparisons and the effects on accuracy of key witness, situation and procedural variables.

[3] Swansea University, Swansea, UK

Here, I briefly review key aspects of previous research into lineup procedures and the theory of lineup performance, before outlining how, in this project, I proposed to extend theory of the memory and decision processes implicated in identification performance, and test novel theory-based methods for reducing eyewitnesses' use of error-prone decision-making strategies. Introductions to the papers presented in Chapters 3 and 4, and to a lesser degree those of Chapters 5-7, review the literature and empirical evidence most relevant for my main research questions. To avoid unnecessary repetition, I will indicate where I have reviewed relevant theory and empirical evidence in each chapter, while briefly outlining the main research questions and experimental work conducted for each paper.

Simultaneous vs sequential. Intense debate surrounds the question of which existing eyewitness identification procedure produces superior accuracy, focussing mainly on comparisons of sequential and simultaneous lineup procedures (e.g., Goodsell, Gronlund, & Carlson, 2010; Gronlund et al., 2015; Wells, Steblay, & Dysart, 2015; Wixted & Mickes, 2014). Both procedures are commonly used by police (National Research Council, 2014; National Viper Bureau, 2009; Wogalter et al., 2004), with use of the sequential procedure growing rapidly over the last thirty years in response to recommendations by identification researchers based on robust evidence of a reduction in false identification rates (e.g., Wells et al., 1998). In contrast to the traditional simultaneous procedure described above, lineup members in the sequential procedure are presented to the witness one at a time. For each lineup member, the witness is asked to indicate whether they are the culprit or not. In practice, procedures vary but, in the main, witnesses will make this yes/no decision for each member of the lineup. Typically, even after they have given a positive response to identify a lineup member as the culprit, they proceed to view the remaining members of the lineup and provide an identification response for each. In cases when more than one lineup member is then identified as the culprit, the witness is usually asked to indicate which positive identification is correct (e.g., Wells, Steblay, & Dysart, 2015).

Wells' (1984, 1993) influential relative judgment theory set the research agenda for comparative investigations of lineup procedures. While there have been some attempts to formalise or refine Wells' original idea (e.g., Charman & Wells, 2014; Clark, 2003, 2008), theories of identification decisions have not developed much further, even as empirical comparisons of simultaneous and sequential procedures abound. Wells proposed that witnesses using a simultaneous procedure tended to adopt relative decision-making processes. That is, eyewitnesses compared lineup members to find the relatively best-matching lineup member to memory for the culprit, and then tended to identify the best-match as the culprit (e.g., Luus & Wells, 1991). In contrast, eyewitnesses using a sequential procedure were unable to use these comparative decision processes to the same extent, and were instead encouraged to make an absolute evaluation of the degree of match between the sole lineup member in view, and memory for the culprit (Lindsay & Wells, 1985). Relative judgment theory does not outline these processes with reference to existing conceptualisations of memory processes in basic recognition memory theory, for which substantial evidence was available. As a result, the theory is over-general, lacking the specificity needed to power rigorous testing for theoretical development. However, the theory predicts that, in general, eyewitnesses should make fewer false identification decisions in response to sequential lineups than simultaneous lineups. Specifically, eyewitnesses would be prevented from taking the dangerous approach of simply picking the evident best-match to the culprit. Thus, they would compare each lineup member more closely to memory for the culprit, and reject lineup members who did not represent a good-enough match for a positive identification. In addition, the relative judgment idea predicts that rates of accurate identification should not differ between the procedures. That is, when the offender was in the lineup, they would tend to be identified accurately from a simultaneous lineup. With sequential presentation, even when it was not immediately clear to the eyewitness who was the best-matching member of the lineup, the offender would be associated with a high degree of accurate match to eyewitness memory. As a result, they would be most likely to be identified from a fair lineup (Lindsay & Wells, 1980), even without the operation of relative decision processes.

Early research concluded that the sequential lineup procedure did encourage superior use of available memory by eyewitnesses, by discouraging comparisons between lineup members and encouraging absolute evaluations of the degree of match between a lineup member and memory for the culprit (Lindsay & Wells, 1985; Wells, 1984, 1993). Further evidence was taken to demonstrate a drop in accurate identifications, but to a lesser extent than the drop in false identifications (e.g., Steblay, Dysart, Fulero, & Lindsay, 2001; Steblay, Dysart, & Wells, 2011). However, later studies and further meta-analyses (Clark, 2012; Clark & Davey, 2005; Palmer & Brewer, 2012) demonstrated that rather than improving eyewitnesses' ability to discriminate innocent from guilty suspects along with some shift in response bias, the sequential procedure affected only response bias. That is, the sequential procedure encouraged eyewitnesses to adopt a more conservative response bias, and make fewer choices from lineups overall, not disproportionately from target-absent lineups. Thus, while false identification rates were reduced with sequential presentation, so too were accurate identifications of guilty suspects.

Given eyewitnesses' notorious tendency to choose from a lineup, even when memory quality is poor (Clark, 2005, 2012; Malpass & Devine, 1981), procedures that encourage a more conservative approach to lineup decision-making are valuable in themselves. However, patterns of identification responses are not consistent across studies with different stimuli and under different witnessing and test conditions (e.g., Carlson & Gronlund, 2011; Carlson, Gronlund, & Clark, 2008; Wells et al., 2015; Goodsell et al., 2010; Wixted & Mickes, 2014). Thus, available empirical evidence does not provide a definitive answer about which procedure is superior. Undeniably, observed rates of identification errors are substantial for both procedures. Therefore, a continuing imperative for identification research that has been overshadowed by the more immediate question of simultaneous-sequential comparison, is the investigation of procedures with the potential to provide a more diagnostic measure of suspect guilt or innocence than these existing procedures. The desired procedure is one that would reduce false identification decisions, while leaving accurate identifications relatively unaffected or, ideally, increased. Maintaining target-present accuracy, while decreasing misidentifications from target-absent lineups, has three main benefits: greater security for the community in providing accurate evidence of guilt for guilty suspects; lower risk of misidentification for innocent suspects; and greater reliability of identification decisions overall. That is, a positive identification from the lineup would be associated with a greater probability of guilt, while a lineup rejection would be more strongly indicative of innocence. The efficacy of such research relies on further theoretical development. But effective evaluation of novel theoretical development, or the effects of procedural reforms and their practical application, depends at least in part on the measures chosen for analysis.

Measures of identification performance. There is a lack of agreement about the most appropriate statistical measures to evaluate lineup comparisons. Eyewitness identification research traditionally analysed independent measures of false identification rates and accurate positive identification rates, frequently supplemented by evaluations of the diagnosticity ratio (i.e., the ratio of correct to false identification rates, e.g., Lindsay & Wells, 1985). The procedure that yielded the largest diagnosticity ratio value was taken to be superior (e.g., Steblay et al., 2011). Increasingly, however, researchers have extended signal-detection theory to the analysis of lineup performance, challenging the usefulness of more traditional analytical approaches (e.g., Ebbesen & Flowe, 2002; Palmer & Brewer, 2012; Wixted & Mickes, 2012). Criticism of traditional measures has centred on the confounding of response bias (the amount of evidence a witness requires before being willing to make a positive identification from the lineup) and discriminability (the ability to distinguish between an innocent and a guilty suspect, regardless of response bias) in the diagnosticity ratio (e.g., Clark, Erickson, & Breneman, 2011; Palmer & Brewer, 2012; Wixted & Mickes, 2012), and on the lack of informativeness of considering false and accurate identification rates in isolation. Specifically, when witnesses make more accurate positive identification decisions, and fewer false identification decisions, the value of the diagnosticity ratio will be larger. However, when witnesses are simply less willing to choose from a lineup overall, the diagnosticity ratio will also increase, even with no change in discriminability (see Clark, 2012). Instead, research has begun to focus more on the use of signal detection measures of discriminability and response bias to evaluate and compare lineup performance (e.g., Mickes, Moreland, Clark, & Wixted, 2014; Wixted & Mickes, 2012, 2015).

Lineup procedures provide a particular challenge for signal detection measures that are usually applied to a set of participant responses to multiple stimuli (Green & Swets, 1966). In contrast, witnesses in real lineup procedures and in laboratory experiments that adopt a typical identification paradigm, each provide a single response to a single lineup. Thus signal detection measures of *d'* (discriminability) and *c* (response bias) are calculated from sample aggregates (e.g., Clark, 2012; Gourevitch & Galanter, 1967, Mickes, Moreland, Clark, & Wixted, 2014) or ROC curves are constructed from aggregated decision confidence data (e.g., Wixted & Mickes, 2012). In addition, the use of signal detection measures for comparisons between lineup procedures has attracted criticism for failing to account for the unique characteristic of lineup tasks: the presence of filler lineup members accompanying the target or foil stimulus at test, (e.g., Lampinen, 2016; Levi, 2016; Wells, Smith, & Smalarz, 2015).

Most importantly, these criticisms point out that eyewitnesses do not just need to differentiate a guilty from an innocent suspect, they must also differentiate a guilty suspect from among multiple foils at test. Thus, possible eyewitness responses include not only correct and incorrect identifications, misses (i.e., failing to identify a guilty suspect), and correct rejections (i.e., rejecting the lineup when it does not include the culprit), but also the possible error of filler identification. While these responses are known errors, patterns of filler identification will be informative about the underlying memorial basis of identification decisions in different procedures, or decision processes (e.g., the idea of filler siphoning, see Wells, Smalarz, & Smith, 2015). Further, evaluations of the overall accuracy of identification decisions provided for a lineup procedure will differ, depending on how these filler identifications are treated in analyses (Wells, Smalarz, & Smith, 2015; see also Palmer & Brewer, 2012, for use of compound signal detection models with lineup tasks). Excluding filler identifications completely from signal detection analyses will yield a quite different result than including them as lineup rejections.

While some researchers have called for a different approach to analyses to the exclusion of signal detection measures (Lampinen, 2016; Wells, Smith, & Smalarz, 2015) the most appropriate indices of lineup performance will depend on the research question in focus. In this project, I compared identification performance between simultaneous procedures to evaluate the relative contribution of different types of memorial information to eyewitnesses' decisions. Variation in the memorial basis of identification decisions was expected to be reflected in eyewitnesses' ability to distinguish between a target-absent lineup and a target-present lineup. In addition, better use of available memory was expected to be reflected not only in a greater ability to differentiate between guilty and innocent suspects, but also in a reduced proportion of filler identification from either target-absent or target-present lineups. Therefore, signal detection measures of discriminability and response bias were the most useful and appropriate indices of identification performance to evaluate these predictions. To account for patterns of filler identification, I conducted these analyses both with and without the inclusion of filler identifications in the dataset. My analytical approach is outlined further in Chapter 2 (as well as more briefly in each of the papers comprising Chapters 3-7).

Experimental paradigms. By far the majority of eyewitness identification research uses an identification paradigm. As noted earlier, I used a face-recognition mini-lineup paradigm (Weber & Brewer, 2004, 2006), as well as a more realistic identification paradigm, to evaluate predictions of the best-match-first dual-process framework for simultaneous lineup decision-making, and investigate the efficacy of key structural and procedural modifications to the standard simultaneous lineup task. There were several reasons for using the face-recognition mini-lineup paradigm, rather than using an identification paradigm throughout. First, a major limitation of the identification paradigm is that it typically affords the collection of only one data point per participant, in response to a single test stimulus set. As a result, a sample of several hundred is needed to gain adequate power to make conclusions about the effects of key factors on identification responses. In contrast, the mini-lineup paradigm collects multiple responses (i.e., in the experiments reported here, 40 or 80 data points) from each participant, and thus needs a much smaller sample (e.g., I aimed to collect data from 30 participants per lineup condition, in most experiments reported here, see Gelman & Hill, 2007).

Second, the mini-lineup paradigm allowed the evaluation of regularities in responding across multiple stimuli (i.e., 40 or 80 targets, corresponding to 40 or 80 test mini-lineups). Randomised allocation of stimuli to blocks, and counterbalanced or randomised allocation of stimuli to within-subjects experimental conditions, allowed me to account for variability in responses from a sample of participants, and variability in responses to multiple stimuli, before evaluating experimental effects. Thus, more confident generalisation of results across stimuli is a valuable benefit of this approach (Westfall, Judd, & Kenny, 2015). By comparison, identification paradigms typically require participants to view only one offender and respond to one lineup. The appearance of the offender, viewing conditions, and patterns of similarity among lineup members will be idiosyncratic. Thus, generalisation across other potential offenders, viewing conditions, and lineup composition, is problematic and limits the conclusions that can be made about broader application of relevant theory (e.g., Wells & Windschitl, 1999). For example, the level and type of similarity between the offender in a target-present lineup, and a target-replacement in a target-absent lineup, will vary widely with offender and lineup characteristics. Hence, patterns of discriminability and response bias will also vary, potentially with large effect, between tests for different offenders. For confident generalisation of any observed effect, repeated replication across different stimuli and viewing conditions is required, at great cost of participant hours, and other resources. These problems have hindered definitive conclusions about many research questions, including simultaneous-sequential comparisons.

Third, the judgments required in the mini-lineup paradigm are more similar to those required in the basic recognition memory paradigms from which I have extended relevant theory. Thus, predictions for regularities of recognition responses (e.g., in patterns of false and accurate identification rates across stimuli for each participant) can be more readily compared to those from recognition paradigms that differ only in the method of presentation of test stimuli. By extrapolation, observed differences in patterns of responding between the mini-lineup paradigm and the full identification paradigm, can serve to highlight potential implications for the application of understandings derived from simple, controlled laboratory tasks, to more complex, noisy, real world recognition demands.

This last point serves to underscore an important disadvantage of the mini-lineup paradigm. That is, it is several steps further from ecological validity than the identification paradigm. However, in this project, I aimed to establish more fundamental patterns in the use of familiarity and recollection, with extrinsic influences on performance more tightly controlled than would be possible in the identification paradigm. Once established in the mini-lineup paradigm, these effects can then be investigated in more veridical paradigms, when task demands are more complex, and played out under varying, and noisier, conditions. Without this initial, exploratory stage of investigation under more tightly controlled experimental conditions, the development of theoretical approaches to eyewitness identification decisions in diverse procedures will continue to be slow and piecemeal (see also, Brewer & Wells, 2011; Clark & Gronlund, 2015; Gronlund, Mickes, Clark, & Wixted, 2015; Lane & Meissner, 2008; Meissner, Tredoux, Parker, & Maclin, 2005).

The Best-Match-First Dual-Process Framework

The lack of more developed theory of the memory and decision processes used by eyewitnesses has constrained the development of lineup innovations to improve identification performance, just as it has limited understanding of differences in identification performance between simultaneous and sequential procedures (cf. Brewer et al., 2012; Weber & Perfect, 2012). While there have been some attempts to apply dual-process theories of recognition memory to lineup tasks (e.g., Gronlund, 2005; Meissner et al., 2005; Mickes, 2015; Palmer, Brewer, Weber, & McKinnon, 2010), and the application of other decision models has been proposed (e.g., Dunning & Stern, 1994; Wixted & Mickes, 2014), much work has continued to refer to the relative judgment idea, with little further elaboration. In this project, I combined a formalised decision model based on the general insights of Wells' (1984, 1993) relative judgment theory of lineup decision-making strategies (see also, Clark, 2003, 2008), with a continuous dual-process account of recognition memory processes (e.g., Wixted & Mickes, 2010). I proposed an original analytical framework to account for patterns of errors in simultaneous lineup decision-making. I tested this account by comparing identification performance in a standard simultaneous lineup procedure, with that observed for modified simultaneous procedures. In Chapter 3, I provide an outline of the proposed best-match-first dual-process framework in detail. Specifically, I review theory and empirical evidence relating to eyewitness identification decision-making for simultaneous lineup procedures, including patterns of response latency. I further review dual-process accounts of the use of familiarity and recollection in recognition decisions, and the effect of task type and decision structure on the use of these memory processes. In two mini-lineup experiments, I used decision response latency to examine the order of component recognition judgments in standard simultaneous lineup decisions (best-match; offender-presence) and two modified procedures in which these component judgments were separated (modified best-match-first procedure); and reversed (modified presence-first procedure). In addition, I compared patterns of identification responses to evaluate whether, as predicted, discriminability would be improved with disruption of best-match-first decision processes in the modified presence-first procedure.

Decision Structure and Metacognition

It is unrealistic to expect that an eyewitness' response to a suspect could ever be based solely on the properties of the suspect and an encoded memory trace of the offender. All cognition is affected by contextual and individual factors, and memory processing is never passive. Top-down processing is an intrinsic feature of remembering (e.g., Kahnemann, 2011; Odegard & Lampinen, 2006) and, for eyewitnesses, retrieval and evaluation of memory for an offender during an identification test is critically affected by preconceptions and expectations of police identification procedures (e.g., Clark 2005; Malpass & Devine, 1981). Personal motivations and experience, the test context, and unconscious memory and decision processes cued by the test itself, also affect memory performance (Palmer, Brewer, & Weber, 2010; Wells & Olson, 2003). At the most basic level, people have a finite capacity to process test stimuli in complex recognition tasks, and attention is unconsciously but selectively directed, for efficient processing. While efficient, this narrowing of attention does not always best serve the demands of a specific task, and accuracy can suffer.

The structure of a recognition task plays an important role in determining the attention, decision-making strategies, and underlying memorial processes that people devote to its completion (e.g., Malmberg & Xu, 2007; Sauer et al., 2008). Task structure thus contributes to accuracy, along with the individual's metacognitive approach to performance. Eyewitness identification research has demonstrated that eyewitnesses tend to be biased toward making a positive identification from a lineup (Clark, 2005; Malpass & Devine, 1981). Using the best-match-first dual-process framework, I proposed that this approach entails an over-reliance on familiarity-based memory, and a neglect of potentially disconfirming recollective detail. Therefore, based on framework predictions, I tested two methodological interventions to disrupt error-prone decision-making strategies, and encourage better use of available familiarity and recollection. Specifically, in a series of four mini-lineup experiments reported in Chapter 4, I further tested the efficacy of: 1) a presence-first decision-structure for simultaneous lineups; independently and in combination with, 2) metacognitive instructions designed to more closely align eyewitnesses' metacognitive approach to the simultaneous lineup task with task requirements for the use of recollection to maximise accuracy.

In Chapter 4, I review relevant theory and evidence related to eyewitnesses' metacognitive predispositions and expectations of lineup tasks. I also further review theory and empirical evidence relevant for variations in the use of familiarity and recollection in different task structures (e.g., standard vs associative recognition tasks) and with the provision of instructions that aimed to improve metacognitive monitoring and control in recognition tasks requiring attention to recollective detail for accurate performance. In addition, I critically review theory and empirical evidence related to both the use of metacognitive instructions in a variety of recognition tasks to improve the use of recollection, and the mechanisms underlying their effectiveness.

Familiarity and Recollection

For one experiment reported in Chapters 5, and followed up in a second experiment in Chapter 6, I further reviewed theory and evidence relevant to the use and neglect of recollection in recognition tasks. Here, my review focused on evidence relating to the conditions under which recall-to-reject (Rotello & Heit, 2000; see also Brainerd, Reyna, Wright & Mojardin, 2003; Reyna & Brainerd, 1995) is likely to be used or neglected. In these experiments, the contributions of familiarity and recollection to recognition were dissociated with orthogonal experimental manipulations that differentially affected each process. I aimed to evaluate eyewitnesses' spontaneous use of recollection in standard simultaneous lineup tasks, and further evaluate predictions for comparatively improved use of recollection in the presence-recollection lineup (combining a presence-first decision structure with metacognitive instructions).

Ecological Validity

Chapter 7 provides a further brief review of the effects of metacognitive expectations on recognition task performance, depending on task structure and the information provided, with specific reference to identification tasks in more controlled (mini-lineup paradigm) and realistic (identification paradigm) contexts. In one experiment, the novel presence-recollection procedure was tested in a more ecologically valid identification paradigm, for comparison with standard simultaneous and standard sequential procedures.

Summary

In nine experiments, I tested and applied a novel best-match-first dual-process framework for the analysis of simultaneous lineup tasks, and tested lineup modifications informed by the framework that were designed to improve the memorial basis of eyewitnesses' identification decisions. I used a face-recognition mini-lineup paradigm to evaluate regularities in response latency for recognition judgments made about simultaneous lineups. These analyses provided evidence about the time course of eyewitness identification decision-making in standard and modified procedures. I also compared patterns of identification responses between standard simultaneous lineups and modified procedures that disrupted eyewitnesses' usual approach to lineup decision-making. Results provided an estimate of differences in the relative contributions of familiarity and recollection to participants' responses. Comparative analyses of lineup types were conducted in both the mini-lineup paradigm and a more realistic eyewitness identification paradigm, with response bias and discriminability as key dependent measures. In addition, orthogonal manipulations of objectively-available familiarity and recollection were used to dissociate and explore the separate contributions of these processes to participants' simultaneous lineup decisions.

Overall, I aimed to investigate the effects of lineup decision structure, and metacognitive predispositions and knowledge, on eyewitnesses' use of familiarity and recollection in simultaneous lineup tasks. This project was guided by two overarching research questions. First, do eyewitnesses tend to adopt a best-match-first approach to standard simultaneous lineups, and consequently neglect the importance of recollection for avoiding false identification, as proposed by the best-match-first dual-process framework? Second, can the framework inform structural and procedural modifications to the standard lineup task that will disrupt this error-prone decision-making and encourage greater discriminability by enhancing eyewitness use of recollection?

References

Australian Human Rights Commission (2011). Inquiry into the Criminal Cases Review Commission Bill 2010: Australian Human Rights Commission submission to the Legislative Review Committee of South Australia. Sydney, NSW: Australian Human Rights Commission.

Brainerd, C. J., Reyna, V. F., Wright, R., & Mojardin, A. H. (2003). Recollection rejection: False-memory editing in children and adults. *Psychological Review*, *110*, 762-784. doi: 10.1037/0033-295x.110.4.762

Brewer, N., Weber, N., Wootton, D., & Lindsay, D. S. (2012). Identifying the bad guy in a lineup using confidence judgments under deadline pressure. *Psychological Science*, *23*, 1208-1214. doi: 10.1177/0956797612441217

Brewer, N., & Wells, G. L. (2011). Eyewitness identification. *Current Directions in Psychological Science*, *20*, 24-27. doi: 10.1177/0963721410389169

Carlson, C. A., & Gronlund, S. D. (2011). Searching for the sequential line-up advantage: A distinctiveness explanation. *Memory*, *19*(8), 916-929. doi: 10.1080/09658211.2011.613846

Carlson, C. A., Gronlund, S. D., & Clark, S. E. (2008). Lineup composition, suspect position, and the sequential lineup advantage. *Journal of Experimental Psychology: Applied*, *14,* 118-128. doi: 10.1037/1076-898x.14.2.118

Charman, S., & Wells, G. L. (2014). Applied Lineup Theory. *Handbook of Eyewitness Psychology 2 Volume Set*, 219-254. doi: 10.4324/9781315805535

Clark, S. E. (2003). A memory and decision model for eyewitness identification. *Applied Cognitive Psychology*, *17*, 629-654. doi: 10.1002/acp.891

Clark, S. E. (2005). A re-examination of the effects of biased lineup instructions in eyewitness identification. *Law and Human Behavior*, *29*, 575-604. doi: 10.1007/s10979-005-7121-1

Clark, S. E. (2008). The importance (necessity) of computational modelling for eyewitness identification research. *Applied Cognitive Psychology*, *22*, 803-813. doi: 10.1002/acp.1484

Clark, S. E. (2012). Costs and benefits of eyewitness identification reform psychological science and public policy. *Perspectives on Psychological Science*, *7*, 238-259. doi: 10.1177/1745691612439584

Clark, S. E., & Davey, S. L. (2005). The target-to-foils shift in simultaneous and sequential lineups. *Law and Human Behavior, 29*, 151-172. doi: 10.1007/s10979-005-2418-7

Clark, S. E., Erickson, M. A., & Breneman, J. (2011). Probative value of absolute and relative judgments in eyewitness identification. *Law and Human Behavior, 35*, 364-380. doi: 10.1007/s10979-010-9245-1

Clark, S. E., & Gronlund, S. D. (2015). The development and evolution of theory in eyewitness identification. In J. G. W., Raajmakers, A. H. Criss, R. L. Goldstone, R, M. Nosofsky, & M. Steyvers (Eds.) *Cognitive Modeling in Perception and Memory: A Festschrift for Richard M. Shiffrin*, Psychology Press: New York. doi: 10.4324/9781315885582

Criminal Cases Review Commission (2016a). Criminal Cases Review Commission. Retrieved March 5, 2016. http://www.ccrc.gov.uk/

Criminal Cases Review Commission (2016b). Criminal Cases Review Commission annual report and accounts 2014/2015. Birmingham, UK: Williams Lee Group on behalf of the Controller of Her Majesty’s Stationary Office.

Cutler, B. L., & Penrod, S. D. (1995). *Mistaken identification: The eyewitness, psychology and the law*. Cambridge University Press. doi: 10.1007/978-1-4615-4891-1

Dioso-Villa, R. (2015). A repository of wrongful convictions in Australia: First steps toward estimating prevalence and causal contributing factors. *Flinders Law Journal, 17,* 163-202.

Dunning, D., & Stern, L. B. (1994). Distinguishing accurate from inaccurate eyewitness identifications via inquiries about decision processes. *Journal of Personality and Social Psychology*, *67*, 818-835. doi: 10.1037//0022-3514.67.5.818

Ebbesen, E. B., & Flowe, H. D. (2002). Simultaneous v. sequential lineups: What do we really know. *Unpublished manuscript*.

Fitzgerald, R. J., Price, H. L., Oriet, C., & Charman, S. D. (2013). The effect of suspect-filler similarity on eyewitness identification decisions: A meta-analysis. *Psychology, Public Policy, and Law*, *19*, 151-164. doi: 10.1037/a0030618

Gelman, A., & Hill, J. (2007). *Data Analysis using Regression and Multilevel/Hierarchical Models.* New York: Cambridge University Press.

Goodsell, C. A., Gronlund, S. D., & Carlson, C. A. (2010). Exploring the sequential lineup advantage using WITNESS. *Law and Human Behavior*, *34*, 445-459. doi: 10.1007/s10979-009-9215-7

Gourevitch, V., & Galanter, E. (1967). A significance test for one parameter isosensitivity functions. *Psychometrika*, *32*, 25-33. doi: 10.1007/bf02289402

Green, D. M., & Swets, J. A. (1966). Signal detection theory and psychophysics. 1966. *New York*, *888*, 889.

Gronlund, S. D. (2005). Sequential lineup advantage: Contributions of distinctiveness and recollection. *Applied Cognitive Psychology,* 19, 23-37. doi: 10.1002/acp.1047

Gronlund, S. D., Mickes, L., Wixted, J. T., & Clark, S. E. (2015). Conducting an eyewitness lineup: How the research got it wrong. In B. H. Ross (Ed.), The psychology of learning and motivation (Volume 63, pp. 1– 43). Waltham, MA: Academic Press. doi: 10.1016/bs.plm.2015.03003

Horry, R., Memon, A., Wright, D. B., & Milne, R. (2011). Predictors of eyewitness identification decisions from video lineups in England: A field study. *Law and Human Behavior*, *36*, 257-265. doi: 1007/s10979-011-9279-z

Innocence Project (2016). Innocence project retrieved March 1, 2016 http://www.innocenceproject.org

Judicial Commission of NSW (2012). *Annual Report 2011-2012*. Sydney, NSW: Judicial Commission of NSW.

Kahnemann, D. (2011). *Thinking, fast and slow*. New York: Macmillan.

Lampinen, J. M. (2016). ROC analysis in eyewitness identification research. *Journal of Applied Research in Memory & Cognition, 5*, 21-33. doi: 10.1016/j.jarmac.2015.08.006.

Lane, S. M., & Meissner, C. A. (2008). A ‘middle road’ approach to bridging the basic–applied divide in eyewitness identification research. *Applied Cognitive Psychology*, *22*, 779-787. doi: 10.1002/acp.1482

Levi, A. (2016). Is ROC analysis a tool that should replace probative analysis in studying lineups? *Journal of Criminal Psychology*, *6*. 42-48. doi: 10.1108/jcp-07-2015-0024

Lindsay, R. C., & Wells, G. L. (1980). What price justice? Exploring the relationship of lineup fairness to identification accuracy. *Law and Human Behavior*, *4*, 303-313. doi: 10.1007/bf01040622

Lindsay, R. C., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology, 70*, 556. doi: 10.1037/0021-9010.70.3.556

Lindsay, R. C., Wells, G. L., & O'Connor, F. J. (1989). Mock-juror belief of accurate and inaccurate eyewitnesses. *Law and Human Behavior*, *13*, 333-339. doi: 10.1007/bf01067033

Luus, C. A., & Wells, G. L. (1991). Eyewitness identification and the selection of distracters for lineups. *Law and Human Behavior, 15*, 43. doi: 10.1007/bf01044829

Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology*, *57*, 335-384. doi: 10.1016/j.cogpsych.2008.02.004

Malmberg, K. J., & Xu, J. (2007). On the flexibility and the fallibility of associative memory. *Memory & Cognition, 35*, 545-556. doi: 10.3758/bf03193293

Malpass, R. S., & Devine, P. G. (1981). Eyewitness identification: Lineup instructions and the absence of the offender. *Journal of Applied Psychology*, *66*, 482. doi: 10.1037/0021-9010.66.4.482

Meissner, C. A., Tredoux, C. G., Parker, J. F., MacLin, O. H. (2005). Eyewitness decisions in simultaneous and sequential lineups: A dual-process signal detection theory analysis. *Memory & Cognition*, 33, 783-792. doi: 10.3758/bf03193074

Mickes, L. (2015). Receiver operating characteristic analysis and confidence-accuracy characteristic analysis in investigations of system variables and estimator variables that affect eyewitness memory. *Journal of Applied Research in Memory and Cognition, 4*, 83-102. doi: 10.1016/j.jarmac.2015.01.003

Mickes, L., Moreland, M. B., Clark, S. E., & Wixted, J. T. (2014). Missing the information needed to perform ROC analysis? Then compute d′, not the diagnosticity ratio. *Journal of Applied Research in Memory and Cognition*, *3*, 58-62. doi: 10.1016/j.jarmac.2014.04.007

National Registry of Exonerations (2016). Current Exonerations. Retrieved February 22, 2016. http://www.law.umich.edu/special/exoneration/Pages/about.aspx

National Research Council. (2014). I*dentifying the Culprit: Assessing Eyewitness Identification.* Washington, DC: The National Academies Press. doi: 10.17226/18891

National VIPER Bureau. (2009). Viper: Video Identification Parade Electronic Recording, retrieved January 16, 2016, http://www.viper.police.uk/

Odegard, T. N., & Lampinen, T. N. (2006). Memory editing: Knowledge, criteria, and alignment. *Memory, 14,* 777-787. doi: 10.1080/09658210600648589

Palmer, M. A., & Brewer, N. (2012). Sequential lineup presentation promotes less-biased criterion setting but does not improve discriminability. *Law and Human Behavior*, *36*, 247-255. doi: 10.1037/h0093923

Palmer, M. A., Brewer, N., McKinnon, A. C., & Weber, N. (2010). Phenomenological reports diagnose accuracy of eyewitness identification decisions. *Acta Psychologica*, *133*, 137-145. doi: 10.1016/j.actpsy.2009.11.002

Palmer, M. A., Brewer, N., & Weber, N. (2010). Postidentification feedback affects subsequent eyewitness identification performance. *Journal of Experimental Psychology: Applied*, *16,* 387-398. doi: 10.1037/a0021034

Pozzulo, J. D., & Lindsay, R. C. L. (1999). Elimination lineups: An improved identification procedure for child eyewitnesses. *Journal of Applied Psychology*, *84*, 167-176. doi: 10.1037/0021-9010.84.2.167

Pryke, S., Lindsay, R. C. L., Dysart, J. E., & Dupuis, P. (2004). Multiple independent identification decisions: a method of calibrating eyewitness identifications. *Journal of Applied Psychology*, *89*, 73-84. doi: 10.1037/0021-9010.89.1.73

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and individual Differences*, *7*(1), 1-75. doi: 10.1016/1041-6080(95)90031-4

Rotello, C. M., & Heit, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory & Cognition*, *28*, 907-922. doi: 10.3758/bf03209339

Sangha, B., & Moles, R. (2012). Post-appeal review rights: Australia, Britain and Canada. *Criminal Law Journal*, 36, 300.

Sauer, J. D., Brewer, N., & Weber, N. (2008). Multiple confidence estimates as indices of eyewitness memory. *Journal of Experimental Psychology: General*, 137, 528-547. doi: 10.1037/a0012712

Semmler, C., Brewer, N., & Douglass, A. B. (2011). Jurors believe eyewitnesses. In B. L. Cutler (Ed.), *Conviction of the innocent: Lessons from psychological research* (pp. 185–209). Washington, DC: American Psychological Association. doi: 10.1037/13085-009

Smalarz, L., & Wells, G. L. (2012). Eyewitness-identification evidence: Scientific advances and the new burden on trial judges. *Court Review: The Journal of the American Judges Association. Paper 385.* 1-21.

Steblay, N., Dysart, J., Fulero, S., & Lindsay, R. C. (2001). Eyewitness accuracy rates in sequential and simultaneous lineup presentations: a meta-analytic comparison. *Law and Human Behavior*, *25*, 459-473. doi: 10.1023/a:1012888715007

Steblay, N. K., Dysart, J. E., & Wells, G. L. (2011). Seventy-two tests of the sequential lineup superiority effect: A meta-analysis and policy discussion. *Psychology, Public Policy, and Law, 17,* 99-139. doi: 10.1037/a0021650

Weber, N., & Brewer, N. (2004). Confidence–accuracy calibration in absolute and relative face recognition judgments. *Journal of Experimental Psychology: Applied, 10,* 156–172. doi: 10.1037/1076-898x.10.3.156

Weber, N., & Brewer, N. (2006). Positive versus negative face recognition decisions: Confidence, accuracy and response latency. *Applied Cognitive Psychology, 20,* 17–31. doi: 10.1002/acp.1166

Weber, N., & Perfect, T. J. (2012). Improving eyewitness identification accuracy by screening out those who say they don't know. *Law and Human Behavior, 36*, 28-36. doi: 10.1037/h0093976

Wells, G. L. (1984). The Psychology of Lineup Identifications1. *Journal of Applied Social Psychology, 14,* 89-103. doi: 10.1111/j.1559-1816.1984.tb02223.x

Wells, G. L. (1993). What do we know about eyewitness identification? *American Psychologist*, *48*, 553-571. doi: 10.1037//0003-066x.48.5.553

Wells, G. L., & Olson, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology*, *54*, 277-295. doi: 10.1146/annurev.psych.54.101601.145028

Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M., & Brimacombe, C. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behavior*, *22*, 603-647. doi: 10.1023/a:1025750605807

Wells, G. L., Smalarz, L., & Smith, A. M. (2015). ROC analysis of lineups does not measure underlying discriminability and has limited value. *Journal of Applied Research in Memory and Cognition*, *4*, 313-317. doi: 10.1016/j.jarmac.2015.08.008

Wells, G. L., Smith, A. M., & Smalarz, L. (2015). ROC analysis of lineups obscures information that is critical for both theoretical understanding and applied purposes. *Journal of Applied Research in Memory and Cognition*, *4*, 324-328. doi: 10.1016/j.jarmac.2015.08.010

Wells, G. L., Steblay, N. K., & Dysart, J. E. (2015). Double-blind photo lineups using actual eyewitnesses: An experimental test of a sequential versus simultaneous lineup procedure. *Law and Human Behavior*, *39*, 1-14. doi: 10.1037/lhb0000096

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin*, *25*, 1115-1125. doi: 10.1177/01461672992512005

Wells, G. L., Yang, Y., & Smalarz, L. (2015). Eyewitness identification: Bayesian information gain, base-rate effect equivalency curves, and reasonable suspicion. *Law and Human Behavior*, *39*, 99-122. doi: 10.1037/lhb0000125

Westfall, J., Judd, C. M., & Kenny, D. A. (2015). Replicating studies in which samples of participants respond to samples of stimuli. *Perspectives on Psychological Science*, *10*, 390-399. doi: 10.1177/1745691614564879

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review, 114*, 152-176. doi: 10.1037/0033-295x.114.1.152

Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review, 117*, 1025-1054. doi: 10.1037/a0020874

Wixted, J. T., & Mickes, L. (2012). The field of eyewitness memory should abandon probative value and embrace receiver operating characteristic analysis. *Perspectives on Psychological Science*, *7*, 275-278. doi: 10.1177/1745691612442906

Wixted, J. T., & Mickes, L. (2014). A signal-detection-based diagnostic-feature-detection model of eyewitness identification. *Psychological Review*, *121*, 262-276. doi: 10.1037/a0035940

Wixted, J. T., & Mickes, L. (2015). Evaluating eyewitness identification procedures: ROC analysis and its misconceptions. *Journal of Applied Research in Memory and Cognition*, *4*, 318-323. doi: 10.1016/j.jarmac.2015.08.009

Wogalter, M. S., Malpass, R. S., & McQuiston, D. E. (2004). A national survey of US police on preparation and conduct of identification lineups. *Psychology, Crime and Law*, *10,* 69-82. doi: 10.1080/10683160410001641873

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*, 441-517. doi: 10.1006/jmla.2002.2864

Zajac, R., & Karageorge, A. (2009). The wildcard: A simple technique for improving children's target-absent lineup performance. *Applied Cognitive Psychology*, *23*, 358-368. doi: 10.1002/acp.1511

CHAPTER 2

Analytical Approach

In this chapter, I outline my approach to data analysis for all experiments reported in this thesis. More traditional approaches to data analysis are problematic when applied to data from experiments that use multiple stimulus sets and memory tests with a number of participants. Therefore, I used mixed-effects models for confirmatory hypothesis testing with data from both face recognition mini-lineup, and eyewitness identification paradigms. Here, I explain the benefits of using mixed-effects modelling in this way, and indicate the simplest approach to model interpretation. As chapters 3-7 comprise complete papers submitted for publication, each includes a truncated explanation of this analytical approach. Therefore, while there is some replication of this material throughout the thesis, it is readily available for ease of reference with each set of reported results.

Overview

In all experiments I report here, a sample of participants were asked to provide responses to memory tests for each stimulus in sets of multiple stimuli within different experimental conditions. The conventional use of ANOVA and logistic regression analyses in applied memory research is not optimal for confirmatory hypothesis testing with this kind of experimental data (e.g., Baayen, Davidson, & Bates, 2008; Barr, Levy, & Scheepers, 2013; Gelman & Hill, 2007; Jaeger, 2008; Judd Westfall, & Kenny, 2012). For these data, ANOVA cannot account concurrently for the several sources of random variation in responses across participants and to diverse stimuli, including variation in the effects of within-subjects manipulations. Failing to model these sources of variance when testing hypotheses for key between- or within-subjects factors is likely to inflate Type I error rates or, conversely, obscure meaningful effects on outcome variables (Barr et al., 2013; Judd, Westfall, & Kenny, 2012; Westfall, Judd & Kenny, 2015). In addition, as all participants respond to the same stimuli (i.e., participant and stimuli effects are crossed), their responses are not statistically independent, as ANOVA models assume (Wright & McDaid, 1996).

Therefore, when appropriate, I used linear mixed-effects modelling for confirmatory hypothesis testing for planned and post hoc comparisons. This approach is recommended by statisticians (e.g., De Carlo, 2009; Gelman & Hill, 2007; Wright, Horry, & Skagerberg, 2009; Wright & London, 2009) and increasingly recognised by applied memory researchers (e.g., Dodson & Dobolyi, 2015; Horry, Halford, Brewer, Milne, & Bull, 2014; Horry, Memon, Wright, & Milne, 2012; Weber & Varga, 2012) as most appropriate for the kinds of data collected in these experimental paradigms. The interpretation of linear mixed-effects models that include all relevant factors (*maximal models*) allows nuanced, but simple, hypothesis testing, and effectively accounts for this variance in providing an estimation of effects (Barr et al., 2013). Conceptually, linear mixed-effects analyses can be interpreted in the same way as regression (ordinary and logistic), and describing the effects of key factors differs little from when interpreting an ANOVA model of the same comparisons.

This chapter is divided into several sections. First, I explain further why ANOVA and logistic regression models are not the most appropriate approach for analysing the data presented in this research. Second, I explain how using mixed-effects models for hypothesis testing addresses the problems encountered with ANOVA and logistic regression models. Third, I set out a rationale for building the random and fixed effects structures of maximal models. The fourth section outlines how the effects of key factors can be tested and interpreted in maximal mixed-effects models. This section includes an extended description of how signal detection measures of response bias and discriminability were evaluated in the papers presented in Chapters 3-7, following De Carlo (2009), and Wright & London (2009). I also briefly describe models used to estimate the contributions of familiarity and recollection to recognition in the experiments reported in Chapters 5 and 6 (cf. more familiar approaches using hit and false alarm rates).

Problems with Traditional Analytical Approaches to Experiments with Multiple Participants and Stimuli

In each of the face recognition and eyewitness identification experiments presented in this thesis, multiple participants studied stimuli sets made up of multiple photos of faces or videos of mock crime scenarios. Participants provided a range of responses to multiple test stimuli, including recognition decisions, ratings of decision confidence, and evaluations of best match to memory. Using multiple stimuli in experimental designs increases the generalisability of findings across stimuli (Wells & Windschitl, 1999; Westfall, Kenny & Judd, 2014). For the same reason, several different sets of stimuli were used in the experiments presented in this thesis, to ensure that the replication of important effects was tested across different stimuli (Westfall et al., 2015). These questions of design and generalisability have not always been adequately addressed in applied eyewitness identification research (e.g., Brewer & Wells, 2011; Gronlund, Mickes, Wixted, & Clark, 2015; Wells & Windschitl, 1999). As a result, much research in the field has a poor record of replication and generalisation. Further, conclusions with important practical ramifications have been made on a thin, often under-powered evidence-base. Adopting a by-participant and by-stimulus analytical approach is not sufficient to address these concerns in itself. Trial-by-trial analyses must be backed up by collecting a sample size larger than that often used for similar research when aggregate measures are analysed (e.g., Asendorpf et al., 2013; Bakker, van Dijk, & Wicherts, 2012; Button et al., 2013; Schimmack, 2012). A larger number of observations per experimental cell will provide enough power to find meaningful effects, especially when multiple factors are tested, as they often are in identification research. These measures are worthwhile because, for these kinds of data, there are several major problems with the use of aggregate statistics (e.g., mean proportion correct, confidence, *d'* and *c*, and *rl*) in tests of the effects of key factors. These problems are outlined below.

First, aggregate statistics may obscure meaningful effects, or yield fallacious effects that are statistically significant, as a result of failing to appropriately account for random variation in responses to individual stimuli both within and between participants (Baayen et al., 2008; Barr et al., 2013; Jaeger, 2008). For example, aggregate measures of signal detection parameters for discriminability and response bias, *d'* and *C*, conflate responses to stimuli that are likely to vary in distinctiveness and memorability (e.g. Bruno, Higham, & Perfect, 2009). A meaningful effect of an experimental factor on discriminability might, therefore, be hidden in aggregates that combine responses to stimuli for which an effect on discriminability was large and meaningful, and stimuli for which discriminability was not as substantially affected. This problem is exacerbated when the size of the effect varies between participants. Similarly, different participants asked to provide decision confidence ratings might have the same mean confidence value, even if they gave responses across quite different ranges of the confidence scale. For example, a mean confidence of 60% might be gained equally from responses made across the whole range of possible confidence values from 0 to 100%, as from responses made within a more restricted range of, say, 50 to 70%. Clearly, while the standard deviation provides an estimate of this variability, the mean model will better represent responses gained from the participant who used the narrower range of the confidence scale. Just as these measures hide important differences, analyses of aggregates may yield a seemingly significant effect by separating data points from the specific participants and stimuli from which they were collected, and re-combining them across these non-independent sources of variation, without appropriately accounting for their variability.

Second, contributing to this problem, ANOVA is unable to account simultaneously for variation in both the effects of experimental factors across stimuli, and between participants within conditions. The importance of this is clear when we consider that the face recognition and eyewitness identification experiments presented here used samples of face photographs. As noted, these stimuli will likely vary in their memorability. But further, the stimuli used do not represent an exhaustive set of all such stimuli. They also differ from other samples of face stimuli used in other research using similar paradigms. Accordingly, there is likely to be significant random variation between stimuli in participants' recognition responses, and in the effects of experimental factors on these responses.

For example, the amount of recollective detail encoded for each face is likely to vary considerably within the sample of stimuli. Participants might encode a stronger memory trace, with more recollectible detail, for more than less memorable faces (e.g., Ingram, Mickes, & Wixted, 2012; Mickes, Wais, & Wixted, 2009). As a result, the recollective evidence for the more memorable faces is likely to be more immediately and automatically retrieved and used in recognition decisions, regardless of other experimental manipulations affecting recollection encoding (e.g., divided attention). Therefore, the use of recollective detail in making recognition responses to memorable faces might vary less with experimental manipulations affecting recollection, than responses to less memorable faces. Take a key manipulation from the research presented in this thesis: pre-test instructions about how to make better use of recollection. A difference in memory performance observed between conditions in which participants are or are not given these instructions may be smaller for more memorable than less memorable faces. An overall pattern of a meaningful effect might thus be lost if responses were aggregated and averaged across these stimuli. Failing to consider this kind of variation in effects between stimuli when traditional ANOVA is used, even when accounting for random variation between and within participants, is likely to inflate Type 1 error rates (Barr et al., 2013). While ANOVA models include consideration of variation in effects across participants (using individual aggregate measures as the outcome variable) they cannot also account for difference in effects between stimuli (Murayama, Sakaki, Yan, & Smith, 2014; Quené, & van den Bergh, 2008). These traditional statistics assume the independence of observations collected from each participant and in response to each stimulus. Clearly, in designs in which multiple participants respond to the same set of multiple stimuli, the collected data are not independent, and more appropriate statistics for their analysis are available.

Finally, aggregate measures for outcome variables for individual participants can only be calculated if participants have provided the responses needed to produce these measures. For example, the signal detection measure of discriminability, *d*', is calculated using aggregate measures of accurate (*hits*) and inaccurate (*false alarms*) positive recognition decisions. Therefore, this measure can only be calculated for participants who have made both accurate and inaccurate positive recognition decisions in response to a set of multiple stimuli, unless adjustments to the original data are made. Excluding participants with missing data points from analyses, or analysing measures produced after post hoc adjustment for missing values, can reduce analytical power or lead to the interpretation of results based on a set of responses less representative of the sample.

Mixed-Effects Models for Hypothesis Testing

With mixed-effects models, each of these problems can be addressed. Linear mixed-effects models are an extension of regression models. Models are built to estimate the change in an outcome variable as a result of changes in one or more predictor variables. For example, the change in identification accuracy (the outcome variable), dependent on target presence (the predictor variable). Unlike ANOVA that relies on aggregate statistics, mixed-effects models are run trial-by-trial on each collected data point. As each data-point from every participant contributes to the creation of the model, models provide a better representation of the full range of participants' responses in the sample data. Importantly, variation between participants and between responses to different stimuli can be explicitly accounted for by inclusion in the models as random effects. In this way, inflation of Type I errors is avoided (Judd et al., 2012; Murayama et al., 2014), without inaccurate assumptions of independence (Wright & McDaid, 1996).

Mixed-effects modelling allows us to consider the effects of multiple predictor variables on an outcome variable in a single model. While ANOVA models also allow consideration of the effects of multiple factors on a dependent variable, they are unable to model all sources of random variability. Therefore, assumed error variance increases more with additional predictors than in mixed-effects models. This increase may hide small, but meaningful effects. Another disadvantage of using ANOVA models is that the effects of outlying responses to specific stimuli or from specific individuals may be exaggerated in calculating aggregate measures. Excluding such outliers would reduce power, and the accuracy of the model's representation of sample data. Mixed-effects models can allow for missing or outlying data points (from individual trials), without requiring their exclusion (Barr et al., 2013).

Mixed-effects models include a mixture of fixed and random effects. Fixed effects are factors for which every level in the population from which they are drawn is measured in the experiment. For example, variables such as gender, and experimental manipulations, would be included in models as fixed effects. Random effects are those factors for which only a sample of the population is used. For example, characteristic random effects in applied memory paradigms include participant and stimulus or item, as both the participants and the stimuli in the experiment are sampled from populations of potential participants and stimuli respectively (Baayen et al., 2008). These random effects may be nested within other factors, or crossed. In typical regression models, a single error term is assumed to account for all variation in the outcome variable not caused by the predictor variable(s). In contrast, mixed-effects models explicitly model the effects of differences in responding between participants, and to different stimuli, as discrete random effects.

When a variable is included in a model as a random effect, the variability (standard deviation) in the outcome variable that is associated with variation in the random effect is estimated. As the model is a regression equation that predicts responses on a curve, variability may be evident in the model intercept, or in the term(s) that produce the overall slope of the regression equation. For example, the formula for a logistic regression model is:

$$y' = b_0 + b_1X_1 + b_2X_2 + \ldots\ldots + b_nX_n$$

where y' is the predicted log odds of the outcome variable (e.g., log odds of making a positive identification decision, rather than a lineup rejection), b_0 is the y intercept, each X is a fixed factor, and each corresponding b value is the change in the slope of the line for that specific fixed factor. The model intercept, b_0, is the baseline value observed for the outcome variable, while the slope represents the predicted change in the observed value of the outcome variable per unit change in the predictor variables. In effect, in estimating the effect on the outcome variable of a fixed factor, the model allows the parameters (intercept and slopes) to vary randomly by the random factors that have been entered. This random variation is not conflated with the model's estimation of the variability due to the fixed factor itself. Random variation can be evident in 1) the tendency to make a given response between participants and to different stimuli (random intercept), 2) the effect of within-subjects fixed factors on participants' responses (random slopes), and 3) responses to factors varied within-stimuli across participants (random slopes). The extent of this random variation, not due directly to the effect of the fixed predictor variables being tested, can then be considered and accounted for within the overall model, alongside the extent of additional variability that is due to fixed factors (Judd et al., 2012).

Random and Fixed Effects Structures in Maximal Models

As noted earlier, maximal mixed-effects models include all relevant factors. The interpretation of the effects of these factors on the outcome variable in a single model allows straightforward confirmatory hypothesis testing that does not focus unnecessarily on goodness of fit. Describing the effects of these key factors on the outcome variable when interpreting a mixed-effects model follows similar logic as interpreting an ANOVA model of the same comparisons. For each experiment in Chapters 3-7, I made planned pairwise comparisons between conditions of main interest (Gelman, Hill, & Yajima, 2012). Here, I describe my approach to building the random and fixed effects structures of these models, in turn, before using examples to illustrate the approach taken to their interpretation throughout this thesis.

Maximal models effectively account for crossed random variability in non-independent observations across participants and stimuli in providing an estimation of the effects of fixed factors of interest (Barr et al., 2013). As models were used for hypothesis testing, rather than evaluating model fit or modelling lineup or recognition decision processes per se, the significance of each addition to the model of a key factor was not relevant. What mattered was the amount of variability that could be ascribed to each factor of main interest, when the contribution of all key factors was accounted for in the model. Therefore, decisions about the inclusion of random and fixed factors depended on their importance for testing predictions, while maintaining overall statistical power to evaluate the effects of experimental factors, while avoiding the inflation of Type I error caused by ignoring stimulus variability (Westfall et al., 2015).

Random effects. The appropriate random effects structure for a model will depend on experimental design and the importance of potential effects. For example, consider an experiment which varies the exposure duration of a single set of face photograph stimuli within-subjects on two levels (short, long) and tests recognition of these faces. Photos would be randomly assigned to a level of exposure duration for each participant. In this design, the random effects of participant and stimulus are crossed (as one random effect is not nested within the other: all participants study the same set of faces, but the exposure duration for a specific photo will vary between participants). A model to predict face recognition accuracy as the outcome variable from the fixed predictor variable of exposure duration can be created with these data. Entering participant as a random effect on the intercept will allow the model to estimate the variability in accuracy due to random variation between *participants* that was not due to the effect of any fixed factors (including the experimental manipulation of exposure duration). Similarly, entering *stimulus* as a random effect on the intercept will allow the model to estimate the variability in accuracy due to random variation in responses to differently memorable face photos in the set. A random slope for variation between *participants* in the effect of the fixed (within-subjects) factor of exposure duration on the slope of the curve predicting accuracy can also be added to the model (exposure|participant). The inclusion of this random slope in the model allows us to explicitly account for random variation between participants in the effect of exposure duration on recognition accuracy, before evaluating the average size of an effect attributable to the experimental manipulation of exposure duration itself. In the same way, a random slope for this effect by *stimulus* (exposure|stimulus) will take into account random variation in the effect of exposure duration between stimuli in the set, when evaluating the effect of the fixed factor itself.

A simple way to conceptualise the contribution of random effects to a mixed-effects model is that, in effect, each model parameter is calculated separately for every level of each random effect. That is, in the case of the example above, for every participant and every stimulus. An intercept, indicating baseline value of the outcome variable before considering the effect of other predictor variables, and a slope coefficient, indicating the change in the outcome variable with a unit change in the predictor, are calculated for each level of each random effect (each person, and each stimulus).

As a general rule, in the experiments reported here, I included random intercept for participant and stimulus. Random slopes were included by participant for key within-subjects fixed factors (e.g., if target-presence was varied within-subjects, a random slope for target presence by participant was added: target-presence|participant). Additional random slopes were considered when each level of a fixed effect (e.g., target presence) would have occurred for each level of a random effect (e.g., when target-presence at test was randomly varied within the set of mini-lineups that was seen by all participants: target-presence|stimulus). In other words, random slopes were included by participant for within-subjects fixed factors (but not for between-subjects fixed factors). For example, when lineup type (e.g., standard simultaneous lineup; modified lineup) was varied within subjects, it was appropriate to include a random slope for lineup type by participant (lineup type|participant). In contrast, when lineup type was varied between-subjects using a single stimulus set for all participants, it was appropriate to include a random slope for lineup type by stimulus (lineup type|stimulus), but not by participant. With between-subjects variation of lineup type, participants responded to a single lineup procedure (e.g., standard simultaneous lineups), so there was no random variation between individuals in the effect of lineup type to be accounted for). When models included a large number of within-subjects or within-stimulus effects, only the most important random effects were included to avoid over-parameterisation. Including too many random factors in models constructed for confirmatory hypothesis testing reduces the model's power to evaluate fixed effects, with negligible improvement in the estimation of the fixed effect (Westfall et al., 2014).

Fixed effects. The fixed effects structure of a model designed for confirmatory hypothesis testing of comparisons between experimental conditions is determined by the independent variables included in the design. Broadly speaking, all independent variables, and their interactions, will be included in the maximal model. The effect of any specific factor can then be evaluated in the context of all effects of all other factors. For example, a main focus of the analyses reported in this thesis, was on pairwise comparisons of identification performance between lineup types (e.g., standard simultaneous lineup vs modified procedure). Specifically, I focused on differences between lineup types in the signal detection parameters of response bias (i.e., participants' tendency to make a positive identification decision, or choose, from a lineup) and discriminability (i.e., the ability to discriminate a previously-seen, guilty suspect from an innocent suspect, regardless of response bias). Analyses followed the approach outlined by De Carlo (2009) and Wright and London (2009; see also Wright et al., 2009). These logistic mixed-effects models predicted the categorical outcome variable of decision type (lineup rejection; positive identification) from a minimum of three fixed factors: lineup type; target-presence; and their interaction. The effect of the fixed factor of lineup type indexed a difference in response bias between lineup procedures. Similarly, the effect of the fixed factor of target presence indexed discriminability, as the difference in the likelihood of making a positive identification decision between target-absent lineups and target-present lineups. To evaluate differences in discriminability between lineup types, the fixed factor of the lineup type × target-presence interaction term was evaluated. This fixed factor structure thus allowed me to evaluate the effect of lineup type on signal detection parameters within a single maximal model that also included appropriate random effects of participant and stimulus (depending on design, for each model).

Model specifications. Mixed-effects models in this thesis were created using the lme4 package (Bates, Maechler, Bolker, & Walker, 2014) in R, an open-source language environment for statistical computing (R Core Team, 2014). The main class of models used for hypothesis testing were those that predicted identification decision type as a dichotomous categorical outcome variable (lineup rejection, positive identification). Therefore, a logit link function was appropriate, and these models represent an extension of logistic regression. Logistic regression differs from Gaussian regression models in a key feature: in Gaussian regression, an outcome variable is predicted from each value of the predictor variable; whereas, in logistic regression, the *likelihood* (log odds) of an outcome is predicted, given a specific value of the predictor variable (Jaeger, 2008). Gaussian models were used to evaluate the effects of key factors of interest on identification decision response latency in Experiments 1 and 2.

Each model included appropriate random intercepts and slopes, as outlined earlier, and fixed predictor variables of interest and their interactions. Running a model generates a regression coefficient for each predictor (b). When using the logit link function, b represents the amount of change in the log odds of an outcome variable with a single unit change in the predictor variable *when all other predictor variables have a value of zero* (i.e., the designated reference conditions). In comparison, for a continuous outcome variable (e.g., response latency), b represents the amount of change in the outcome variable per unit change in the predictor (also when all other predictor variables have a value of zero). In the models created to test predictions in this thesis, many of the predictor variables were also dichotomous. For example, target presence (target-absent, target-present), and lineup type (standard, modified). For these predictors, the coefficient represents the change in the slope of the regression line between the two values of the predictor (an assigned reference condition = 0, and a comparison condition = 1). Hence, the value of the coefficient for the effect of target-presence on decision type represents the amount of change in the outcome between the target-absent condition (0) and the target-present condition (1), when the value of all other predictors is zero. When there were a number of dichotomous predictors included in a model predicting a dichotomous outcome variable, the coefficient for each added predictor was interpreted as representing the change in the likelihood of the specific outcome (e.g., a positive identification decision) between the two levels of the predictor variable – *at the levels of other predictor variables that had arbitrarily been given the zero value*. For example, for the dichotomous predictor variable of lineup type, the coefficient represented the change in likelihood of the outcome variable between the standard procedure (reference condition = 0) and the modified procedure (comparison condition = 1). Therefore, to evaluate the change in likelihood of the outcome variable due to the combined effect of several predictors (or factors), b values must be interpreted as additive across the conditions being considered.

Model Interpretation

There are a number of ways to approach the interpretation of effects estimated by mixed-effects models (Barr et al., 2013; Judd et al., 2012). The most appropriate approach will depend on the specific research question. For confirmatory hypothesis testing, when the fit of the model to the data is not the main concern, there is little value in assessing improvements to model fit with the addition of each factor. A better approach is to evaluate whether regression coefficients for factors of interest and their interactions, in a maximal model that includes all of these factors, differ significantly and meaningfully from zero (Barr et al., 2013).

Before interpreting each maximal model, I used likelihood ratio tests to evaluate the extent to which the fit of a baseline model (including only random effects) was improved by the addition of all fixed factors (the maximal model) (Bates et al., 2014). Then, for hypothesis testing, the significance of effects of individual fixed factors on the outcome variable were evaluated by examining their coefficients. I used an alpha level of .05 for all inferential analyses. Effects were taken to be significant if the 95% confidence intervals (CI = b ± [1.96*SE]) around the coefficient excluded 0. When the distance of one limit of the CI from zero was negligible, I considered the range of the confidence interval and the size of the effect in making inferences about whether effects were meaningful.

As noted, logistic regression models that predict a binary categorical outcome variable give the log-odds for observing a given outcome. Log-odds values are symmetrical around 0. Values greater than 0 indicate the predicted outcome (e.g., a positive identification decision) is more likely, and values lower than 0 indicate the absence of the predicted outcome (e.g., a lineup rejection) is more likely. The difference between two conditions in the likelihood of the predicted outcome being observed can be evaluated as a log odds ratio (*lnOR),* or the ratio of the log odds gained for each condition. For ease of interpretation, I also reported odds ratios (*ORs)*, the exponents of log odds ratios, when discussing differences between levels of fixed factors. Confidence intervals for *ORs* are, likewise, the exponents of CIs for *lnORs* (*lnOR* ± [1.96 * *SE*]). I interpreted ORs to be statistically significant if their CIs excluded 1. For Gaussian regression models (e.g., those with response latency as the outcome variable), coefficients directly indexed the predicted difference in the outcome variable between levels of each binary categorical fixed factor. Aggregate statistics were also presented when relevant and useful to more fully describe identification performance (e.g., to indicate the number and proportion of observations for each type of identification response: accurate positive identification; missed identification; filler identification from target-absent and target-present lineups; and correct rejection).

Testing signal detection hypotheses: an example. In relation to the analyses in this thesis, an example of a model used to evaluate signal detection parameters of response bias and discriminability will usefully demonstrate the interpretation of a maximal model in greater detail. Table 1 shows coefficients for factors from a model created for data from Experiment 2 in Chapter 3 to test the effects of lineup type (standard simultaneous lineup; modified presence-first lineup) on discriminability and response bias. Maximal models are most easily interpreted with reference to a plot of the model (see Figure 1). Here, with reference to Table 1, I first explain how the effects of individual fixed factors are indexed. Then, with reference to Figure 1, I outline how differences between experimental cells are evaluated in a mixed design.

To test this pairwise comparison of lineup conditions I constructed a model (see Table 1) predicting decision type (lineup rejection or positive identification). Lineup type (standard simultaneous = 0; modified presence-first = 1), target presence (target-absent = 0, target-present = 1), and their interaction were fixed factors (De Carlo, 2009; Wright & London, 2009). Lineup type was varied between subjects. For each fixed factor, Table 1 shows the regression coefficient (b), with its standard error and CI. Variation attributed to random effects is also listed.

As lineup rejection was the baseline category (lineup rejection = 0, positive identification = 1), the model estimated the change in likelihood (log odds) of a positive identification with each factor. The intercept value shows that the baseline log odds of a positive identification were -1.91 for the reference lineup type (standard simultaneous = 0). The intercept term thus provides an estimate of response bias without considering any differences between lineup types.

Table 1

Fixed Effect Coefficients (and Random Effects SDs) for Mixed-Effects Model Predicting Decision Type (Lineup Rejection; Positive Identification) by Lineup Type (#Standard; Presence-First) in Chapter 3, Experiment 2 (filler identifications excluded)

	Effects	*b*	SE_b	95% CI_b (*SD*)
Random	Participant			(0.67)
	TP|Participant			(0.81)
	Stim			(0.98)
	TP|Stim			(1.24)
	Lineup|Stim			(0.20)
	TP*Lineup|Stim			(0.66)
Fixed	Intercept	-1.91		
	Lineup	-0.66*	0.24	-1.11, -0.19
	TP	2.81*	0.23	2.34, 3.27
	Lineup × TP	0.48*	0.25	0.00, 0.96

Note: * indicates significant predictors, # indicates the reference lineup condition, + Standard Deviation (*SD)* reflects variability due to the random effect. Number of observations = 2143, *n* = 64.

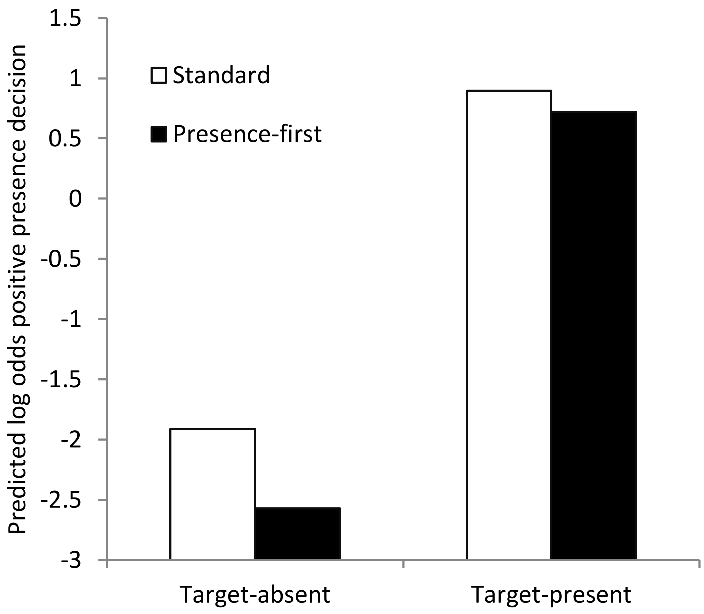


Figure 1. Plot of predicted log odds of a positive identification decision in Chapter 3, Experiment 2 by lineup type and target presence. Values greater than 0 indicate a positive identification from the mini-lineup is more likely, and values lower than 0 indicate a lineup rejection is more likely. Coefficient standard errors, indicative of the variability of the estimates, are displayed in Table 1.

First, discriminability in the reference condition (standard lineup) was indexed by the fixed effect of target-presence, where discriminability is the difference in the likelihood of making a positive identification between the two levels of target presence (target-absent; target-present). The positive coefficient of b = 2.81[2.34, 3.27] (i.e., with CIs that excluded 0) for target presence indicates that, as expected, if the guilty suspect was in the lineup (target-present), a positive identification was significantly more likely than if the suspect was innocent (target-absent).

Second, the negative coefficient of b = -0.66 [-1.11, -0.19] for lineup type indicates a change in the log odds of a positive identification between the two levels of lineup type (with CIs that do not include 0). As the standard simultaneous lineup was coded as 0, this reveals that the log odds of making a positive identification was significantly lower when the modified presence-first lineup procedure was followed, as compared with the standard simultaneous procedure. However, this term can only be taken to index the difference in response bias between lineup procedures, when the lineup type × target-presence interaction term is not also a significant predictor. That is, the effect of lineup type indexes overall differences in choosing, but is not informative about how this varies with target-presence (i.e., whether discriminability is different between the two lineup type conditions).

Third, to evaluate differences in discriminability between lineup types, the interaction term must be considered. In the model shown in Table 1, the positive coefficient for the lineup-type × target-presence interaction term indicates significantly greater discriminability for the modified presence-first procedure than the simultaneous procedure (i.e., the CI did not include 0). In other words, participants' ability to discriminate guilty suspects from innocent suspects, is greater with the change from a standard procedure to the modified procedure. The positive and significant coefficient for the interaction indicates that when the modified procedure (1) was used, rather than the standard simultaneous procedure (0), the positive effect of target presence (absent = 0, present = 1) on the log odds of a positive presence decision was greater.

Finally, to estimate the total difference in the likelihood of the outcome over several factors in a model, the coefficient values for each factor are combined. These can be plotted for specific experimental cells (see Figure 1) to most clearly demonstrate the overall pattern of effects given by the model. For example, the difference between the log odds of a positive identification in response to a target-absent (0) standard lineup (0), and that in a target-absent (0) presence-first lineup (1), considering the interaction of target presence and lineup type, would be represented by adding the values of the coefficients for each dichotomous factor that has been given a value of 1 (i.e., the comparison condition), rather than 0 (i.e., the reference condition), using the regression equation given by the model. Therefore, the size of this difference in likelihood would be evaluated by considering the sum of the coefficients for the fixed factors of lineup type, and the interaction term. For the following regression model equation:

$$y' = b_0 + b_{\text{lineup}} X_{\text{lineup}} + b_{\text{target-presence}} X_{\text{target-presence}} + b_{\textit{interaction term}} X_{\text{lineup}} X_{\textit{target-presence}}$$

the log odds of a positive identification for target-absent standard simultaneous lineups is given by substituting the corresponding values from Table 1:

$$y' = -1.91 + -0.66*0 + 2.81*0 + 0.48*0*0 = -1.91$$

whereas the log odds of a positive identification for target-absent modified presence-first lineups is given by:

$$y' = -1.91 + -0.66*1 + 2.81*0 + 0.48*1*0 = -2.57$$

In effect, a truncated regression equation is compared with the baseline equation (that includes no fixed effects) to evaluate the size of the combined effects. The truncated equation includes only the regression coefficients for fixed factors directly relevant to the specific conditions of interest in the comparison. In this example, the b values for the factors of lineup procedure and the interaction would be included in the truncated equation (but not target presence itself, as we are interested in comparing only target-absent (0) conditions).

Comparison with other measures of signal detection parameters. A recent focus of methodological innovation and debate in applied eyewitness identification research has been the application of signal detection measures to patterns of eyewitness choosing and accuracy (e.g., Gronlund et al., 2015; Mickes, 2015; Wells, Smith, & Smalarz, 2015; Wixted & Mickes, 2014, 2015). As outlined above, evaluation of signal detection measures of discriminability and response bias using measures aggregated across observations for stimuli and/or participants, is not the most appropriate approach for these kinds of data. For most experiments reported here, decision confidence was not measured (cf. Experiment 1 in Chapter 5), ruling out the use of ROC analyses.

In lineup tasks, mistaken positive identifications of filler members of the lineup represent a response type in addition to the positive recognition and negative rejection decisions that are possible responses in simple item recognition tasks in which a single stimulus is presented at test. The importance of considering filler identifications in analyses largely depends on the research question being investigated. For my purposes, patterns of filler identifications were informative for evaluating participants' likely use of familiarity and recollection in providing identification responses (e.g., Malmberg & Xu, 2007). However, for the most practical, applied question of which lineup procedure best diagnosed suspect guilt or innocence (regardless of patterns of filler identification), and how likely participants were to make a positive identification for a suspect overall, patterns of filler identification were not relevant. Therefore, I conducted mixed-effects analyses both with and without filler identifications, and interpreted models according to the specific research question at hand.

Testing dual-process hypotheses. In each experiment presented in this thesis, I investigated research questions that focused on participants' use of familiarity-based and recollective evidence from memory in eyewitness identification-type tasks. In these experiments, the pattern of errors made by participants was taken to indirectly index the extent to which decisions had been informed by the consideration of recollective evidence, in addition to familiarity-based evidence (e.g., Wixted & Mickes, 2010). Specifically, false identification decisions (or false alarms) made for innocent suspects in target-absent trials, were more likely when participants' recognition decisions were informed less by recollection and more predominantly by global match evidence of match to memory for the previously-seen target (e.g., Hockley & Consoli, 1999; Jacoby, 1991; Malmberg, 2008; Rotello & Heit, 2000; Yonelinas, 2002). Correct positive identifications were also more likely, or more likely to be made with greater confidence, if greater positive recollection of the target informed recognition decisions than if these identifications were based primarily on familiarity.

For the experiments reported in Chapters 5 and 6, I manipulated the number of times targets were presented for study (repetitions: two; six) to vary the strength of target familiarity. Orthogonally, I also varied the amount of attention participants could use to encode these targets (attention: divided; full attention) to vary the strength of recollection that was potentially available to participants in making their identification decisions. Typically, research on the use of recollection in recognition presents analyses of differences between conditions in mean hit rates (i.e., accurate positive recognition decisions) and false alarm rates (i.e., inaccurate positive recognition decisions) that are based on aggregate proportions of these kinds of recognition decisions for each participant in target-present and target-absent trials respectively (for reviews, see Malmberg, 2008; Mandler, 2008; Yonelinas, 2002). Similarly, in identification research, aggregate proportions of decisions made in each identification response category are generally compared. Here, I describe several clear patterns of recognition responses that can be described when comparing levels of experimental manipulations of ‘memory strength’ (weak, strong). Then, I outline how these patterns were evaluated using mixed-effects models.

First, higher hit rates and false alarm rates with increased ‘memory strength’ can be taken to indicate a lack of use of recollection when unstudied foils are highly similar to studied targets (at least in target-absent trials, if not in both target-absent and target-present trials). This result occurs when the greater familiarity of a stimulus leads to stronger global match to memory, and an increased likelihood of a positive recognition decision across target presence (e.g., Stretch & Wixted, 1998; Xu & Malmberg, 2007). That is, recollective evidence of mismatch is not being used to offset the misleadingly greater familiarity of unstudied foils that are highly similar to studied targets when memory strength for targets is strong. Without the use of recall-to-reject (Rotello & Heit, 1999) in this way, higher false alarm rates are evident.

Second, an increase in hit rates, without a concomitant increase in false alarm rates, indicates that participants are using recollected evidence related to studied stimuli in two ways: a) in target-present trials, to better positively identify studied targets; and b) in target-absent trials, to offset misleading familiarity associated with foils, and so offset a potential increase in false alarms due to greater familiarity (e.g., Stretch & Wixted, 1998).

Third, a strength-based mirror effect (Glanzer & Adams, 1985, 1990; Glanzer, Adams, Iverson, & Kim, 1993) may be observed with increases in familiarity and recollection (e.g., Cary & Reder, 2003; Higham, Perfect, & Bruno, 2009; Reder et al., 2000). That is, hit rates are higher, and false alarm rates are lower, with greater availability and use of both familiarity and recollection.

Aggregate measures of hit and false alarm rates are frequently used to calculate signal detection measures of response bias (e.g., *c*), and discriminability (e.g., *d*'), which is expected to increase when participants make greater use of recollection, or more recollection is available (e.g., Wixted, 2007). The mixed-effects models I used for hypothesis testing in these experiments predicted the log odds of a positive identification, and so could potentially yield estimates of these same patterns of effects. I considered model estimates (coefficients) for 1) the fixed effects of target presence, lineup procedure, and their interaction, and 2) the fixed effects of manipulated strength variables when relevant, on the likelihood of participants making a positive identification (a hit or a false alarm) for a single trial. In this way, inferences about participants' use of recollective evidence in different conditions were possible, in terms of established patterns of choosing and discriminability.

Models used to test the effects of repetitions and attention manipulations in Chapters 5 and 6 thus predicted decision type (lineup rejection; positive identification) as the outcome variable. Attention (divided = 0, full = 1), repetitions (two = 0, six = 1), target presence (target-absent = 0, target-present = 1), and their interactions were fixed factors. In Chapter 6 analyses, lineup type (standard simultaneous = 0, presence-recollection = 1) was also a fixed factor. The coefficient for the fixed factor of attention indexed any difference in witnesses' tendency to choose (response bias) between divided and full-attention conditions. Similarly, the coefficient for repetitions indexed differences in witnesses' response bias depending on the number of repetitions of the target at study. As in previously described models, the coefficient for target presence indexed the accuracy of positive decisions, independent of bias (discriminability). Variation in discriminability between levels of attention was thus indexed by the attention × target presence interaction term (Wright & London, 2009). Similarly, the repetitions × target presence interaction term indexed variation in discriminability between levels of repetitions. Finally, difference in the effect of attention on discriminability depending on repetitions condition was indexed by the interaction term for all fixed factors (repetitions × attention × target presence).

Summary

Throughout this thesis, data were analysed using Gaussian and logistic mixed-effects models. This approach allowed me to avoid the shortcomings of traditional approaches to the analysis of both categorical and continuous outcome variables when observations are gained from multiple participants to multiple stimuli. Importantly, this approach also has the benefit of increasing generalisability beyond the specific stimuli tested, and so is more likely to support replicability of results. For ease of interpretation and to provide a full description of patterns of responses, more familiar aggregate statistics are also presented for select comparisons.

References

Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J. A., Fiedler, K., Fiedler, S., Funder, D.C., Kliegl, R., Nosek, B.A., and Wicherts, J. M. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality*, 27, 108-119. doi: 10.1002/per.1919

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390-412. doi: 10.1016/j.jml.2007.12.005

Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science*, *7*, 543-554. doi: 10.1177/1745691612459060

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structures for confirmatory hypothesis testing: Keeping it maximal. *Journal of Memory and Language,* 68, 255-278. doi: 10.1016/j.jml.2012.11.001

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2014). _lme4: Linear mixed-effects models using Eigen and S4_. Retrieved from http://CRAN.R-project.org/package=lme4

Brewer, N., & Wells, G. L. (2011). Eyewitness identification. *Current Directions in Psychological Science, 20*, 24-27. doi: 10.1177/0963721410389169

Bruno, D., Higham, P. A., & Perfect, T. J. (2009). Global subjective memorability and the strength-based mirror effect in recognition memory. *Memory & Cognition*, *37,* 807-818. doi: 10.3758/mc.37.6.807

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafo, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience,* 14, 355-376. doi: 10.1038/nrn3475

Cary, M., & Reder, L. M. (2003). A dual-process account of the list-length and strength-based mirror effects in recognition. *Journal of Memory and Language,* 49, 231-248. doi: 10.1016/S0749-596X(03)00061-5

De Carlo, L. T. (1998). Signal detection theory and generalized linear models. *Psychological Methods, 3*, 186-205. doi: 10.1037//1082-989x.3.2.186

Dodson, C. S., & Dobolyi, D. G. (2015). Misinterpreting eyewitness expressions of confidence: The featural justification effect. *Law and Human Behavior*, *39*, 266-280. doi: 10.1037/lhb0000120

Gelman, A., & Hill, J. (2007). *Data Analysis using Regression and Multilevel/Hierarchical Models.* New York: Cambridge University Press.

Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness*, *5*(2), 189-211. doi: 10.1080/19345747.2011.618213

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition,* 13, 8-20. doi: 10.3758/BF03198438

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 36, 5-16. doi: 10.1037//0278-7393.16.1.5

Glanzer, M., Adams, J. K., Iverson, G. J., & Kim, K. (1993). The regularities of recognition memory, *Psychological Review,* 100, 546-567. doi: 10.1037//0033-295X.100.3.546

Gronlund, S. D., Mickes, L., Wixted, J. T., & Clark, S. E. (2015). Conducting an eyewitness lineup: How the research got it wrong. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Volume 63, pp. 1– 43). Waltham, MA: Academic Press. doi: 10.1016/bs.plm.2015.03003

Higham, P. A., Perfect, T. J., & Bruno, D. (2009). Investigating strength and frequency effects in recognition memory using type-2 signal detection theory. *Journal of Experimental Psychology: Learning Memory and Cognition, 35*, 57-80. doi: 10.1037/a0013865

Hockley, W. E., & Consoli, A. (1999). Familiarity and recollection in item and associative recognition. *Memory & Cognition, 27,* 657-664. doi: 10.3758/bf03211559

Horry, R., Halford, P., Brewer, N., Milne, R., & Bull, R. (2014). Archival analyses of eyewitness identification test outcomes: what can they tell us about eyewitness memory? *Law and Human Behavior,* 38, 94-108. doi: 1 0.1037/lhb0000060

Horry, R., Memon, A., Wright, D. B., & Milne, R. (2012). Predictors of eyewitness identification decisions from video lineups in England: A field study. *Law and Human Behavior,* 36, 257-265. doi: 10.1007/s10979-011-9279-z

Ingram, K. M., Mickes, L., & Wixted, J. T. (2012). Recollection can be weak and familiarity can be strong. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 325-339. doi: 10.1037/e520592012-514

Jacoby, L. L. (1991). A process dissociation framework - separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*, 513-541. doi: 10.1016/0749-596x(91)90025-f

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434-446. doi: 10.1016/j.jml.2007.11.007

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology, 103*, 54-69. doi: 10.1037/a0028347

Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology*, *57*, 335-384. doi: 10.1016/j.cogpsych.2008.02.004

Malmberg, K. J., & Xu, J. (2007). On the flexibility and the fallibility of associative memory. *Memory & Cognition, 35*, 545-556. doi: 10.3758/bf03193293

Mandler, G. (2008). Familiarity breeds attempts: A critical review of dual-process theories of recognition. *Perspectives on Psychological Science,* 3, 390-399. doi: 10.1111/j/1745-6924.2008.00087x

Mickes, L. (2015). Receiver operating characteristic analysis and confidence-accuracy characteristic analysis in investigations of system variables and estimator variables that affect eyewitness memory. *Journal of Applied Research in Memory and Cognition, 4,* 83-102. doi: 10.1016/j.jarmac.2015.01.003

Mickes, L., Wais, P. E., & Wixted, J. T. (2009). Recollection is a continuous process: Implications for dual-process theories of recognition memory. *Psychological Science, 20*, 509-515. doi: 10.1111/j.1467-9280.2009.02324.x

Murayama, K., Sakaki, M., Yan, V. X., & Smith, G. M. (2014). Type I error inflation in the traditional by-participant analysis to metamemory accuracy: A generalized mixed-effects model perspective. *Journal of Experimental Psychology: Learning Memory and Cognition, 40,* 1287-1306. doi: 10.1037/a0036914

Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modelling with crossed random effects and with binomial data. *Journal of Memory and Language*, *59, 413-425*. doi: 10.1016/j.jml.2008.02.002

R Core Team. (2014). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org.

Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember-know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning Memory and Cognition*, *26*, 294-320. doi: 10.1037/0278-7393.26.2.294

Rotello, C. M., & Heit, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory & Cognition*, *28*, 907-922. doi: 10.3758/bf03209339

Schimmack, U. (2012). The ironic effect of significant results on the credibility of multiple-study articles. *Psychological Methods*, *17*, 551-566. 10.1037/a0029487

Stretch, V., & Wixted, J.T. (1998). On the Difference Between Strength-Based and Frequency-Based Mirror Effects in Recognition Memory. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 24, 1379-1396. doi: 10.1037/0278-7393.24.6.1379

Weber, N., & Varga, M. (2012). Can a modified lineup procedure improve the usefulness of confidence? *Journal of Applied Research in Memory and Cognition*, 1, 152-157. doi: 10.1016/j.jarmac.2012.06.007

Wells, G. L., Smith, A. M., & Smalarz, L. (2015). ROC analysis of lineups obscures information that is critical for both theoretical understanding and applied purposes. *Journal of Applied Research in Memory and Cognition*, *4*, 324-328.doi: 10.1016/j.jarmac.2015.08.010

Wells, G. L., & Windschitl, P.D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin,* 25, 1115-1125. doi: 10.1177/01461672992512005.

Westfall, J., Judd, C. M., & Kenny, D. A. (2015). Replicating studies in which samples of participants respond to samples of stimuli. *Perspectives on Psychological Science, 10*, 390-399. doi: 10.1177/1745691614564879

Westfall, J., Kenny, D. A., Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General,* 143, 2020-2045. doi: 10.1037/xge0000014

Wixted, J. T. (2007). Dual-Process Theory and Signal-Detection Theory of Recognition Memory. *Psychological Review,* 114, 152-176. doi: 10.1037/0033-295X.114.1.152

Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review*, *117*, 1025-1054. doi: 10.1037/a0020874

Wixted, J. T., & Mickes, L. (2014). A signal-detection-based diagnostic-feature-detection model of eyewitness identification. *Psychological Review*, *121*, 262-276. doi: 10.1037/a0035940

Wixted, J. T., & Mickes, L. (2015). Evaluating eyewitness identification procedures: ROC analysis and its misconceptions. *Journal of Applied Research in Memory and Cognition*, *4*, 318-323. doi: 10.1016/j.jarmac.2015.08.009

Wright, D. B., & McDaid, A. T. (1996). Comparing system and estimator variables using data from real lineups. *Applied Cognitive Psychology,* 10, 75-84. doi: 10.1002/(SICI)1099-0720(199602)10:1%3C75::AID-ACP364%3E3.0.CO;2-E

Wright, D.B., & London, K. (2009). Multilevel modelling: Beyond the basic applications. *British Journal of Mathematical and Statistical Psychology,* 62, 439-456. doi: 10.1348/000711008X327632

Wright, D.B., Horry, R., & Skagerberg, E.M. (2009). Functions for traditional and multilevel approaches to signal detection theory. *Behavior Research Methods,* 41, 257-267. doi: 10.3758/BRM.41.2.257

Xu, J., & Malmberg, K. J. (2007). Modeling the effects of verbal and nonverbal pair strength on associative recognition. *Memory & Cognition, 35*, 526-544. doi: 10.3758/bf03193292

Yonelinas, A.P. (2002) The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language,* 46, 441-517. doi: 10.1006/jmla.2002.2864

CHAPTER 3

Split Decision: An Account of Eyewitness Decision-Making in Simultaneous Lineups and a Novel Presence-First Identification Procedure[1]

Abstract

In *best-match-first* accounts of simultaneous lineups, eyewitnesses make a sequence of decisions to arrive at their identification response. When at least one lineup member meets criterion for selection as best-match for the offender, eyewitnesses first select a best-match, before deciding whether they are the offender. We propose a framework that combines a best-match-first account of lineup decisions with a *dual-process* account of memory, to hypothesise that best-match-first decision-making entails over-reliance on familiarity and neglect of the importance of recollection. In this way, simultaneous lineups promote choosing and decrease the discriminability of guilty from innocent suspects. In two mini-lineup experiments, we investigated the timing of lineup decisions in simultaneous procedures and tested a novel presence-first lineup designed to disrupt best-match-first decision-making. For each target, participants completed a standard lineup or one of two modified two-step procedures: 1) the *presence-first* procedure, in which they first indicated target presence (*present/not present)*, before indicating which lineup member was the target (for *present* decisions); or, 2) a *best-match-first* procedure, in which they first selected the best-match, before indicating (*yes/no*) whether the best-match was the target. In Experiment 1, best-match decisions in the best-match-first procedure were made faster than standard lineup decisions

[1] Under review at *Acta Psychologica,* as Guerin, N., Weber, N., & Horry, R. (2016). Split Decision: An Account of Eyewitness Decision-Making in Simultaneous Lineups and a Novel Presence-First Identification Procedure. As first author, I designed and conducted the experiments and analyses, interpreted results, and wrote the manuscript. Nathan Weber, as my PhD supervisor, advised on the experimental and analytical design, and contributed to the interpretation of results and manuscript editing. Ruth Horry, as Adjunct Supervisor, contributed to the interpretation of results and the manuscript.

for choosers, but slower for non-choosers. Results are consistent with best-match-first decision-making when at least one lineup member exceeds a best-match criterion. Experiment 2 demonstrated greater discriminability for presence-first than standard lineups, consistent with better use of memory when best-match-first decision-making was disrupted.

Introduction

Eyewitness identification decisions are fraught with serious consequences for the course of criminal investigations and prosecutions and the individuals involved. Yet, it has long been known that identification evidence is unreliable (e.g., Brewer & Palmer, 2010; Innocence Project, 2015; National Research Council, 2014). Despite the attention given by cognitive researchers to many aspects of identification procedures, field studies show that substantial rates of identification error persist (e.g., Horry, Memon, Wright, & Milne, 2011; Wells, Steblay, & Dysart, 2015) and improvements to identification tests have produced only modest gains in diagnosticity (e.g., Brewer & Palmer, 2010). While we must accept the limits and vulnerabilities of eyewitness memory, the potential costs of identification errors lend both theoretical and practical importance to improved understanding of the decision processes that contribute to these errors. Given that eyewitness identifications are, at their core, recognition decisions, an adequate account must explain how basic memory processes operate in these complex tasks. The development of identification procedures that make the best use of eyewitness memory rests on this imperative (e.g., Brewer, Weber, Wootton, & Lindsay, 2012; Brewer & Wells, 2011).

With few notable exceptions (e.g., Dunning & Stern, 1994; Wixted & Mickes, 2014), theorising about identification decision processes has centred on Wells' (1984, 1993; see also Lindsay & Wells, 1985) relative judgment account. The idea underlying Wells' account is that an identification procedure that includes the suspect and *n* known-innocent fillers allows comparisons between these faces. To the extent that these comparisons can be made, eyewitnesses are likely to make a relative decision about match to memory for the offender (Charman & Wells, 2014). That is, eyewitnesses compare the degree of match to memory between lineup members to find the best-match, before deciding whether the best-match is the offender (Luus & Wells, 1991). Relative decision strategies contrast with absolute comparisons of each face to memory (Wells, 1984, 1993). When lineup members are presented simultaneously, comparisons between them can readily be made, and relative processes are thought to dominate (Charman & Wells, 2014; Wells, 1984). In contrast, when lineup members are presented sequentially, comparison is more difficult, and absolute match to memory is thought to be more directly evaluated, one-to-one. Absolute judgments are thus conceived to be superior to relative judgments, with lower error rates. Therefore, performance is argued to be better in sequential lineups, than simultaneous. However, the generality of the relative judgment idea limits the specificity of testable predictions it can generate. Further, it cannot account for a range of key findings from identification research (e.g., Clark & Davey, 2005; Clark & Gronlund, 2015) and its use remains largely uninformed by recognition memory and decision theory that has garnered substantial empirical support. Therefore, we present and test a framework that formalises the valuable, but general, intuitions that underlie relative judgment theory, and combines them with key tenets from dual-process models of recognition memory, to explain memory and decision processes in simultaneous lineups.

Basic memory (e.g., Malmberg & Xu, 2007) and eyewitness identification research (e.g., Brewer et al., 2012; Sauer, Brewer, & Weber, 2008; Weber & Perfect, 2012) has shown that the structure of both simple and complex recognition tasks affects how people use available memory to make recognition decisions, and what kinds of mistakes they make. Most eyewitness identification tasks ask eyewitnesses to consider multiple stimuli. It follows that eyewitnesses' use of memory for an offender to make an identification decision will depend on the multiple recognition decisions that contribute to their final response, and the order in which they are made. Therefore, a useful account of eyewitness memory in identification procedures must attend to 1) the component recognition decisions that make up an identification response, and 2) the use of memory in each of these component decisions, in the order they are made. The framework we propose draws on best-match-first accounts of simultaneous eyewitness lineup decision-making (e.g., Clark, 2003, 2008; Pozzulo & Lindsay, 1999; Wells 1984) and continuous dual-process models of recognition memory (e.g., Wixted & Mickes, 2010). We argue that eyewitnesses using best-match-first decision-making are prone to errors as a result of over-reliance on *global-match* memory processes and neglect of critical *recollective* evidence. Then, we outline our rationale, based on this framework, for a novel two-step presence-first simultaneous procedure designed to disrupt the natural, best-match-first order of simultaneous lineup decision-making and encourage better use of evidence from memory.

Best-Match-First Decision-Making

In a simultaneous lineup procedure, an eyewitness is presented with several lineup members at the same time, and must indicate the guilty offender, if present, or indicate that no lineup member matches their memory for the offender. Some researchers have conceptualised this task as being made up of two simpler, component recognition decisions (e.g., Palmer, Brewer, McKinnon, & Weber, 2010; Pozzulo & Lindsay, 1999): 1) *a best-match decision*, in which witnesses decide on the lineup member that most closely matches memory for the offender; and 2) *a presence decision*, in which witnesses judge whether or not the guilty offender is present in the lineup. Several accounts of simultaneous lineup tasks develop Wells' (1984, 1993) relative judgment idea and hold that eyewitnesses adopt a generally *best-match-first* decision-making process (e.g., Clark, 2003; Pozzulo & Lindsay, 1999). That is, eyewitnesses decide first which lineup member best matches the offender, before deciding whether or not the best-match is, in fact, the offender.

While discussions of eyewitness decision-making often touch on best-match-first accounts, we found only one direct test of the idea that eyewitnesses make separate component recognition decisions in a specific order. Pozzulo, Crescini, and Lemieux (2008) tested their best-match-first account of the order of simultaneous lineup component decision-making, *Two Judgment Theory* (see also Pozzulo & Lindsay, 1999), using a single target-absent lineup in an identification paradigm. A standard procedure was compared with a modified procedure that limited lineup exposure to 2 s. The limited exposure procedure was designed to 1) allow eyewitnesses time to make the first (best-match) decision, but 2) curtail the time available to make the second (offender-presence) decision. Eyewitnesses interrupted by limited exposure were predicted to be more likely to positively identify their (innocent) best-match selection than non-limited eyewitnesses. No significant or meaningful differences in identification performance were found, with identical rates of false identifications (.39 [95% CI: .25, .55], n=14, n=15) and correct rejections (.61 [95% CI: .45, .75], n=22, n=23) for limited and non-limited procedures respectively, but the study was underpowered. Therefore, there is no reliable empirical evidence about whether simultaneous identification decisions follow a best-match-first order of component recognition decisions. Setting a lineup exposure duration to test decision order in this way is problematic. Context and witness factors vary considerably between lineups for different offenders, and witness response latencies vary with them (e.g., Weber, Brewer, Wells, Semmler, & Keast, 2004). Pozzulo et al.'s (2008) choice of a 2 s deadline was somewhat arbitrary, made without evidence about the length of response lag that would allow for a best-match decision, and yet cut off the later presence decision. The same would be true for any other nominated response lag. Therefore, we adopted a different approach to testing the order of component recognition decisions.

Response latency and identification decision models. We used response latency as an index of the temporal order of component recognition decisions (cf. Brewer & Weber, 2008). Reaction time data have been important in the development of theories of recognition memory (e.g., Ratcliff & Rouder, 1998; Ratcliff & Starns, 2009). Eyewitness identification researchers have examined decision response latency as a potential marker of identification accuracy (e.g., Brewer, Caon, Todd, & Weber, 2004; Dodson & Dobolyi, 2016) or decision strategy (e.g., Brewer, Gordon, & Bond, 2000). But testing predictions for response latency also provides a unique way to evaluate decision-making in paradigms in which each participant completes multiple recognition tests. This approach allowed us to evaluate consistencies in responses to diverse face stimuli, while also accounting for individual response variability. The goals of a dynamic account of recognition and decision processes in identification procedures are to predict how decision processes progress over time before an identification response is given, and to understand what evidence informs both accurate eyewitness decisions and errors (e.g., Ratcliff, 1978; Ratcliff & Starns, 2009; Wagenmakers, Van Der Maas, & Grasman, 2007). A dynamic account will consider 1) the likelihoods of positive identifications (choosing) and lineup rejections (non-choosing), and 2) decision latencies. Predictions about the kinds of recognition decisions made in a specific task are thus testable by examining response latencies (Ratcliff & Rouder, 1998). In this paper, our aim is not to provide a formally specified dynamic model of simultaneous decisions. Rather, we propose that patterns of response latency across decision types can be analysed to evaluate macro predictions about identification decision structure. Specifically, response latency can provide a useful preliminary index of the order of simpler recognition decisions made by eyewitnesses to arrive at an overall identification decision. This early evidence can then inform the development of more detailed models (Brewer & Weber, 2008; Clark, 2008; Gronlund, Mickes, Wixted, & Clark, 2015).

We took Clark's (2003, 2008) WITNESS model of best-match-first decision-making in simultaneous identification tasks as a starting point. In WITNESS, three kinds of lineup responses are possible: lineup rejections; best-match identifications; and best-match don't-knows. For all responses, the match of lineup members to memory for the offender is evaluated. In situations when no lineup member is a good-enough match to be meet criterion for judgment as a best-match candidate, the lineup is rejected outright. If at least one lineup member is a good-enough match, evaluation of the best-match will lead to a positive identification, or a don't-know response. In a simultaneous lineup, the match quality of all lineup members can be evaluated concurrently, and a positive identification might be informed not only by the degree of match between the best-match and memory for the offender, but by factors such as the magnitude or type of difference in match between some or all lineup members (Clark, Erickson, & Breneman, 2011; Goodsell, Gronlund, & Carlson, 2010; Sauer et al., 2008; Wixted & Mickes, 2014). Clark (2003) argued that positive best-match identifications will be made faster than don't knows, and both will be faster than lineup rejections. When there is enough evidence of positive match to the offender to reach criterion for a positive identification of the best-match as the offender (a best-match-identification), this evidence will accumulate faster than when there is insufficient evidence to reach criterion and a don't-know response is given. Evidence will take still longer to accumulate when it is insufficient for a best-match candidate to be chosen at all (lineup rejection). These assumptions generally fit reported patterns of identification response latency (e.g., Dunning & Peretta, 2002; Sporer, 1993, but see also Weber et al., 2004). However, our central concern is with how the decision processes in which accumulating memorial evidence is evaluated unfold in real time, and WITNESS does not provide an explicitly temporal account of eyewitness decision rules if the lineup is not rejected outright.

WITNESS (Clark, 2003, 2008; see also Clark et al., 2011) does not specify whether the best-match decision precedes the overall identification decision when the lineup has not been rejected. In other words, if a best-match can be selected, there is no later option of rejecting the lineup (by rejecting the best-match), only of providing a don't-know response. Lineup rejections are therefore always based on a lack of sufficient evidence to consider choosing a best-match:

> *…identification decisions are made prior to rejection decisions, and if a positive identification decision is made, the second step (possible rejection) is not considered. Thus, identification decisions are primary, rejection decisions are secondary, and don't know is a response of last resort* (Clark, 2003, p.633)

For our purposes, the problem with the WITNESS model space is that it allows us to consider only that eyewitnesses make a single, overall identification decision (pick the best-match as the offender, or reject the lineup) or give a don't-know response. Therefore, WITNESS cannot test our two key questions about eyewitness decision-making. First, do eyewitnesses tend to first decide on a best-match, if at least one lineup member provides sufficient evidence of match, and only then make a second, final decision about whether that best-match is the offender? And, second, do eyewitnesses evaluate match to memory differently, or using different memorial evidence, after a best-match has been chosen? The question of when and how best-match-rejections are made is important as witnesses' willingness to choose operates in the context of lineups that are designed to yield plausible best-match candidates (e.g., Wells & Turtle, 1986).

Police lineups, and those used by eyewitness identification researchers, are generally constructed by selecting known-innocent fillers who match the description of the offender (description-matched), or the appearance of a suspect (suspect-matched) (Clark, Howell, & Davey, 2008). In a fair lineup all members will resemble the offender to a similar degree, so that the suspect is not conspicuous (Wells, 1993), but without being so similar that the task is impossible (Lindsay & Wells, 1980). Fillers are chosen to ensure a relatively high degree of similarity between members, and to the available description of the offender. As a consequence, at least one lineup member is likely to exceed an eyewitness' criterion for choosing a best-match. A familiarity-based selection of a known-innocent filler or an innocent suspect as the best-match should later be rejected by the witness as a match with available memory for the offender, rather than a don't-know response being made, if: 1) familiarity does not reach criterion for identification; or, 2) recollection of the offender is inconsistent with a match. To accurately reject best-match candidates who are not the offender, recollection of the actual offender is critical. Therefore, to further evaluate decision-making steps, and the use of familiarity and recollection in simultaneous lineups, we needed a model that includes best-match-rejections.

We adapted WITNESS (Clark, 2003) to describe the temporal order of lineup decisions for three possible lineup responses[2]: 1) outright-rejection; 2) best-match-rejection; and 3) best-match-identification (see Figure 1 for schematic model). An eyewitness will first select a best-match candidate, if there is sufficient evidence of match from at least one lineup member (if not, an outright-rejection will be made). The eyewitness will then decide whether or not the best-match candidate is the offender. If the best-match candidate does not meet criterion for match to the offender, a best-match-rejection is made. Both global-match and dual-process models (e.g., Malmberg, 2008) predict that best-match-rejections will tend to take longer than best-match-identification. Positive evidence for match will accumulate more slowly for best-match-rejections, as will recollected evidence of mismatch, consistent with observed patterns of identification response latency. Likewise, outright-rejections will tend to

[2] We do not include the possibility of *don't know* responses, as we have not tested them here.

be made slower than both best-match-rejections and best-match-identifications, as positive match evidence will accumulate more slowly still.

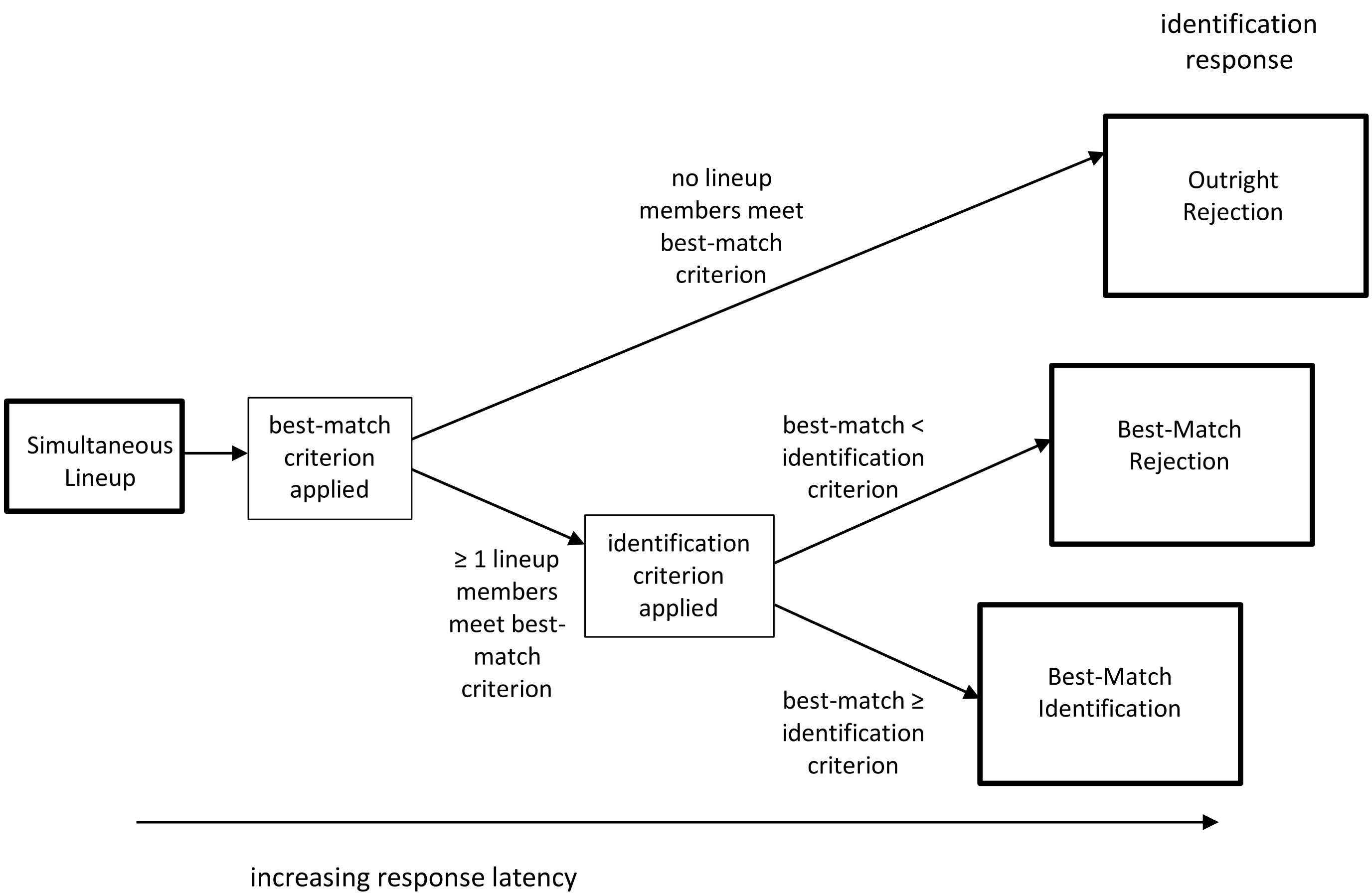


Figure 1. A schematic model of eyewitness identification decision processes for standard simultaneous lineups in the best-match-first dual-process framework.

Testing the order of lineup decisions. We set out to test the order of lineup decisions by comparing a standard simultaneous procedure with two modified procedures in which decisions were made in two steps. Each modified two-step procedure was made up of the two simpler component recognition decisions we identified as making up the standard simultaneous identification decision: a best-match decision; and a target-presence decision.

The modified procedures differed only in the order of the two decision steps. In the modified best-match-first procedure, mock eyewitnesses 1) indicated the lineup member that best matched memory for the target, before 2) indicating whether or not the best-match was the offender. In the modified presence-first procedure, we reversed this best-match-first order of lineup decisions. Mock eyewitnesses 1) indicated whether or not the target was present in the lineup, then 2) indicated which lineup member was the target (if a positive presence decision had been made). Participants were assigned to one lineup condition (standard, best-match-first, presence-first), studied a series of photos of faces, and were tested with four-person simultaneous lineups.

Our best-match-first account makes testable predictions about differences in response latency between standard identification decisions and first-decisions in the modified best-match-first procedure (i.e., the best-match decision). Overall, we expected lineup rejections to be *slower* than positive identifications. However, as the time-course of decision processes for positive identifications differs from that of rejections (of both kinds), we made separate predictions for choosers and non-choosers (see Figure 2 for a plot of framework predictions for first-decision response latency).

The best-match first account predicts that eyewitnesses using a standard procedure would first choose a best-match, then decide about offender presence, before offering an identification response. Therefore, for choosers, we expected that identification responses for standard lineups entailing these two sequential decisions (best-match + presence) would be made *slower* than first-decision responses for best-match-first lineups that required only a single decision (best-match).

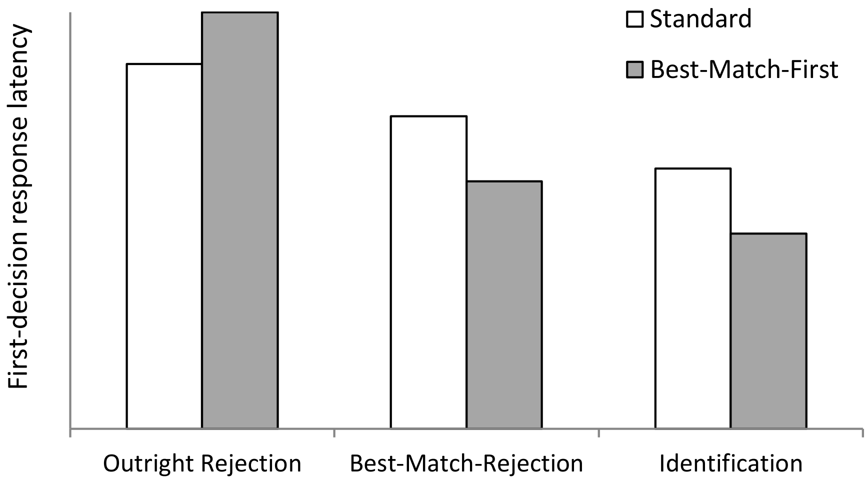


Figure 2. Best-match-first dual-process framework predictions for first-decision response latency by lineup type and decision type.

For non-choosers, the expected difference in response latency depended on the proportion of rejections arrived at via each of the two pathways for rejection. When a best-match-rejection would typically be observed for a given lineup in a standard procedure, predictions would clearly match those for choosers. Best-match-first witnesses would be able to reach their first decision (best-match) *faster* than standard witnesses would offer their final identification responses (best-match + presence). However, this prediction does not hold for rejections of test lineups that did not include a member who exceeded the best-match-selection threshold. Rather, test lineups with no plausible best-match would be rejected outright by standard lineup participants in a single step. Best-match-first participants, however, would be forced to select a best-match candidate regardless. Therefore, they would need to make an additional decision, whether by guessing or by adjusting the best-match-selection criterion, to select a best-match candidate (no-candidate + forced best-match). As a result, we expected forced best-match responses in the best-match-first procedure to be *slower* than equivalent outright-rejections in the standard procedure. It was likely that there would be some proportion of outright-rejections for our sets of stimuli. Therefore, overall, we predicted that the latency advantage for best-match-first lineups would be smaller for non-choosers than choosers. With a greater proportion of lineups that did not include a plausible best-match candidate and so would typically attract outright-rejections, the best-match-first latency advantage for non-choosers would shrink, or even be reversed.

As outlined below, we expected the presence-first procedure to pre-empt or disrupt the best-match-first decision-making processes that would characterise the standard and modified best-match-first procedures. We are not aware of reported tests of recognition memory that involve a binary *yes/no* response made for an array of multiple stimuli. Thus, there was no basis for specific predictions about the micro decision-processes presence-first participants would use, or their comparative response latency. However, we expected that presence-first participants would not take the two, ordered steps that characterise familiarity-based best-match-first processes (i.e., best-match + presence). Therefore, a departure from best-match-first decision-making for positive identifications would be suggested if presence-first participants did not take significantly longer than best-match-first participants to come to their first decision for positive identifications. For non-choosers, we had no specific expectation for the pattern of latency between best-match-first and presence-first procedures. Similarly, we made no predictions about differences in first-decision response latency between presence-first and standard procedures. However, exploratory comparisons were expected to be informative about the effects of the modified procedure on simultaneous identification decision-making.

A Dual-Process Account of Best-Match-First Accuracy

Alongside the best-match-first account of decision processes, a continuous dual-process account (e.g., Wixted, 2007; Wixted & Mickes, 2010) of recognition memory processes provides the second leg of the framework we propose to account for eyewitness performance in simultaneous identification procedures. Dual-process models (e.g., Malmberg, 2008; Yonelinas 2002) propose that both *familiarity* (global match between a test stimulus and previously studied target) and *recollection* (recall of specific target detail) can inform recognition decisions. The relative contributions of familiarity and recollection to recognition vary with test type and content (e.g., Cook, Marsh, & Hicks 2005; Malmberg & Xu, 2007), including face recognition (Marcon, Susa, & Meissner, 2009), and with instructions about task and memory processes (e.g., Lane, Roussel, Villa, & Morita, 2007; Lane, Roussel, Starns, Villa, & Alonzo, 2008; Rotello, Macmillan & Van Tassell, 2000; Starns, Lane, Alonzo, & Roussel, 2007). Recollection retrieval is relatively slower and more effortful than familiarity retrieval, particularly when not cued by the presence of the studied target stimulus itself (e.g., Guerin, S.A., Robbins, Gilmore, & Schacter, 2012). In some tasks, people neglect recollection, even when available and needed to avoid false recognition decisions (e.g., Malmberg & Xu, 2007). Here, we focus on two major research questions. First, do best-match-first decision processes lead eyewitnesses to rely too much on potentially misleading familiarity and weigh available recollection too lightly in simultaneous identification decisions? Second, will the modified presence-first procedure disrupt potentially maladaptive processes to outperform a standard procedure by increasing the weight of recollective evidence in eyewitnesses' decision-making?

Several lines of evidence suggest that best-match-first decision processes are likely to encourage over-reliance on familiarity: 1) eyewitnesses' tendency to choose from a lineup; 2) the similarity of known-innocent fillers and innocent suspects to an offender; and 3) the structure of component decisions. We discuss each, before outlining how the presence-first procedure would encourage greater use of recollection.

First, eyewitness identification research has established that witnesses tend to be overly willing to choose from a lineup across a variety of viewing and test conditions (e.g., Brewer & Wells, 2006; Malpass & Devine, 1981). High rates of false identification, indicative of a liberal response criterion, have been interpreted to reflect a tendency to assume the offender will be in the lineup (e.g., Wells & Seelau, 1995). An extension to this interpretation rests on the effect of lineup composition on eyewitnesses' use of familiarity and recollection. Description- and suspect-matched lineups include multiple filler lineup members that are very similar to each other, the offender, and/or the innocent suspect. As recognition tests, these multiple test stimuli will evoke a substantial sense of familiarity, due to their similarity to the offender. This strong familiarity of lineup members is likely to encourage eyewitnesses focused on finding the offender to make familiarity-based positive recognition decisions for targets, fillers, and innocent suspects (Meissner, Tredoux, Parker, & MacLin, 2005) and neglect the importance of recollection (Gronlund, 2005).

Second, this high target-filler similarity that characterises lineup tasks is akin to that of two common lab task types: a) associative recognition, in which unstudied test stimuli are highly similar to studied targets; and b) item recognition, in which unstudied test stimuli evoke considerable target-related familiarity (although, less so than associative tasks). Under conditions that encourage a lack of attention to available recollection, tests of both associative (e.g., Malmberg & Xu, 2007) and item recognition (e.g., Cook, Marsh, & Hicks, 2005) reliably find substantial rates of false recognition due to the misleading familiarity of unstudied test stimuli that closely resemble studied targets. In such conditions, participants need to recall mismatching detail related to the studied target (i.e., *recall-to-reject*, Rotello & Heit, 2000; see also *recollection rejection*, Brainerd, Reyna, Wright & Mojardin, 2003), or appropriately consider the absence of matching recollected detail (e.g. Gallo, Bell, Beier, & Schacter, 2006; Guerin, S.A., et al., 2012), to accurately reject misleadingly familiar, but unstudied, foils. Participants have been shown to neglect potentially available recollection that would help to avoid false recognition when they underestimate the likely similarity of new foils to old targets (Malmberg and Xu, 2007) or have not considered how recollection would be useful (e.g., Lane et al., 2007; Lane et al., 2008; Rotello, Macmillan, & Van Tassel, 2000). Moreover, when vital recollective detail is not available, participants might make positive recognition decisions on the basis of familiarity alone, even when they would have used recollection if available (e.g., Gronlund & Ratcliff, 1989; Guerin, S.A., et al., 2012; Hintzmann & Curran, 1994). This possibility is particularly important for eyewitness identification evidence, as the recollective detail available to eyewitnesses might often be limited by poor viewing conditions at the time of the crime, or lengthy delays before a suspect is apprehended.

In addition, Malmberg and Xu (2007) found that variability in the degree of target-foil similarity within a test-list affected false recognition rates. When the test-list included both foils with high target-foil similarity, and foils with relatively lower target-foil similarity, participants made more false recognition decisions for high-similarity foils, than when all foils on the test-list had high similarity. The presence of low-similarity foils seems to have misled participants to assume that target familiarity would be reliably higher than foil familiarity, and so familiarity alone would support accurate discrimination of targets from foils. As a result, participants did not use recall-to-reject to counter the misleading familiarity of high-similarity foils and avoid false recognition. For the same task, with an additional brief delay before participants could provide their recognition response to a test stimulus, false recognition rates for high-similarity foils were lower. This result indicates greater use of recollection to avoid the risk of false recognition created by substantial, but misleading, foil familiarity. The variability of target-foil similarity in the test-list parallels that of many identification lineups in that their functional size (i.e., the number of members who closely resemble the offender) is frequently lower than their nominal size (Brigham, Meissner, & Wasserman, 1999). In a lineup task, then, even if some lineup members were less highly similar to the offender (and so more readily rejected as a potential best-match), familiarity-based responding might yet be cued.

Finally, the two simpler component recognition decisions that together make up a simultaneous identification decision in the best-match-first account can be likened to an *n*-alternative forced-choice recognition task (the best-match decision) and a yes/no recognition task (the offender-presence decision). Hintzmann (1988) explained that when foils and targets are similar in forced-choice tasks, their familiarity distributions will largely overlap, but targets would still be expected to be more familiar than foils. Familiarity-based responding can thus support accurate performance in forced-choice tasks even when participants do not make use of recollective detail (e.g., Migo, Montaldi, Norman, Quamme, & Mayes, 2009). Cook et al. (2005) demonstrated that recollection tends to be used in a specific task only if and when it is perceived to be useful. As recollection of target detail is not needed to find the closest global match in the lineup, eyewitnesses are likely to rely on familiarity to make their initial best-match decision, even if recollection of the offender is potentially available.

In itself, this initial reliance on familiarity would not increase the likelihood of false identification. However, we argue that after eyewitnesses have made a best-match decision relying on familiarity, they will not automatically re-set expectations for the subsequent yes/no presence decision. This idea is supported by the observation that people will neglect effortful recollection in challenging tasks if they are not informed of this strategy (e.g., Lane et al., 2007; Lane et al., 2008; Rotello, Macmillan, & Van Tassel, 2000; Starns et al., 2007). The component decisions are not structurally or temporally separate in the standard procedure, providing no cues to re-set task expectations after the best-match is selected. If eyewitnesses do not adjust to the different evidential needs of the second component task, and continue to rely on the familiarity evidence used to find the best-match, recollection will be neglected when judging whether or not the best-match is the offender.

Familiarity alone will not support accurate presence decisions for highly similar, innocent best-match candidates from target-absent lineups. In the presence decision, eyewitnesses do not evaluate the comparative match strength of simultaneously presented alternatives, but the match to memory of a single test stimulus. In old/new recognition tasks with substantial target-foil similarity, reliance on the misleading familiarity of new test stimuli is known to lead to their false recognition as old (e.g., Hintzmann & Curran, 1994; Rotello & Heit, 2000). In the case of eyewitness identification, we propose that best-match-first decision processes would leave eyewitnesses prone to familiarity-based identification decisions. They would be likely to falsely identify a known-innocent filler or an innocent suspect who has similar-enough appearance to the offender (Gronlund, 2005; Meissner et al., 2005).

Evidence for eyewitnesses' use of familiarity and recollection in identification tests is limited. Dual-process models have only rarely been extended to eyewitness identification tasks (e.g., Carlson & Gronlund, 2011; Gronlund, 2005; Lane & Meissner, 2008; Meissner et al., 2005; Mickes, 2015; Palmer et al., 2010; Sauerland & Sporer, 2009). Studies directly investigating the contributions of familiarity and recollection to simultaneous lineup decisions have relied heavily on subjective *Remember-Know* ratings (e.g., Gardiner & Richardson-Klavehn, 2000; Tulving 1985). Remember-Know ratings do not map directly on to underlying familiarity and recollection retrieval processes, and their use is context-dependent (e.g., Bodner & Lindsay, 2003; Tulving, 1989). This seriously limits the interpretation of these results as a reliable index of the retrieval processes that have contributed to identification responses, and in what proportions. If both familiarity and recollection may contribute, and be weighted, in any combination in a recognition decision (e.g., Ingram, Mickes, & Wixted, 2012; Mickes, Wais, & Wixted, 2009), binary ratings of the experience of remembering will provide a coarse representation of the evidence used, and their use will vary with individuals, contexts, and task-types (e.g., Tousignant, Bodner, & Arnold, 2015). While there is evidence that witnesses who report using recollection are more likely to be accurate (e.g., Mickes, 2015; Palmer et al., 2010), there is no direct evidence about the use of recollection in eyewitness identification decisions.

The Presence-First Lineup: Improving Use of Recollection

Decades of eyewitness identification research have centred on the observation that if the structure of the simultaneous lineup task encourages poor use of eyewitness memory in the high-consequence decision-making context of criminal investigations, the most obvious way to better test suspect guilt is to re-structure the task (e.g., Brewer et al., 2012; Lindsay & Wells, 1985; Sauer et al., 2008; Wells & Luus, 1990). Leading researchers have noted that effective restructuring of identification tasks must be informed by empirically tested theory of basic memory and decision processes (e.g., Brewer & Wells, 2011; Gronlund et al., 2015). Yet, surprisingly little research has taken this approach. The development of novel identification tests after the innovation of the sequential lineup (Lindsay & Wells, 1985) has been minimal, and largely ad hoc. A great deal more empirical evidence has demonstrated that non-structural procedural modifications designed to improve people's use of memory, such as warnings and instructions, do not have reliable effects across tasks, test contexts, and individual variability in either recognition (e.g., Lane et al., 2007, 2008; Neuschatz, Benoit, & Payne, 2003; Neuschatz, Payne, Lampinen, & Toglia, 2001) or identification tasks (e.g., Brewer & Palmer 2010; Clark, 2005). However, structural differences between tasks designed to test memory for the same stimuli or events have consistently been shown to shape people's approach to decision-making in predictable ways (e.g., Brewer & Palmer, 2010, 2012; Malmberg, 2008). Therefore, we aimed to modify the structure of the simultaneous task to cue more effective decision and memory processes, while retaining the potential benefits for memory retrieval of presenting multiple test stimuli together (e.g., Goodsell et al., 2010).

We designed a novel presence-first procedure informed by our theoretical framework with two aims. First, we aimed to test the dual-process account of the neglect of recollection in standard simultaneous lineup tasks. Second, and most importantly, we aimed to test the idea that re-ordering the component best-match and presence recognition decisions made for a lineup would improve identification performance by facilitating better use of recollection. In re-ordering component recognition decisions, the presence-first procedure also explicitly separates them into two distinct decisions. We consider the potential effects of these two key structural changes to the simultaneous task (separating and re-ordering decisions), before outlining predictions for patterns of witnesses' identification responses for standard and modified procedures.

Separating component decisions. We have argued that once an eyewitness chooses a best-match from a simultaneous lineup based on familiarity, they are unlikely to re-set task expectations to decide on offender-presence. Researchers have tested the idea that simply splitting the task into two separate recognition steps or tasks, without changing the sequence of best-match and presence decisions, might effectively cue eyewitnesses to adjust task expectations for the second decision (Pozzulo & Balfour, 2006; Pozzulo, Dempsey, Corey, Girardi, Lawandi & Aston, 2008; Pozzulo & Lindsay, 1999; Weber & Varga, 2012). However, tests by Pozzulo and colleagues, and Weber and Varga (2012), of modified best-match-first lineup tasks provide no substantive evidence that splitting the simultaneous decision in this way resulted in better use of memory. In a three-step procedure in a mini-lineup design, Weber and Varga's (2012) participants made a best-match selection, rated their confidence that the best-match was the target, then indicated (*yes/no*) whether the best-match was the target. Greater discriminability than for a standard simultaneous lineup was reported, largely due to less false identification in the modified procedure. However, the small accuracy advantage disappears when the higher proportion of known-innocent filler identifications from target-present lineups in the modified procedure is considered. Results are more consistent with the focus on best-match selection having reinforced familiarity-based responding, rather than having cued participants to re-set task expectations to attend to the recollection needed to reject high-similarity fillers and innocent suspects. However, as the manipulation of decision separation was confounded by the ecphoric confidence rating step, and the greater length of the procedure, we cannot confidently anticipate that the same pattern of results would be observed when decisions were separated in the absence of confounding factors.

Pozzulo and colleagues tested a similar two-step best-match-first procedure, an *elimination lineup*, in a series of experiments using an identification paradigm (Pozzulo & Balfour, 2006; Pozzulo, Dempsey, et al., 2008; Pozzulo & Lindsay, 1999). In the elimination procedure, participants first selected a best-match, then indicated whether the best-match was the offender. Overall, while results suggest the procedure might have affected response bias and/or discriminability, these measures were not evaluated. All studies were substantially underpowered and, in two, the lineup manipulation was confounded by marked differences in instructions about memory evaluation (Pozzulo & Lindsay, 1999; Pozzulo & Balfour) and the strength of cautions against false identification (Pozzulo & Balfour, 2006). The potential effect of confounding instructions on ecphoric comparison (e.g., Brewer et al., 2012; Sauer et al., 2008), internal study-context reinstatement (e.g., Evans, Marcon, & Meissner, 2009; Hanczakowski, Zawadzka, & Coote, 2014), and/or willingness to choose (Malpass & Devine, 1981), prevents useful interpretation of participants' use of memorial evidence.

Overall, tests of modified best-match-first procedures have not provided conclusive evidence about how eyewitnesses' use of information from memory was affected by separating component decisions. This is an important question for the development of improved lineup procedures, as the effect of task structure on eyewitnesses' use of familiarity and recollection is likely to be more automatic (and less susceptible to witness variables) than their responses to other task modifications (e.g., Malmberg & Xu, 2007). Indeed, the sequential (Lindsay & Wells, 1985) and confidence rating (Brewer et al., 2012) lineup formats also disaggregate the iconic simultaneous task into several focal and delimited decisions. Greater understanding of the effects of decision type and order in these tasks is needed to support predictions across varying conditions. The experiments we present here are thus an important test of the effect of separating component decisions in both two-step procedures on eyewitnesses' use of memorial evidence.

Re-ordering component decisions. We have argued that best-match-first decision-making contributes to eyewitnesses' neglect of the importance of recollected detail. Therefore, pre-empting a best-match-first approach by asking witnesses to make different decisions about a simultaneous lineup, would also potentially pre-empt witnesses' neglect of the importance of recollective detail in making their identification decisions. Akin to decision-making in *n*AFC tasks, selecting a best-match is likely to focus attention on positive evidence of match to the offender, rather than evidence of mismatch. Concentration on the degree of positive match between lineup members is likely to encourage over-reliance on their global match or familiarity, and a lack of attention to mismatching detail that might diagnose innocence. As the amount of encoded or retained recollection may be limited for many witnesses, and more effortful to retrieve than familiarity, neglect is particularly likely (e.g., Guerin, S.A., et al., 2012). In contrast, tasking a witness to make only a binary old/new decision about whether the offender is present in the lineup, implicitly entails weighing positive match evidence against evidence of mismatch, and is more likely to evoke the use of recall-to-reject (Malmberg, 2008), the distinctiveness heuristic (Gallo et al., 2006), or other post-retrieval memory monitoring mechanisms. Importantly, the lineup stimuli being evaluated have been presented simultaneously, with the potential benefit of multiple cues for maximal retrieval and maintenance in working memory of available memory for the offender (e.g., Humphreys, Bain & Pike, 1989). Retrieved memorial evidence is available when evaluating every lineup member. This compares with the cumulatively cued memory that is available (but may be difficult to maintain in working memory) when lineup members are presented sequentially. By explicitly delaying the potential selection of a single offender from the lineup, a presence-first procedure would thus pre-empt the over-concentration on positive global match evidence to find a best-match that bedevils standard simultaneous decision-making. The presence-first procedure would also pre-empt the premature focus on, and commitment to (e.g., Deffenbacher, Bornstein, & Penrod, 2006; Goodsell, Neuschatz, and Gronlund, 2009; Schreiber & Sergent, 1998), a single lineup member, that is a risk when the purpose of the task is framed as finding and evaluating the best-match.

Evaluating the presence-first lineup. We conducted two experiments designed as a preliminary test of our theoretical framework and the effectiveness of the novel presence-first procedure in promoting better use of recollection. We argue that once eyewitnesses have made a best-match judgment, they are prone to decide that the lineup member judged as the best-match is also the guilty offender, even if the quality of match to memory for the offender is relatively low. That is, the criterion for identifying the best-match as the offender tends to be set relatively low. As a result, when a target-absent lineup includes an innocent suspect or filler who resembles the offender closely enough to allow a best-match choice, the eyewitness is likely to go on to make a false positive presence decision. In this way, false identifications are encouraged by cues to decision-making embedded in the task structure itself. We expected the presence-first procedure to discourage this focus on the best-match, and encourage better use of recollected detail of the offender. Therefore, we expected greater discriminability for the presence-first procedure than for either the standard simultaneous or modified best-match-first procedures.

The Experiments

We used a face recognition, mini-lineup paradigm to compare standard, best-match-first, and presence-first simultaneous lineup procedures. Participants studied a series of photos of faces, and gave identification responses for a series of four-person lineups, one per target face. The mini-lineup paradigm allowed us to collect multiple data points per participant, and evaluate regularities of responding across multiple participants and stimuli. Gathering large amounts of data is particularly important when testing a new theoretical explanation for observed patterns of responses in a specific task, or evaluating performance in a previously untested task (cf., Gronlund et al., 2015; Sauer et al., 2008; Weber et al., 2004). In contrast, in an eyewitness identification paradigm that uses video or real life mock crime stimuli to more closely simulate police identification contexts, witnesses make an identification decision for only a small number of offenders (generally, only one or two), providing limited data points per participant. Such an approach requires a large sample to power valid conclusions. Moreover, substantial differences in eyewitness responses for different lineups, crime scenarios, and offenders, seriously constrain the generalisation of findings from one experiment using the more veridical identification paradigm.

Participants in our experiments followed the standard, best-match-first, or presence-first procedure for all lineups. We had two major research questions, and recap key predictions for each. First, would participants in the standard simultaneous procedure follow a best-match-first order of decision-making, as outlined in our best-match-first dual-process framework? If the best-match-first dual-process framework explains patterns of responses from standard simultaneous identification tasks, we expected to find the first direct evidence that witnesses first make a best-match judgment, before deciding on the guilt of the best-match candidate.

Specifically, we predicted that participants in the best-match-first procedure would provide their first decisions (best-match) faster than participants in the standard procedure would provide overall identification decisions (best-match + offender-presence). However, we anticipated that this difference would be smaller for lineups that were rejected than lineups from which a positive identification was made. This was because those following the best-match-first procedure would need to make an additional decision (outright-rejection + forced best-match) to select a best-match candidate from lineups that would have been rejected outright in the standard procedure (outright rejection). Additionally, we compared first-decision response latency from the best-match-first and presence-first procedures to evaluate framework predictions that participants in the novel presence-first procedure would not have used best-match-first processes.

Our second major research question was would participants following the presence-first procedure make better use of recollection to demonstrate greater discriminability of innocent from guilty suspects than participants using a standard procedure? We predicted greater discriminability would be observed for the presence-first procedure than either the standard simultaneous procedure or the modified best-match-first procedure. In addition, we compared discriminability in the best-match-first and standard procedures to directly evaluate the effect of separating, but not re-ordering, component recognition decisions.

Method

Except where indicated, both experiments followed the same method.

Participants and Design

A 3 (lineup type: standard, best-match-first, presence-first) × 2 (target presence: target-absent, target-present) mixed design was used, with target presence manipulated within-subjects.

A priori, we aimed to collect data from 30 participants per experimental condition. Ninety-four novice undergraduate students, who had not previously taken part in an eyewitness identification-type task paradigm, from Flinders University (62 female, age: *M* = 23 years, *SD* = 7 years) participated in Experiment 1 for payment or course credit. Ninety-six novice students (73 female, age: *M* = 24 years, *SD* = 8 years) completed Experiment 2. All collected data were analysed.

Materials

We used 40 sets of five description-matched colour photos constructed by Weber and Brewer (2004, 2006). Photos showed a front view of the head and neck with a neutral facial expression and background (200 × 200 pixels) and were presented on an 18” monitor (resolution: 1920 × 1080 pixels). In each set, we randomly designated a target, a target-replacement (for target-absent mini-lineups), and three foils. Face sets were divided into two blocks of 20. In each block, half of the tests were target-present (target plus foils) and half target-absent (target-replacement plus foils). All mini-lineups were presented in a row of four, with response buttons below. To enable participants to look for a specific target in each mini-lineup (analogous to a witness looking for a specific offender), rather than any studied face (Weber & Brewer, 2006), a text cue (name or occupation, by block) was presented with each target (study) and mini-lineup (test). The cue was centred above photos. Photo order in each mini-lineup, trial order in each block, and cue and target-presence assignment, were randomly determined for each participant.

Procedure

Participants completed the entire experiment on PC in individual cubicles using a mouse to click on-screen buttons in response to written prompts. After giving informed consent, they completed two blocks of trials (after four practise trials), with block order, and assignment of name or occupation cues to blocks, counterbalanced. Each block included a study phase, a visual distractor task, and a test phase. In the study phase, target photos were presented sequentially, in the centre of the screen. Each cue was presented for 1000 ms before presentation of the target, and throughout the target's exposure (1000 ms). Following each study phase, participants worked on a spatial memory task for 3 min. In each test phase, participants completed 20 mini-lineup trials. Participants were told that each mini-lineup may or may not include a target (i.e., unbiased instructions). The decision itself and response latency were recorded for all decisions. In each experiment, we manipulated only target presence, and lineup type on three levels, between-subjects.

Experiment 1. In the standard lineup condition, participants provided a single response. They identified a lineup member as the studied target (by clicking on the photo), or clicked the *Not present* button.

In the presence-first condition, participants made an identification decision in two steps. First, they indicated whether the studied face was present or not-present in the mini-lineup. Second, if participants had indicated that the studied face was present, they clicked on the face identified as the target. Otherwise, they proceeded immediately to the next lineup.

The best-match-first condition also required two steps. First, participants were asked to click on the lineup member that best matched the studied face. Once a best-match was selected, the remaining faces in the mini-lineup disappeared from the screen, leaving the best-match photo on-screen for an additional 1.5 *s*. Participants then indicated whether or not their displayed best-match selection was the studied target.

Experiment 2. The manipulation of lineup type followed that in Experiment 1, with one difference. For presence-first and best-match-first conditions, the two identification decisions were blocked, rather than interleaved. Participants first completed both blocks of trials making only their first decision for each mini-lineup (a presence or best-match decision, respectively). Before making this first decision, participants were not told that they would later be asked to make a second decision for each mini-lineup. After all trials for both blocks were completed, presence-first participants were presented sequentially with lineups for which they had made a positive presence decision, and asked to indicate the best-match lineup member. Similarly, best-match-first participants were presented sequentially with each chosen best-match photo, and asked to indicate whether or not it was the studied target.

Results & Discussion

Overview

We used an alpha level of .05 for all inferential analyses. For response latency and identification decision performance, our focus was on trial-by-trial analysis using mixed-effects models. Conceptually, these analyses can be interpreted in the same way as (ordinary and logistic) regression, but have several advantages over more common approaches. The approach allowed us to evaluate the effects of lineup type while accounting for random variability within and between non-independent observations for multiple stimuli from multiple individuals (e.g., Baayen, Davidson, & Bates, 2008; Judd, Westfall, & Kenny, 2012; Murayama Sakaki, Yan & Smith, 2014). Maximal Gaussian and logistic mixed-effects regression models that included all fixed factors of interest were evaluated to test predictions for response latency, and choosing and identification accuracy, respectively (Barry, Levy, Scheepers, & Tily, 2013; Wright & London, 2009). Data were analysed using the lme4 package (Bates, Maechler, Bolker, & Walker, 2014) for R (R Core Team, 2014).

Models were used for confirmatory hypothesis testing, rather than evaluating model fit or modelling lineup decision processes per se. Each model included fixed factors of key interest. The random effects structure varied for each model, and allowed us to evaluate fixed factors without the risk of statistical occlusion or enhancement of effects entailed in ANOVA with aggregate measures (Barr et al., 2013; Gelman & Hill, 2007; Wright & McDaid, 1996), and to maintain statistical power while avoiding inflated type I error caused by ignoring stimulus variability (Westfall, Judd, & Kenny, 2015). We prioritised the inclusion of random effects that improved fit, and avoided over-parameterising models so that our power to evaluate planned comparisons would not be limited. For models predicting first-decision response latency, we included only random intercept and random slopes for target presence, by participant and stimulus. Random effects in models that predicted identification decision type (lineup rejection; positive identification) typically included random intercept by participant and stimulus, random slopes by participant for within-subjects manipulations, and random slopes by stimulus for within-stimulus manipulations. Tables of standard deviations for random effects are included in an online supplement (Tables 10 and 11).[3]

We made planned pairwise comparisons between conditions of main interest (Gelman, Hill & Yajima, 2012) and evaluated the significance of effects of individual fixed factors on the outcome variable by examining their coefficients in the model. Effects were taken to be significant if the 95% confidence interval (*CI*) around the coefficient excluded 0. Logistic regression models, predicting a binary categorical outcome variable, give the log-odds for observing a given outcome. Log-odds *v*alues are symmetrical around 0. Values greater than 0 indicate the predicted outcome (e.g., a positive identification decision) is more likely, and values lower than 0 indicate the absence of the predicted outcome (e.g., a lineup rejection) is more likely. For ease of interpretation, we also report odds ratios (*ORs)*, the exponents of log odds ratios (*lnORs*), when discussing differences between levels of fixed factors. Confidence intervals (*CIs*) for *ORs* are, likewise, the exponents of CIs for *lnORs* (*lnOR* ± [1.96 * *SE*]). We interpreted these ORs to be statistically significant if their CIs excluded 1. For Gaussian regression models (here, those with response latency as the outcome variable), coefficients directly index the predicted difference in the outcome variable between levels of each binary

[33] Supplemental materials for this chapter are presented in Appendix 1

categorical fixed factor. Aggregate descriptive statistics are also presented when relevant as an index of identification performance (see Table 9).

In practice, positive identifications of known-innocent filler lineup members in lineup tasks are known errors. Therefore, these responses need not be considered for practical questions about which lineup procedure most reliably tests suspect guilt. However, patterns of filler identification have theoretical implications insofar as they indicate differences in the use of memory or decision processes (Wells, 2008; Wells & Turtle, 1986). Therefore, when decision type (lineup rejection; positive identification) was our outcome variable, we conducted analyses both with and without filler identifications. When response latency was our outcome variable, filler identifications were included in all analyses.

Response Latency and Decision Type

We examined the effect of lineup condition on first-decision response latency. To correct for positive skew, we applied an *ln* transformation to response latencies (*lnrl*). For standard, best-match-first, and presence-first lineups respectively, first decisions were: a) identification decisions; b) best-match decisions; and c) presence decisions. Our main focus of interest was whether first-decision response latency would be significantly faster in the best-match-first procedure than in the standard procedure. For each pairwise comparison we evaluated a model with first-decision response latency (*lnrl)* as the outcome variable. Lineup type, target presence (target-absent = 0; target-present = 1), decision type (lineup rejection = 0; positive identification = 1) and their factorial combinations were fixed factors. Except where noted, the same analyses were conducted for both experiments.

Experiment 1. First, we compared standard and best-match-first lineup procedures to test predictions of our best-match-first dual-process framework. For lineups from which a positive identification was later made, we predicted best-match-first witnesses would provide their best-match judgments faster than standard witnesses would make their positive identifications. Lineup, decision type, and the interaction terms for decision type with each of target presence and lineup were significant predictors of response latency in the model comparing standard and best-match-first lineups (see Tables 1 & 2). Results are most easily interpreted by considering the plotted model (Gelman & Hill, 2007). As predicted, best-match decisions were made faster than standard decisions (see Figure 3, Panel B). Consistent with targets generally providing stronger evidence of match than innocent suspects, responses were faster for target-present than target-absent lineups, but the effect of lineup did not differ with target presence.

Table 1

Model Fit Statistics for Mixed-Effects Model Predicting First-Decision Response Latency (lnrl) by Lineup Type in Experiments 1 & 2

Comparison	Expt	χ^2	*(df)*	*p*
Standard vs Best-Match-First (BMF)	1	192.84	7	< .001
	2	62.94	7	< .001
Standard vs Presence-First (PF)	1	92.78	7	< .001
	2	102.33	7	< .001
BMF vs PF	1	169.03	7	< .001
	2	29.28	7	< .001

Note. Number of observations for Expt 1: Standard vs BMF = 2440, *n* = 61; Standard vs PMF = 2600, *n* = 65; BMF vs PF = 2480; *n* = 62. For Expt 2: Standard vs BMF = 2600, *n* = 65; Standard vs PMF = 2560, *n* = 64; BMF vs PF = 2520; *n* = 63.

Table 2

Fixed Effect Coefficients for Mixed-Effects Model Predicting First-Decision Response Latency (lnrl) by Lineup Type (#Standard; Best-match-first) in Experiments 1 & 2

	Expt 1			Expt 2		
Fixed Effect	*b*	SE_b	95% CI_b	*b*	SE_b	95% CI_b
Intercept	8.45			8.53		
Lineup	0.19*	0.09	0.03, 0.36	0.10	0.05	-0.08, 0.28
TP	-0.02	0.04	-0.09, 0.06	-0.11*	0.04	-0.21, -0.01
Lineup × TP	0.02	0.06	-0.14, 0.09	-0.10	0.06	-0.21, 0.02
Decision Type (Dec)	0.13*	0.04	0.06, 0.20	0.03	0.04	-0.04, 0.11
Dec × TP	-0.29*	0.05	-0.39, -0.19			
Lineup × Dec	-0.27*	0.05	-0.37, -0.17			
Lineup × TP × Dec	0.04	0.07	-0.10, 0.19			

Note. * indicates significant predictors, # indicates the reference lineup condition. Number of observations for Expt 1: 2440, *n* = 61. For Expt 2: 2600, *n* = 65. Models that included the standard lineup procedure did not include random effects by stimulus due to a recording error during the experiment.

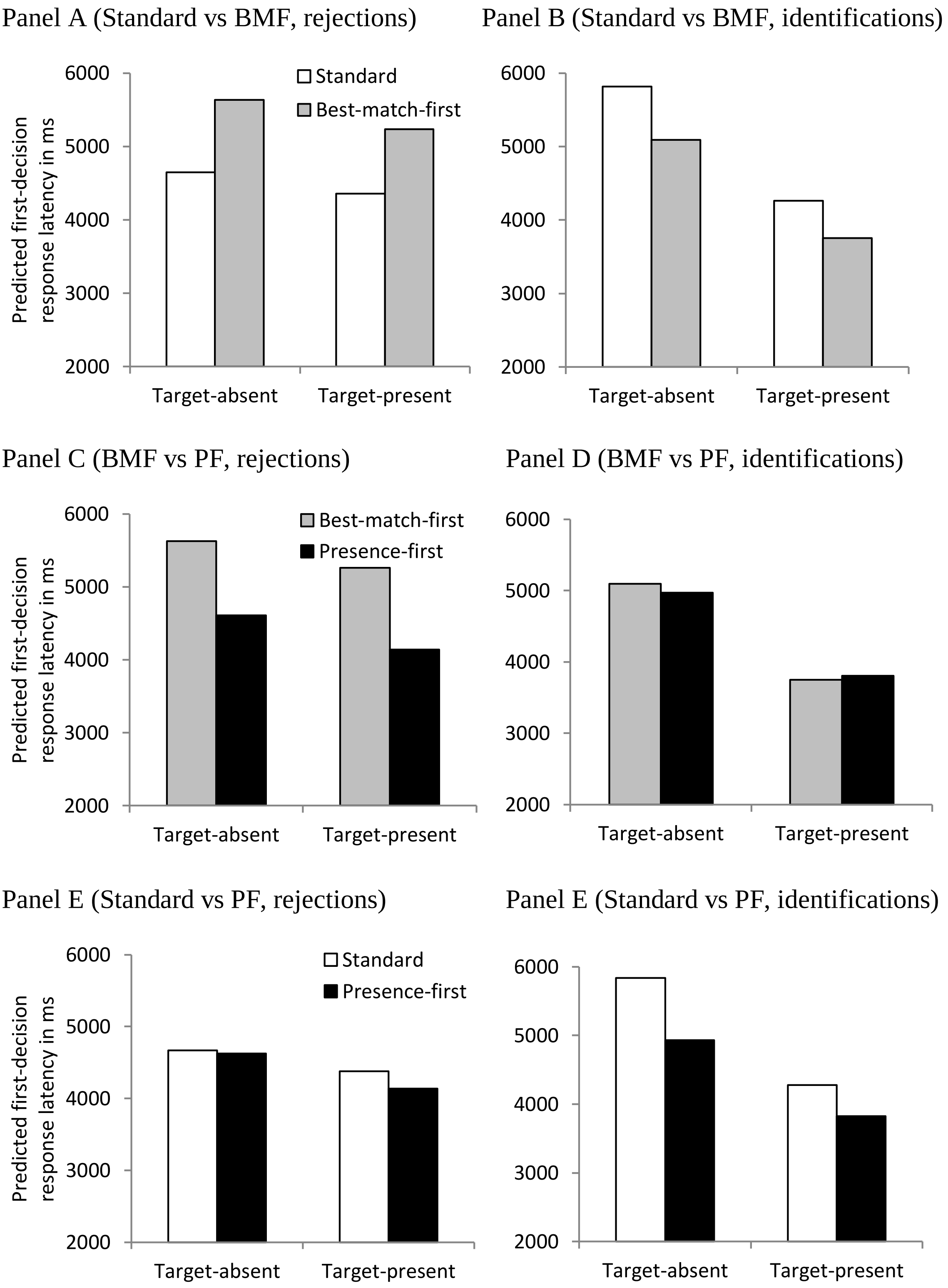


Figure 3. Plots of predicted first-decision response latency (ms) in Experiment 1 by lineup type, target presence, and decision type (lineup rejection; positive identification). Coefficient standard errors, indicative of the variability of the estimates, are displayed in Tables 2-4.

For lineup rejections, we predicted that the best-match-first advantage over the standard procedure in response speed would be smaller than for positive identifications, and potentially reversed, depending on the balance of best-match-rejections and outright-rejections in the sample. Figure 3 (Panel A) shows that the best-match-first advantage was reversed, with a significant effect of lineup such that best-match decisions were significantly slower than standard rejections. This finding is consistent with predictions of our best-match-first dual-process framework, and suggests a substantial proportion of lineups for which outright-rejections would ordinarily be made (in a standard simultaneous procedure), rather than best-match-rejections.

In sum, best-match decisions in the best-match-first procedure were faster than standard identification decisions for lineups from which a positive identification was made, but slower than standard rejections for lineups that would be rejected. Results are consistent with predictions of our best-match-first dual-process framework.

We made exploratory comparisons of the best-match-first and presence-first procedures (see Tables 1 & 3). We proposed that presence-first participants who made an initial presence decision for each lineup, without needing to point out any targets, would avoid best-match-first decision-making processes, and make better use of recollection than participants using the other procedures. There was no theoretical or evidential basis for predictions about the nature or timing of processes that would contribute to their presence decisions. However, had presence-first witnesses used best-match-first decision processes, they would first have made a best-match judgment before offering their positive presence decision based on evaluation of the best-match candidate. Therefore, presence-first participants would have made two decisions (best-match + presence) while best-match-first witnesses made only one (best-match). Accordingly, presence-first step 1 responses would have been made more slowly, than best-match-first best-match judgments. As Figure 3 (Panel D) clearly shows, there was no meaningful difference between the two modified procedures in first-decision response latency for positive identifications. Thus, our finding provides no evidence of best-match-first decision-making by presence-first participants. For non-choosers, best-match-first participants' best-match judgments were made significantly more slowly than presence-first witnesses' rejection responses, for both target-absent and target-present lineups (see Figure 3, Panel C). This result is not informative about best-match-first decision-making in the presence-first procedure, as the ratio of best-match-rejections to outright-rejections in the sample is unknown.

Table 3

Fixed Effect Coefficients for Mixed-Effects Model Predicting First-Decision Response Latency (lnrl) by Lineup Type (#Best-match-first; Presence-first) in Experiments 1 & 2

	Expt 1			Expt 2		
Fixed Effect	*b*	SE_b	95% CI_b	*b*	SE_b	95% CI_b
Intercept	8.65			8.62		
Lineup	-0.21*	0.08	-0.38, -0.05	-0.15	0.08	-0.32, 0.01
TP	-0.06	0.04	-0.14, 0.03	-0.24*	0.03	-0.31, -0.17
Lineup × TP	-0.05*	0.06	-0.17, -.06	0.10*	0.04	0.02, 0.19
Dec	-0.14*	0.04	-0.22, -0.06			
Dec × TP	-0.22*	0.05	-0.33, -0.11			
Lineup × Dec	0.23*	0.05	0.12, 0.33			
Lineup × TP × Dec	0.06	0.08	-0.09, 0.21			

Note. * indicates significant predictors, # indicates the reference lineup condition. Number of observations for Expt 1: 2480; *n* = 62. For Expt 2: 2520; *n* = 63.

Table 4

Fixed Effect Coefficients for Mixed-Effects Model Predicting First-Decision Response Latency (lnrl) by Lineup Type (#Standard; Presence-First) in Experiments 1 & 2

	Expt 1			Expt 2		
Fixed Effect	*b*	SE_b	95% CI_b	*b*	SE_b	95% CI_b
Intercept	8.45			8.53		
Lineup	-0.02	0.08	-0.17, 0.12	-0.06	0.08	-0.21, 0.10
TP	-0.02	0.04	-0.09,0.05	-0.11*	0.04	-0.19, -0.03
Lineup × TP	-0.07	0.06	-0.19, 0.03	0.09	0.06	-0.02, 0.20
Dec	0.13*	0.04	0.06, 0.20	0.02	0.03	-0.05, 0.09
Dec× TP	-0.29*	0.05	-0.38, -0.19	-0.19*	0.05	-0.28, -0.09
Lineup × Dec	-0.03	0.05	-0.14, 0.07	-0.07	0.05	-0.18, 0.03
Lineup × TP × Dec	0.10	0.07	-0.04, 0.24	0.05	0.07	-0.09, 0.19

Note: * indicates significant predictors, # indicates the reference lineup condition. Number of observations for Expt 1: 2600, *n* = 65; For Expt 2: 2560, *n* = 64. Models that included the standard lineup procedure did not include random effects by stimulus due to a recording error during the experiment.

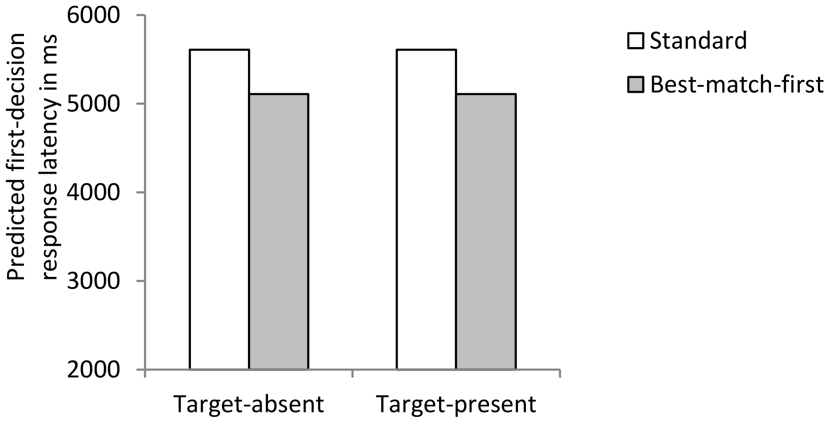
Finally, we also compared first-decision response latency between standard and presence-first procedures (see Tables 1 & 4). The framework does not make predictions for this comparison, and no significant effect of lineup was observed (see Tables 1 and 4, and Figure 3, Panels E & F). In isolation, results do not support conclusions about presence-first decision processes within component decisions.

Summary. Overall, in Experiment 1, for positive identifications, best-match judgments for the best-match-first procedure were made significantly faster than identifications for the standard procedure. For lineup rejections, best-match judgments for the best-match-first procedure were made significantly slower than those for standard (and presence-first) lineups. Therefore, predictions of the best-match-first, dual-process account for differences in first-decision response latency between best-match-first and standard procedures were supported. Comparisons of first-decision latency between best-match-first and presence-first procedures were suggestive of presence-first witnesses avoiding best-match-first decision processes.

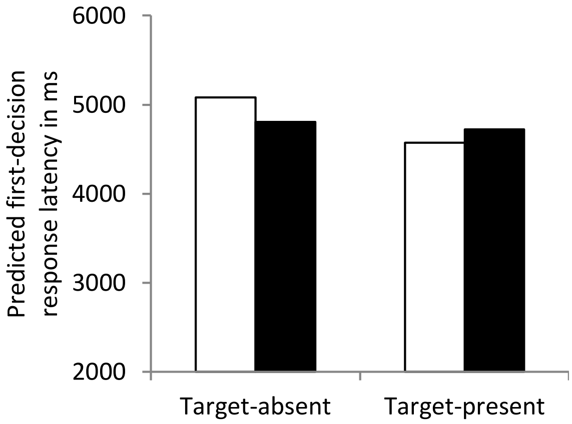
Experiment 2. Delayed second-decision accuracy for the best-match-first procedure was near-chance in Experiment 2, when two-step decisions were blocked. Therefore, for comparisons of first-decision response latency involving the best-match-first procedure, we collapsed across decision type (lineup rejections and positive identifications). Without being able to reasonably separate choosers from non-choosers, it was not possible to test our main hypothesis, but we report results for completeness (see Tables 1, 2 & 3, and Figure 4). While low second-decision performance limits the conclusions that can sensibly be drawn, results provide no evidence to contradict findings from Experiment 1 that were consistent with framework predictions.

We also compared standard and presence-first procedures (Tables 1 & 4) and evaluated the effect of lineup for positive identifications and rejections separately. As Figure 4 (Panels B and C) shows, and replicating the general pattern observed for Experiment 1, no significant effects of lineup were observed. In sum, findings matched across experiments.

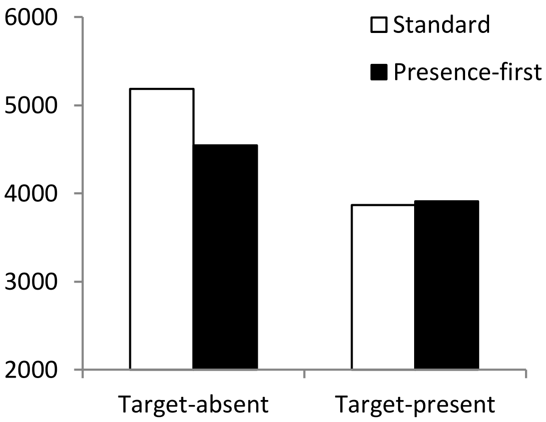
Panel A (Standard vs BMF)



Panel B (Standard vs PF, rejections)



Panel C (Standard vs PF, identifications)



Panel D (BMF vs PF)

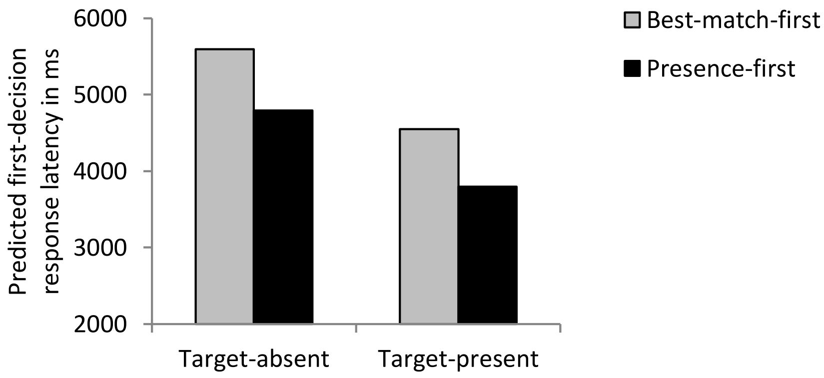


Figure 4. Plots of predicted first-decision response latency (ms*)* in Experiment 2 by lineup type and target presence in Panels A and D, and by lineup type, target presence, and decision type (lineup rejection; positive identification) in Panels B and C. Coefficient standard errors, indicative of the variability of the estimates, are displayed in Tables 2-4.

Decision Type and Accuracy (Choosing and Discriminability)

We compared the likelihood of choosing (positive identifications of suspects and fillers) pairwise between each of the standard and modified (best-match-first; presence-first) lineup procedures. We also examined the effect of lineup on the accuracy of lineup decisions, using a measure of comparative discriminability. Analyses were conducted both with and without filler identifications. Both measures are important for the evaluation of witnesses' memory performance, and procedural reliability, respectively (Wells, 2008). Therefore, we report statistics for all analyses for both experiments. As patterns of effects were similar with or without the inclusion of filler identifications, we discuss these patterns once, but make specific reference to patterns of filler identification when pertinent. When filler identifications were included, our measure of discriminability indexed the accuracy of target *presence* decisions, regardless of whether the lineup member identified was the suspect (innocent or guilty) or a filler. When filler identifications (from both target-absent and target-present lineups) were excluded, our measure of discriminability indexed the accuracy of positive identification decisions (i.e., witnesses' ability to discriminate guilty from innocent suspects).

Logistic regression models, including mixed-effects logistic regression (e.g., Wright & London, 2009), can be used to estimate signal detection parameters of discriminability and response bias (De Carlo, 1998). To test each pairwise comparison of lineup conditions we constructed a model predicting decision type (lineup rejection or positive identification, see Tables 5-8). Lineup type, target presence (target-absent = 0, target-present = 1), and their interaction were fixed factors. The effect of target presence provided an index of discriminability (the accuracy of decisions, independent of bias), while the interaction term indicated how much discriminability varied between lineup procedures (Wright & London, 2009). In the absence of a significant interaction, the effect of lineup indexed any difference in response bias (witnesses' tendency to choose). Identical analyses were carried out with data from both experiments. However, identification decisions were made in different contexts in each experiment. In Experiment 1, component decisions in the modified two-step procedures were made consecutively for each lineup before proceeding to the next lineup. On the other hand, in Experiment 2, the two decisions were made in separate blocks. Here, we report results from both experiments, and discuss how patterns of results differed when two-step decisions were blocked, rather than made in immediate succession.

Experiment 1. In Experiment 1, contradicting our prediction that discriminability would be significantly better in the presence-first procedure than either the standard or the best-match-first procedure, the lineup × target presence interaction term did not significantly predict a difference in discriminability between lineup procedures (see Table 5 for model fit statistics, Tables 6-8 for model coefficients, and Figure 5 for model plots). Examining Figure 5, the pattern of effects on discriminability for all comparisons follows our predictions, but no lineup procedure demonstrated significantly greater discriminability than any other. However, consistent with lower response bias for the presence-first lineup procedure, there was a significant effect of lineup in the model comparing the presence-first and best-match-first procedures, when filler identifications were excluded. As this difference was not significant when filler identifications were included, the finding suggests that presence-first witnesses were not making significantly better use of recollection to avoid mistakenly identifying fillers at a higher rate than best-match-first witnesses. However, as noted, all comparisons were in the direction of predictions. There was no evidence that lineup type significantly predicted response bias (i.e., the likelihood that a witness would make a positive identification from a lineup), for other pairwise comparisons between lineup procedures (with or without filler identifications included).

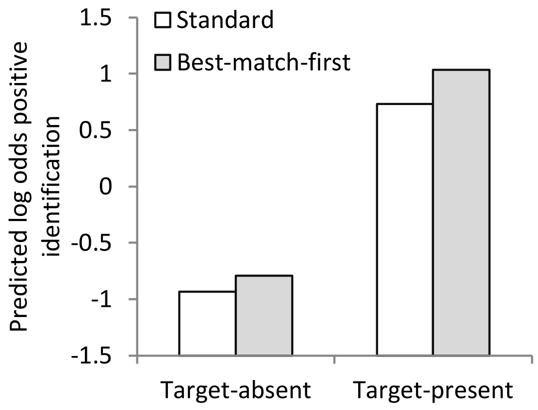
Table 5

Model Fit Statistics for Mixed-Effects Model Predicting Decision Type (Positive Identification; Lineup Rejection) by Lineup Type in Experiments 1 & 2

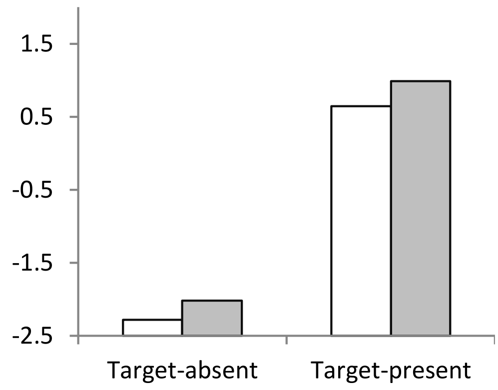
Filler Identifications	Comparison	Expt	χ^2	*df*	*p*
Included	Standard vs BMF	1	73.65	3	< .001
		2	40.96	3	< .001
	Standard vs PF	1	96.19	3	< .001
		2	67.14	3	< .001
	BMF vs PF	1	69.03	3	< .001
		2	63.64	3	< .001
Excluded	Standard vs BMF	1	106.41	3	< .001
		2	64.97	3	< .001
	Standard vs PF	1	126.86	3	< .001
		2	77.78	3	< .001
	BMF vs PF	1	85.13	3	< .001
		2	107.79	3	< .001

Number of observations for Expt 1, fillers included (excluded): Standard vs BMF = 2440 (2054), n = 61 (61); Standard vs PF = 2600 (2063), n = 65 (61); BMF vs PF = 2480 (2066), n = 62 (61). For Expt 2: Standard vs BMF = 2600 (2028), n = 65 (65); Standard vs PF = 2560 (2143), n = 64 (64); BMF vs PF = 2520 (2055), n = 63 (63).

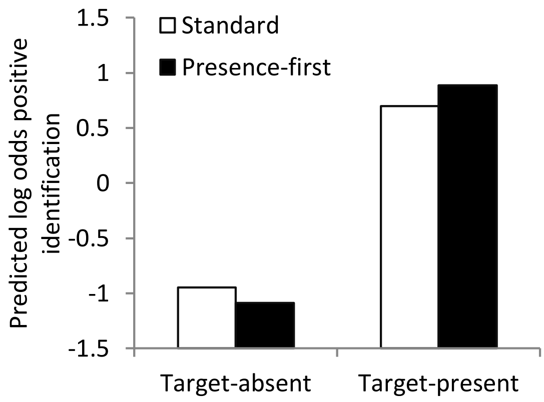
Panel A (Standard vs BMF, fillers included)



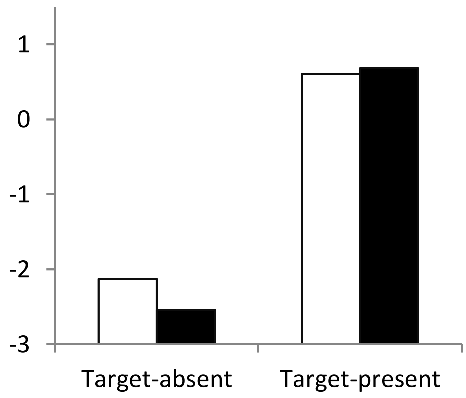
Panel B (Standard vs BMF, fillers excluded)



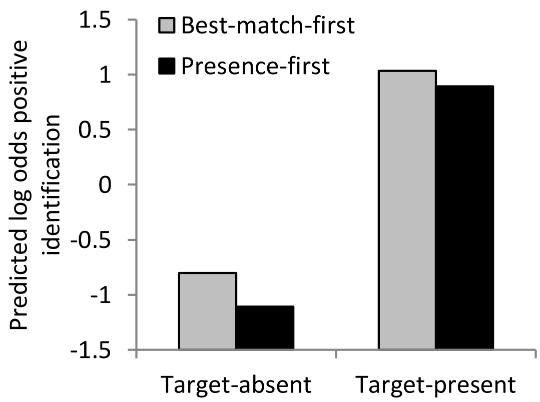
Panel C (Standard vs PF, fillers included)



Panel D (Standard vs PF, fillers excluded)



Panel E (BMF vs PF, fillers included)



Panel F (BMF vs PF, fillers excluded)

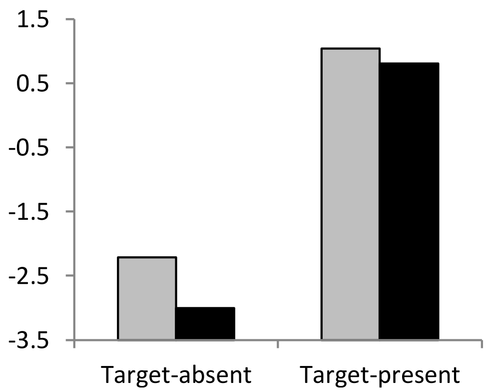


Figure 5. Plots of predicted log odds of a positive identification decision in Experiment 1 by lineup type and target presence. Values greater than 0 indicate a positive identification from the mini-lineup is more likely, and values lower than 0 indicate a lineup rejection is more likely. Coefficient standard errors, indicative of the variability of the estimates, are displayed in Tables 6-8.

Table 6

Fixed Effect Coefficients for Mixed-Effects Model Predicting Decision Type (Lineup Rejection; Positive Identification) by Lineup Type (#Standard; Presence-First) in Experiments 1 & 2

		Expt 1			Expt 2		
Filler Identifications	Fixed Effect	*b*	SE_b	95% CI_b	*b*	SE_b	95% CI_b
Included	Intercept	-0.94			-0.55		
	Lineup	-0.19	0.18	-0.54, 0.16	-0.71*	0.19	-1.07, -0.34
	TP	1.72 *	0.18	1.36, 2.07	1.69*	0.21	1.29, 2.10
	Lineup × TP	0.33	0.25	-0.17, 0.83	0.45*	0.20	0.07, 0.84
Excluded	Intercept	-2.13			-1.91		
	Lineup	-0.41	0.26	-0.91, 0.09	-0.66*	0.24	-1.11, -0.19
	TP	2.73*	0.20	2.33, 3.12	2.81*	0.23	2.34, 3.27
	Lineup × TP	0.49	0.31	-0.11, 1.09	0.48*	0.25	0.00, 0.96

Note: * indicates significant predictors, # indicates the reference lineup condition. Number of observations for Expt 1: Standard vs PF = 2600 (2063), *n* = 65 (61). For Expt 2: Standard vs PF = 2560 (2143), *n* = 64 (64). Expt 1 Models that included the standard lineup procedure did not include random effects by stimulus due to a computer error during the experiment.

Table 7

Fixed Effect Coefficients for Mixed-Effects Model Predicting Decision Type (Lineup Rejection; Positive Identification) by Lineup Type (#Standard; Best-Match-First) in Experiments 1 & 2

		Expt 1			Expt 2		
Filler Identifications	Fixed Effect	*b*	SE_b	95% CI_b	*b*	SE_b	95% CI_b
Included	Intercept	-0.98			-0.61		
	Lineup	0.07	0.26	-0.44, 0.57	0.49	0.27	-0.04, 1.01
	TP	1.80 *	0.22	1.37, 2.24	1.74*	0.23	1.28, 2.02
	Lineup × TP	0.28	0.33	-0.36, 0.92	-1.41*	0.27	-1.93, -0.88
Excluded	Intercept	-2.28			-1.97		
	Lineup	0.26	0.31	-0.34, 0.86	0.54 *	0.27	0.00, 1.07
	TP	2.92*	0.25	2.42, 3.41	2.89*	0.28	2.32, 3.44
	Lineup × TP	0.09	0.36	-0.62, 0.80	-1.56*	0.32	-2.20, -0.93

Note: * indicates significant predictors, # indicates the reference lineup condition. Number of observations for Expt 1: Standard vs BMF = 2440 (2054), *n* = 61 (61). For Expt 2: Standard vs BMF = 2600 (2028), *n* = 65 (65). Expt 1 Models that included the standard lineup procedure did not include random effects by stimulus due to a computer error during the experiment.

Table 8

Fixed Effect Coefficients for Mixed-Effects Model Predicting Decision Type (Lineup Rejection; Positive Identification) by Lineup Type (#Best-Match-First; Presence-First) in Experiments 1 & 2

		Expt 1			Expt 2		
Filler Identifications	Effect	*b*	SE_b	95% CI_b	*b*	SE_b	95% CI_b
Included	Intercept	-0.80			-0.11		
	Lineup	-0.36	0.28	-0.91, 0.19	-1.17*	0.21	-1.58, -0.76
	TP	2.20*	0.29	1.63, 2.77	0.33*	0.14	0.05, 0.60
	Lineup × TP	0.10	0.38	-0.63, 0.84	1.84*	0.22	1.41, 2.27
Excluded	Intercept	-2.21			-1.41		
	Lineup	-0.79*	0.34	-1.46, -0.11	-1.10*	0.24	-1.58, -0.62
	TP	3.25 *	0.34	2.59, 3.91	1.29*	0.20	0.90, 1.69
	Lineup × TP	0.56	0.45	-0.32, 1.45	1.94*	0.26	1.43, 2.45

Note: * indicates significant predictors, # indicates the reference lineup condition. Number of observations for Expt 1: BMF vs PF = 2480 (2066), *n* = 62 (61). For Expt 2: BMF vs PF = 2520 (2055), *n* = 63 (63).

Finally, it is also noteworthy that, in itself, splitting the standard simultaneous task into two structurally separate tasks in the modified procedures was not associated with better discriminability than in the standard task. Thus we found no evidence for Pozzulo and colleagues' (Pozzulo et al., 2008; Pozzulo & Lindsay, 1998) idea that separating the best-match from the offender-selection decision would encourage better decision-making in the second judgment.

Rationale for Experiment 2. The structure of the presence-first procedure in the mini-lineup paradigm in Experiment 1 meant that witnesses were making their two-step identification decisions for each lineup, then proceeding to the next lineup, until each block was completed. Making presence and best-match decisions (when witnesses had indicated the target was present) in immediate succession in this way might have cued best-match-first decision and memory processes across trials for both component decisions. That is, when presence-first witnesses alternated between best-match and presence decisions with every response, cues provided by the Present/Not Present task to forestall best-match-first processes would have been in direct competition with cues that encouraged familiarity-based best-match-first decision-making for the second, offender-selection decision, when made. These intermittent cues to best-match-first processes might have heightened the salience of familiarity evidence for positive match, useful for judging best-match candidates. Results suggest that cues to best-match-first decision-making might have prevented presence-first participants from making better use of recollective evidence, to some extent, when the two decisions were made in close succession for multiple lineups.[4] Overall, the combination of anticipating the need to pick an offender from each lineup that was given a positive presence decision, and having made best-match decisions for some lineups, might have encouraged over-reliance on familiarity in decision-making in general. As a result, identification

[4] Though, no significant effects of trial or block were observed.

performance was not significantly different from that observed for either the best-match-first or the standard lineup procedure (for which we predicted witnesses would take a best-match-first approach). However, results were in the predicted direction, suggesting a potential effect. Therefore, in Experiment 2 we separated the two decisions for each of the modified procedures into separate test blocks.

Experiment 2. Results from Experiment 2 tell a different story than those for Experiment 1, and support our predictions for the presence-first procedure. When we asked witnesses in the modified procedures to make their two identification decisions for the same lineups in separate blocks, memory performance in both of the modified procedures differed significantly from the standard lineup procedure.

When we compared identification accuracy between standard and presence-first lineups, significant differences in discriminability (the effect of the lineup × target presence interaction) and response bias (the effect of lineup) emerged. As predicted, the presence-first procedure demonstrated significantly greater discriminability than the standard procedure ($OR_{\text{fillers included}}$ = 1.57 [1.07, 2.34]; $OR_{\text{fillers excluded}}$ = 1.62 [1.00, 2.61]), but also lower response bias (i.e., a lower tendency to choose from the lineup). As Figure 6 shows, there was a significant difference between procedures in witnesses' tendency to make a positive identification, with those in the presence-first procedure less likely to make a positive identification than those in the standard procedure. However, the difference between lineup procedures in the likelihood of making a false identification decision for an innocent suspect was significantly greater than the difference in the likelihood of accurately identifying a guilty suspect from a target-present lineup. That is, presence-first witnesses' greater ability than those in the standard procedure to discriminate guilty from innocent suspects was largely due to a higher likelihood of avoiding false identifications from target-absent lineups, as was expected with greater use of recollection of mismatching detail. In sum, as predicted, when participants in the presence-first procedure did not know that they would later be asked to select a lineup member as the offender (second decision) after indicating whether or not the offender was present in the lineup (first decision), discriminability was significantly better than for a standard simultaneous procedure.

Best-match-first second decision performance. Best-match-first witnesses were led to a drastically lower presence performance than that of witnesses in the other procedures, by making their presence decisions separately, in a later block than their best-match decisions (see Figure 6). Pairwise comparisons involving the best-match-first procedure are reported here for completeness (see Tables 5, 6, & 8, and Figure 6 (Panels A, B, E, & F), but results are not interpretable in terms of framework predictions.

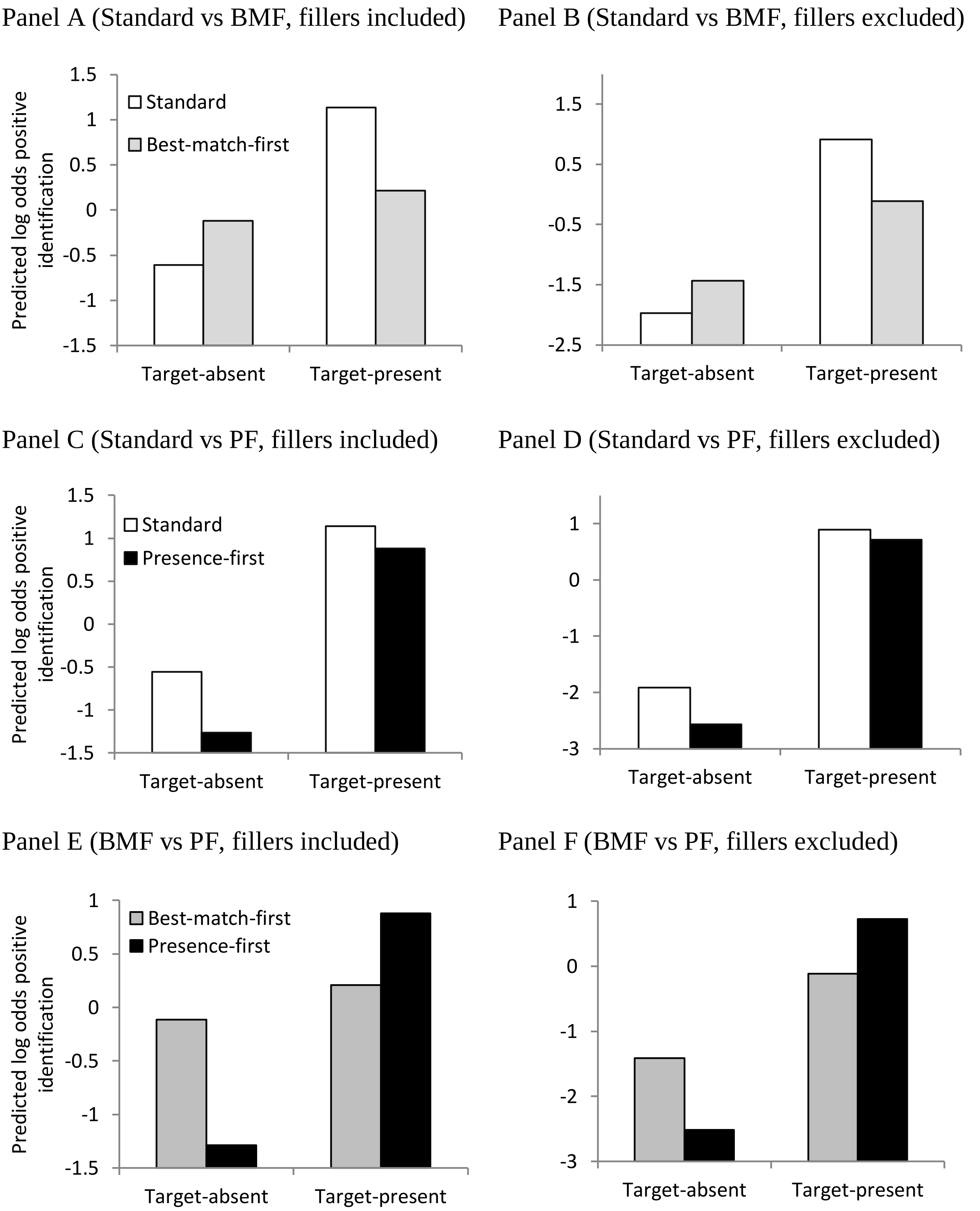


Figure 6. Plots of predicted log odds of a positive identification decision in Experiment 2 by lineup type and target presence. Values greater than 0 indicate a positive identification from the mini-lineup is more likely, and values lower than 0 indicate a lineup rejection is more likely. Coefficient standard errors, indicative of the variability of the estimates, are displayed in Tables 6-8.

The most likely explanation for the disparity in performance between the best-match-first procedure and the remaining two procedures was the impact of blocking the two decisions in the modified procedure on participants' ability to make the delayed second component decision. To evaluate this explanation for their poor discriminability, we examined the accuracy of best-match decisions made by best-match-first witnesses for target-present lineups in Experiments 1 and 2. The proportion correct for best-match selections that ended in positive identifications (i.e., the offender was selected as best-match) was similar across lineup procedures in Experiment 1 (Standard: *M* = .84 [.83, .85], *SD* = .13; Presence-first: *M* = .86 [.85, .87], *SD* = .15; Best-match-first: *M* = .86 [.85, .87], *SD* = .13), and Experiment 2 (Standard: *M* = .82 [.81, .83], *SD* = .17; Presence-first: *M* = .87 [.86. .88], *SD* = .13), except for the best-match-first condition in Experiment 2 (*M* = .73 [.72, .74], *SD* = .17). Further, the mean number of positive presence decisions made was higher for standard (Expt 1: *M* = 11.22 [11.05, 11.39], *SD* = 3.16; Expt 2: *M* = 14.64 [14.49, 14.79], *SD* = 2.87) and presence-first (Expt 1: *M* = 12.16 [11.96, 12.36], *SD* = 3.46; Expt 2: *M* = 13.87 [13.74, 14.00], *SD* = 2.35) procedures, than for the best-match-first procedure in Experiment 2 (Expt 1: *M* = 12.52 [12.28, 12.76], *SD* = 4.45; Expt 2: *M* = 10.97 [10.78, 11.16], *SD* = 3.45). However, when we examine the proportion correct for best-match selections from *all* target-present trials for best-match-first participants in both experiments, regardless of their second presence decision, the difference is negligible (Experiment 1: *M* = .74 [.73, .75], *SD* = .18; Experiment 2: *M* = .71 [.70, .72], *SD* = .16). Overall, the accuracy of best-match selections was similar for best-match-first participants in both experiments, but of targets selected as the best-match, considerably less were later identified as the target in Experiment 2, than Experiment 1. These results suggest that presenting the two decision steps in separate blocks made it more difficult for best-match-first witnesses to make accurate presence decisions about their best-match candidates.

One explanation for this is that when best-match-candidates were presented later, and outside the context of the full lineup, for second decisions in Experiment 2, witnesses could not recall the corresponding lineup from which they had been selected. Or, even if witnesses remembered their best-match-candidate, they may have been less able to accurately discriminate its source (e.g., Johnson, Hashtroudi, & Lindsay, 1993; Memon, Hope, Bartlett, & Bull, 2002). That is, whether the face had been studied as a target, or seen as a filler during the earlier best-match decision test blocks. As a result, presence decision accuracy was at near-chance levels. The absence of a similar drop in identification performance for the presence-first procedure in Experiment 2 supports this explanation. Because presence-first participants were presented again with the whole lineup from which to make their offender selection decision, they did not need to recall the lineup. Therefore, the accuracy of their second decisions (offender-selection) about targets in target-present lineups did not suffer from their two-step decisions being blocked.

Summary. In Experiment 2, we aimed to test whether separating the first presence decision from the later offender-selection decision would allow presence-first witnesses to make better use of recollection than in Experiment 1. Blocking the two recognition decisions witnesses made in the modified procedures removed competing structural cues to best-match-first decision-making created by alternating first and second decisions in the previous experiment. Results are consistent with presence-first participants having been successfully diverted from best-match-first decision-making, and an over-reliance on positive familiarity. That is, presence-first witnesses were able to focus less on the idea of picking out a specific offender, if present, and more on weighing up the evidence both for *and against* the offender being present in the lineup. Consequently, as predicted, they made better use of evidence from memory, consistent with better use of recollective evidence, leading to more accurate final identification decisions. If, as we proposed, evidence of mismatch with memory for the offender was more likely to be neglected in best-match-first decision-making (e.g., insufficient use of recall-to-reject, Rotello & Heit, 2000; Reyna & Brainerd, 1995, and the distinctiveness heuristic, Gallo, Bell, Beier, & Schacter, 2006), then making witnesses less aware that an offender might need to be picked was likely to improve their consideration of evidence against positive identification. Importantly, though, the pattern of results across both experiments suggests that the presence-first procedure did not as effectively cue witnesses to make better use of recollective evidence when other factors impinged on the framing of the identification task. The effect we observed of the presence-first procedure on performance only reached significance when witnesses 1) were not cued to anticipate making a full identification decision; and 2) did not experience pointing a lineup member out as the offender until after presence decisions were made.

Table 9

Descriptive Statistics for Identification Responses (Hit, Miss, Filler ID, Correct Rejection, and False ID Rates) by Lineup Type for Experiments 1 & 2

				Response Proportion							
				Target-present				Target-absent			
Expt	Lineup Type	Statistic	*n*	Hits	Filler IDs (TP)	Misses	Number of observations	Correct Rejections	False IDs	Filler IDs (TA)	Number of observations
1	Standard	*M*	32	.56	.11	.33	640	.71	.09	.20	640
		SD		.16	.10	.17		.14	.08	.12	
		95% CI		.55, .57	.10, .12	.31, .34		.70, .72	.08, .10	.19, .21	
	BMF	*M*	29	0.63	0.11	0.27	580	0.68	0.11	0.21	580
		SD		0.22	0.11	0.19		0.24	0.09	0.18	
		95% CI		.61, .64	.10, .12	.25, .28		.67, .70	.10, .11	.20, .22	
	PF	*M*	33	0.61	0.10	0.30	660	0.74	0.07	0.19	660
		SD		0.17	0.09	0.13		0.14	0.05	0.12	
		95% CI		.6, .62	.09, .10	.29, .31		.73, .75	.06, .07	.18, .20	

2	Standard	*M*	33	.60	.13	.27	660	.62	.12	.27	660
		SD		.17	.13	.14		.21	.08	.17	
		95% CI		.59, .61	.12, .14	.26, .28		.60, .63	.11, .13	.25, .28	
	BMF	*M*	32	.40	.15	.45	640	.53	.14	.33	640
		SD		.13	.12	.17		.19	.08	.17	
		95% CI		.39, .41	.14, .16	.44, .46		.51, .54	.14, .15	.32, .34	
	PF	*M*	31	.61	.09	.31	620	.76	.08	.16	620
		SD		.16	.08	.12		.12	.07	.10	
		95% CI		.59, .62	.08, .09	.30, .32		.75, .77	.08, .09	.16, .17	

General Discussion

In this paper, we tested 1) a best-match-first dual-process framework to explain memory and decision processes in standard simultaneous lineup procedures; and 2) a novel presence-first simultaneous lineup procedure informed by the framework, designed to improve the use of recollection by eyewitnesses. We investigated three key predictions of the framework. First, that eyewitnesses first select a best-match candidate from a simultaneous lineup, before deciding whether or not the best-match candidate is the offender. Second, that best-match-first decision-making would encourage an over-reliance on familiarity, and neglect of the importance of recollected detail related to the offender. Finally, that reversing the order of component recognition decisions that contribute to an eyewitness' identification response, would pre-empt best-match-first processes and allow eyewitnesses to make better use of recollection. In two experiments, we modified the standard simultaneous lineup procedure such that mock eyewitnesses using the modified procedures made the two component decisions separately in one of two possible orders: best-match-first; or presence-first. Results from Experiment 1 provide the first direct evidence for the best-match-first time-course of component recognition decisions made for simultaneous lineups. In Experiment 2, when the first and second component decisions in the modified procedures were made in separate blocks, we found greater discriminability in the presence-first than a standard simultaneous procedure. Results thus support framework predictions about the use of familiarity and recollection for patterns of eyewitness identification performance in these tasks. We discuss theoretical and applied implications of these key findings in turn.

Best-match-first decision-making. We examined patterns of first-decision response latency to evaluate the likely order of component decisions made for simultaneous lineups. Consistent with predictions, we found evidence that eyewitnesses were likely to first choose a best-match candidate, if at least one lineup member was a good-enough match to memory for the offender, before deciding on offender presence. Our first prediction was that in taking a best-match-first approach to making identification decisions for a simultaneous lineup, eyewitnesses would demonstrate slower first-decision response latency for positive identifications from standard than best-match-first lineups as first-decision best-match responses would be available earlier in the time-course of an identification decision. As predicted, first decisions in Experiment 1[5] were made earlier by eyewitnesses in the best-match-first than the standard procedure for positive identification decisions. This is consistent with the idea that best-match responses required one decision (the best-match selection) while standard decisions required two decisions (best-match + offender-presence). We predicted that this difference would be smaller (or potentially reversed) for lineup rejections, as best-match selections for lineups that would otherwise be rejected outright (in a standard procedure) would require an additional decision (i.e., a guess or forced best-match-selection). This prediction was supported when tested in Experiment 1. For lineup rejections, first-decision response latency was slower for best-match-first than standard lineups. This finding is consistent with a proportion of rejections having been made for lineups that did not include a plausible best-match candidate. Rejections for these lineups would have required two steps in the best-match-first procedure (outright lineup rejection + forced best-match selection) while standard rejections required only one step (outright lineup rejection).

These results provide the first direct evidence for the best-match-first order of simultaneous lineup decision processes first proposed in Wells' (1984) relative judgment idea, and taken as the starting point for the development of our analytical framework. While

[5] Decision type was not examined separately for this comparison in Experiment 2, due to poor second-decision performance, but patterns across decision type were consistent with this effect.

often assumed in explanations for simultaneous lineup decision-making performance, and instantiated in Clark's (2003) WITNESS model of identification decisions, little past research has been directed at testing this conceptualisation of component identification decisions, or their likely order.

An interesting finding emerged from comparison of first-decision response latency for the presence-first procedure with the best-match-first procedure. Positive presence decisions by presence-first witnesses were not made slower than best-match judgments for lineups later deemed to be target-present by best-match-first witnesses in Experiment 1. That is, there was no evidence that presence-first participants took a best-match-first approach and made two ordered judgments (best-match + presence) while best-match-first participants made only one (best-match). Most importantly, this finding demonstrated that there was no unique and inevitable approach to making recognition decisions for these simultaneous lineup tasks. That is, eyewitness decision processes for the same set of simultaneously presented stimuli depended on the task context, and what questions were asked.

Further research into how eyewitnesses evaluate lineup stimuli to make these component decisions is needed to specify decision-making processes in simultaneous lineup tasks in more detail. A formal model of decision-making for an array of multiple test stimuli would potentially inform the development of more effective identification tasks by 1) identifying finer-grained decision processes that contribute to identification errors; and 2) informing structural or instructional modifications capable of ameliorating error-prone processes. Moreover, results from these experiments have implications for structural task analysis of other identification lineup formats, most obviously variants on the sequential procedure, and Brewer and colleagues' (Brewer et al., 2012; Sauer et al., 2008) novel confidence rating procedure. Both of these formats involve a disruption of best-match-first decision-making, but further research is needed to evaluate any structural or procedural effects on witnesses' use of familiarity and recollection.

Recollection and the novel presence-first procedure. Comparisons of discriminability between standard and modified lineup procedures in Experiment 2 were also consistent with framework predictions of over-reliance on familiarity with best-match-first decision-making, and better use of recollection with the reversed order of component decisions in the novel presence-first procedure. Eyewitnesses were better at differentiating between innocent and guilty suspects with the presence-first procedure than either the best-match-first or standard procedures, consistent with more appropriate weighting of familiarity and recollection. This greater discriminability was driven by a higher likelihood of avoiding false identifications. In this, presence-first participants demonstrated a pattern typically observed when more use is made of recollection to counter the misleading familiarity of highly similar foils. Investigating ways to reduce eyewitnesses' tendency to make false identification decisions when the offender is not in the lineup, without unduly reducing witness confidence in making accurate positive identifications of offenders, has been a major focus of research into eyewitness identification procedures. Therefore, this finding is important for future identification research. Also consistent with the framework, separating the simultaneous task into its two component decisions, without redressing the best-match-first sequence of decisions, was not sufficient to produce this effect. However, while the pattern of effects was in the predicted direction in Experiment 1, discriminability did not differ significantly between lineup procedures. Therefore, the effect of the presence-first procedure was evident most clearly only in Experiment 2, when eyewitnesses were not aware they would later be asked to make a second offender-selection decision for lineups recognised as target-present.

The greater efficacy of the presence-first procedure than the standard simultaneous procedure in Experiment 2 has important implications for the development of identification procedures that can better diagnose suspect innocence or guilt than those currently used by police. When eyewitnesses have encoded and retained recollectible detail related to the offender, it is important that this detail is used to avoid false identifications of innocent suspects who closely resemble the offender, as well as to increase the quality of evidence informing accurate positive identifications. Moreover, when the amount of recollectible detail encoded or able to be retrieved is limited, it is important that eyewitnesses weight familiarity-based information from memory appropriately in the absence of recollection (cf. Sacher, Taconnat, Souchay, & Isingrini, 2009; Souchay, Moulin, Clarys, Taconnat, & Isingrini, 2007). Therefore, if the presence-first procedure encouraged use of the distinctiveness heuristic (e.g., Gallo et al., 2006), or memory editing mechanisms other than recall-to-reject (Rotello & Heit, 2000), identification performance is likely to be enhanced, even in the absence of substantial recollection. The presence-first procedure also retains the potential benefit for maximal retrieval of memory for the offender of presenting several cues (lineup members) at the same time (e.g., Humphreys et al., 1989). Thus, the presence-first procedure holds the promise of improved use of available familiarity and recollection, even when memory quality is low and retrieved memorial evidence is predominantly familiarity-based. However, resolution of the question of whether simultaneous or sequential presentation is of greater benefit for memory retrieval, and under what conditions, is elusive. Therefore, further investigation of the use of familiarity and recollection in sequential formats, and whether it can be improved, also has immediate practical import.

Nevertheless, if the difference in patterns of significance observed for Experiments 1 and 2 does reflect a genuine difference in discriminability patterns, this is clearly a problem for the use of the modified presence-first procedure by police. All participants were novices who had not previously participated in any eyewitness identification-type experimental tasks. We could be confident that our mock witnesses were free of prior experience-based learning or expectations about identification decision-making when making recognition decisions about test lineups. Therefore, one possible explanation for the observed difference in identification performance for the presence-first procedures in Experiments 1 and 2 is that alternating best-match decisions between offender-presence decisions within a block in Experiment 1 cued best-match-first decision-making processes across both decisions for all trials. As a result, Experiment 1 presence-first participants were not as effectively deterred from finding and focusing on a single best-match candidate, and did not make as much use of recollection as was observed in Experiment 2. In practice, eyewitnesses would generally be asked to make only a single identification decision for a single lineup. Therefore, alternating decision types would not be an issue. However, multiple factors could potentially undermine the efficacy of the presence-first procedure in diverting eyewitnesses from best-match-first processes. Therefore, its usefulness in the mini-lineup lab task may not be directly transferrable to genuine criminal investigations.

A second possible explanation for the observed difference in presence-first efficacy, with more serious implications for the applied utility of the novel procedure, is that mock witnesses could only make better use of recollection when they were not aware that they would need to make the second offender-selection decision for lineups identified as including the offender. While many eyewitnesses would not have previous first-hand experience of an identification lineup, most people would bring to the task a clear expectation of being asked to point out a single person, if recognised as the offender. Expectations have been shown to affect people's retrieval and evaluation of memorial evidence (e.g., Dodson & Schacter, 2002; Johnson, Hashtroudi, & Lindsay, 1993; Odegard & Lampinen, 2006; Lampinen, Odegard, & Bullington, 2003). If common expectations prime eyewitnesses to adopt a best-match-first approach to evaluating lineup members, regardless of the actual recognition questions they are asked (e.g., for an offender-presence first decision), no benefit in the reliability of identification evidence will be gained from the modified procedure. Similar effects on decision-making might be anticipated to arise from witness perceptions of the urgency of assisting police investigations by identifying a guilty suspect, overt or implicit administrator bias; or other witness-specific factors (e.g., Clark, 2005; Malpass & Devine, 1981).

Given these limitations, further research into the efficacy of the presence-first procedure itself should take two separate tacks. First, to further investigate the procedure’s effectiveness across a variety of test conditions when eyewitnesses know they may need to select a single lineup member as the offender. And, second, to investigate alternative methods, whether structural or other, to cue or encourage eyewitnesses to make better use of recollection. For example, providing participants with pre-test information or warnings about the nature of potential memory errors (e.g., Dunning & Stern, 1994; McCabe & Smith, 2002), or markers of memory accuracy (e.g., Lane et al., 2007; Neuschatz et al., 2001), has been demonstrated to improve use of recollection across a range of recognition and recall tasks in which the familiarity of unstudied stimuli is likely to be misleading (e.g. Gallo, Roberts, & Seamon, 1997; Lane et al., 2007; McDermott & Roediger, 1998; Schooler, Gerhard, & Loftus, 1986; Starns et al., 2007).

Conclusions. In theoretical terms, our results demonstrate that our framework provides a useful dual-process explanation for eyewitnesses’ use of memorial evidence in simultaneous lineups that encourage best-match-first decision-making. In this, these experiments provide further impetus for the extension of theoretical memory models developed to explain performance in simple laboratory tasks, to real-world tasks that entail far more complex demands on recognition memory (cf. Brewer et al., 2012; Brewer & Wells, 2011; Gronlund et al., 2015). Moreover, we found evidence that the modified presence-first procedure informed by key understandings implemented in the framework, led to more effective use of memorial evidence by eyewitnesses making simultaneous lineup decisions. In this way, the reliability of their identification responses was enhanced. Our framework thus provides a basis for more formal process modelling of extant and novel identification tasks. Greater use of established knowledge from basic memory research in models (e.g., Ratcliff & Starns, 2009; Starns, Ratcliff, & McKoon, 2012; Voskuilen & Ratcliff, 2016) that might explain complex patterns of eyewitness choosing, accuracy, response latency, and other recognition responses and ratings in this way is vital. Both innovative test design that can better diagnose suspect guilt or innocence, and a stronger evidence-base for concrete practical recommendations, rely on theory-driven research to advance the just use of identification evidence at all levels of the criminal justice system.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390-412. doi: 10.1016/j.jml.2007.12.005

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structures for confirmatory hypothesis testing: Keeping it maximal. *Journal of Memory and Language*, 68, 255-278. doi: 10.1016/j.jml.2012.11.001

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2014). _lme4: Linear mixed-effects models using Eigen and S4_. Retrieved from http://CRAN.R-project.org/package=lme4

Bodner, G. E., & Lindsay, D. S. (2003). Remembering and knowing in context. *Journal of Memory and Language*, *48*, 563-580. doi: 10.1016/s0749-596x(02)00502-8

Brainerd, C. J., Reyna, V. F., Wright, R., & Mojardin, A. H. (2003). Recollection rejection: False-memory editing in children and adults. *Psychological Review*, *110*, 762-784. doi: 10.1037/0033-295x.110.4.762

Brigham, J. C., Meissner, C. A., & Wasserman, A. W. (1999). Applied issues in the construction and expert assessment of photo lineups. *Applied Cognitive Psychology*, *13*, S73-S92. doi: 10.1002/(sici)1099-0720(199911)13:1+%3Cs73::aid-acp631%3E3.3.co;2-w

Brewer, N., Caon, A., Todd, C., & Weber, N. (2006). Eyewitness identification accuracy and response latency. *Law and Human Behavior*, *30*, 31-50. doi: 10.1007/s10979-006-9002-7. doi: 10.1007/s10979-006-9002-7

Brewer, N., Gordon, M., & Bond, N. (2000). Effect of photoarray exposure duration on eyewitness identification accuracy and processing strategy. *Psychology, Crime and Law*, *6*, 21-32. doi: 10.1080/10683160008410829

Brewer, N., & Palmer, M. A. (2010). Eyewitness identification tests. *Legal and Criminological Psychology*, *15*, 77-96. doi: 10.1348/135532509x414765

Brewer, N., & Weber, N. (2008). Eyewitness confidence and latency: Indices of memory processes not just markers of accuracy. *Applied Cognitive Psychology*, *22*, 827-840. doi: 10.1002/acp.1486

Brewer, N., Weber, N., Wootton, D., & Lindsay, D. S. (2012). Identifying the bad guy in a lineup using confidence judgments under deadline pressure. *Psychological Science*, *23*, 1208-1214. doi: 10.1177/0956797612441217

Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied*, *12*, 11-30. doi: 10.1037/1076-898X.12.1.11

Brewer, N., & Wells, G. L. (2011). Eyewitness identification. *Current Directions in Psychological Science*, *20*, 24-27. doi: 10.1177/0963721410389169

Carlson, C. A., & Gronlund, S. D. (2011). Searching for the sequential line-up advantage: A distinctiveness explanation. *Memory*, *19*(8), 916-929. doi: 10.1080/09658211.2011.613846

Carlson, C. A., Gronlund, S. D., & Clark, S. E. (2008). Lineup composition, suspect position, and the sequential lineup advantage. *Journal of Experimental Psychology: Applied*, *14*, 118-128. doi: 10.1037/1076-898x.14.2.118

Charman, S., & Wells, G. L. (2014). Applied Lineup Theory. *Handbook Of Eyewitness Psychology 2 Volume Set*, 219-254. doi: 10.4324/9781315805535

Clark, S. E. (2003). A memory and decision model for eyewitness identification. *Applied Cognitive Psychology*, *17*, 629-654. doi: 10.1002/acp.891

Clark, S. E. (2005). A re-examination of the effects of biased lineup instructions in eyewitness identification. *Law and Human Behavior*, *29*, 575-604. doi: 10.1007/s10979-005-7121-1

Clark, S. E. (2008). The importance (necessity) of computational modelling for eyewitness identification research. *Applied Cognitive Psychology*, *22*, 803-813. doi: 10.1002/acp.1484

Clark, S. E., & Davey, S. L. (2005). The target-to-foils shift in simultaneous and sequential lineups. *Law and Human Behavior, 29*, 151-172. doi: 10.1007/s10979-005-2418-7

Clark, S. E., Erickson, M. A., & Breneman, J. (2011). Probative value of absolute and relative judgments in eyewitness identification. *Law and Human Behavior*, *35*, 364-380. doi: 10.1007/s10979-010-9245-1

Clark, S. E., & Gronlund, S. D. (2015). The development and evolution of theory in eyewitness identification. In J. G. W., Raajmakers, A. H. Criss, R. L. Goldstone, R, M. Nosofsky, & M. Steyvers (Eds.) *Cognitive Modeling in Perception and Memory: A Festschrift for Richard M. Shiffrin*, Psychology Press: New York. doi: 10.4324/9781315885582

Clark, S. E., Howell, R. T., & Davey, S. L. (2008). Regularities in eyewitness identification. *Law and Human Behavior*, *32*, 187-218. doi: 10.1007/s10979-006-9082-4

Cook, G. I., Marsh, R. L., & Hicks, J. L. (2005). Revisiting the role of recollection in item versus forced-choice recognition memory. *Psychonomic Bulletin & Review*, *12*, 720-725. doi: 10.3758/bf03196763

De Carlo, L. T. (1998). Signal detection theory and generalized linear models. *Psychological Methods*, *3*, 186-205. doi: 10.1037//1082-989x.3.2.186

Deffenbacher, K. A., Bornstein, B. H., & Penrod, S. D. (2006). Mugshot exposure effects: Retroactive interference, mugshot commitment, source confusion, and unconscious transference. *Law and Human Behavior, 30,* 287-307. doi: 10.1007/s10979-006-9008-1

Dodson, C. S., & Dobolyi, D. G. (2016). Confidence and Eyewitness Identifications: The Cross-Race Effect, Decision Time and Accuracy. *Applied Cognitive Psychology*, *30*, 113-125. doi: 10.1002/acp.3178

Dodson, C. S., & Schacter, D. L. (2002). When false recognition meets metacognition: The distinctiveness heuristic. *Journal of Memory and Language*, *46*, 782-803. doi: 10.1037/e501882009-058

Dunning, D., & Perretta, S. (2002). Automaticity and eyewitness accuracy: A 10-to 12-second rule for distinguishing accurate from inaccurate positive identifications. *Journal of Applied Psychology*, *87*, 951-962. doi: 10.1037//0021-9010.87.5.951

Dunning, D., & Stern, L. B. (1994). Distinguishing accurate from inaccurate eyewitness identifications via inquiries about decision processes. *Journal of Personality and Social Psychology*, *67*, 818-835. doi: 10.1037//0022-3514.67.5.818

Evans, J. R., Marcon, J. L., & Meissner, C. A. (2009). Cross-racial lineup identification: Assessing the potential benefits of context reinstatement. *Psychology, Crime & Law*, *15*, 19-28. doi: 10.1080/10683160802047030

Gallo, D. A., Bell, D. M., Beier, J. S., & Schacter, D. L. (2006). Two types of recollection-based monitoring in younger and older adults: Recall-to-reject and the distinctiveness heuristic. *Memory*, *14*, 730-741. doi: 10.1080/09658210600648506

Gallo, D. A., Roberts, M. J., & Seamon, J. G. (1997). Remembering words not presented in lists: Can we avoid creating false memories? *Psychonomic Bulletin & Review*, *4*, 271-276. doi: 10.3758/bf03209405

Gardiner, J. M., & Richardson-Klavehn, A. (2000). Remembering and knowing. In E. Tulving & F. I. M. Craik (Eds.), The Oxford handbook of memory (pp. 229-244). London: Oxford University Press. doi: 10.1002/acp.822.

Gelman, A., & Hill, J. (2007). *Data Analysis using Regression and Multilevel/Hierarchical Models.* New York: Cambridge University Press.

Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness*, *5*(2), 189-211. doi: 10.1080/19345747.2011.618213

Goodsell, C. A., Gronlund, S. D., & Carlson, C. A. (2010). Exploring the sequential lineup advantage using WITNESS. *Law and Human Behavior*, *34*, 445-459. doi: 10.1007/s10979-009-9215-7

Goodsell, C. A., Neuschatz, J. S., & Gronlund, S. D. (2009). Effects of mugshot commitment on lineup performance in young and older adults. *Applied Cognitive Psychology*, *23*, 788-803. doi: 10.1002/acp.1512

Gronlund, S. D. (2005). Sequential lineup advantage: Contributions of distinctiveness and recollection. *Applied Cognitive Psychology,* 19, 23-37. doi: 10.1002/acp.1047

Gronlund, S. D., Mickes, L., Wixted, J. T., & Clark, S. E. (2015). Conducting an eyewitness lineup: How the research got it wrong. In B. H. Ross (Ed.), The psychology of learning and motivation (Volume 63, pp. 1– 43). Waltham, MA: Academic Press. doi: 10.1016/bs.plm.2015.03003

Guerin, S. A., Robbins, C. A., Gilmore, A. W., & Schacter, D. L. (2012). Retrieval failure contributes to gist-based false recognition. *Journal of Memory and Language*, *66*, 68-78. doi: 10.1016/j.jml.2011.07.002

Hanczakowski, M., Zawadzka, K., & Coote, L. (2014). Context reinstatement in recognition: memory and beyond. *Journal of Memory and Language*, *72*, 85-97. doi: 10.1016/j.jml.2014.01.001

Hintzmann, D. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, 95, 528–551. doi: 10.1037//0033-295x.95.4.528

Hintzmann, D.L. & Curran, T. (1994) Retrieval dynamics of recognition and frequency judgements: Evidence for separate processes of familiarity and recall. *Journal of Memory and Language*, 33, 1-18. doi: 10.1006/jmla.1994.1001

Horry, R., Memon, A., Wright, D. B., & Milne, R. (2011). Predictors of eyewitness identification decisions from video lineups in England: A field study. *Law and Human Behavior, 36,* 257-265. doi: 1007/s10979-011-9279-z

Humphreys, M. S., Bain, J. D., & Pike, R. (1989). Different ways to cue a coherent memory system: A theory for episodic, semantic, and procedural tasks. *Psychological Review*, *96*, 208-233. doi: 10.1037/0033-295x.96.2.208

Ingram, K. M., Mickes, L., & Wixted, J. T. (2012). Recollection can be weak and familiarity can be strong. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38,* 325-339. doi: 10.1037/e520592012-514

Innocence Project (2015). Innocence project retrieved October 31, 2015 http://www.innocenceproject: .

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3-28. doi: 10.1037//0033-2909.114.1.3

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology, 103*(1), 54-69. doi: 10.1037/a0028347

Lampinen, J. M., Odegard, T. N., & Bullington, J. L. (2003). Qualities of memories for performed and imagined actions. *Applied Cognitive Psychology, 17*, 881-893. doi: 10.1002/acp.916

Lane, S. M., & Meissner, C. A. (2008). A 'middle road' approach to bridging the basic–applied divide in eyewitness identification research. *Applied Cognitive Psychology*, *22*, 779-787. doi: 10.1002/acp.1482

Lane, S. M., Roussel, C. C., Villa, D., & Morita, S. K. (2007). Features and feedback: Enhancing metamnemonic knowledge at retrieval reduces source-monitoring errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 1131-1142. doi: 10.1037/0278-7393.33.6.1131

Lane, S. M., Roussel, C. C., Starns, J. J., Villa, D., & Alonzo, J. D. (2008). Providing information about diagnostic features at retrieval reduces false recognition. *Memory*, *16*, 836–851. http://doi: 10.1080/09658210802337341

Lindsay, R. C., & Wells, G. L. (1980). What price justice? Exploring the relationship of lineup fairness to identification accuracy. *Law and Human Behavior*, *4*, 303-313. doi: 10.1007/bf01040622

Lindsay, R. C., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology*, *70*, 556. doi: 10.1037/0021-9010.70.3.556

Luus, C. A., & Wells, G. L. (1991). Eyewitness identification and the selection of distracters for lineups. *Law and Human Behavior, 15*, 43. doi: 10.1007/bf01044829

Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology*, *57*, 335-384. doi: 10.1016/j.cogpsych.2008.02.004

Malmberg, K. J., & Xu, J. (2007). On the flexibility and the fallibility of associative memory. *Memory & Cognition*, *35*, 545-556. doi: 10.3758/bf03193293

Malpass, R. S. (2006). A policy evaluation of simultaneous and sequential lineups. *Psychology, Public Policy, and Law, 12*, 394-418. doi: 10.1037/1076-8971.12.4.394

Malpass, R. S., & Devine, P. G. (1981). Eyewitness identification: Lineup instructions and the absence of the offender. *Journal of Applied Psychology*, *66*, 482. doi: 10.1037/0021-9010.66.4.482

Marcon, J. L., Susa, K. J., & Meissner, C. A. (2009). Assessing the influence of recollection and familiarity in memory for own-versus other-race faces. *Psychonomic Bulletin & Review*, *16*, 99-103. doi: 10.3758/pbr.16.1.99

McDermott, K. B., & Roediger, H. L. (1998). Attempting to avoid illusory memories: Robust false recognition of associates persists under conditions of explicit warnings and immediate testing. *Journal of Memory and Language*, *39*, 508-520. doi: 10.1006/jmla.1998.2582

Meissner, C. A., Tredoux, C. G., Parker, J. F., MacLin, O. H. (2005). Eyewitness decisions in simultaneous and sequential lineups: A dual-process signal detection theory analysis. *Memory & Cognition,* 33, 783-792. doi: 10.3758/bf03193074

Memon, A., Hope, L., Bartlett, J., & Bull, R. (2002). Eyewitness recognition errors: The effects of mugshot viewing and choosing in young and old adults. *Memory & cognition, 30*, 1219-1227. doi: doi.10.3758/bf03213404

Mickes, L., Wais, P. E., & Wixted, J. T. (2009). Recollection is a continuous process: Implications for dual-process theories of recognition memory. *Psychological Science, 20,* 509-515. doi: 10.1111/j.1467-9280.2009.02324.x

Mickes, L. (2015). Receiver operating characteristic analysis and confidence-accuracy characteristic analysis in investigations of system variables and estimator variables that affect eyewitness memory. *Journal of Applied Research in Memory and Cognition*, *4,* 83-102. doi: 10.1016/j.jarmac.2015.01.003

Migo, E., Montaldi, D., Norman, K. A., Quamme, J., & Mayes, A. (2009). The contribution of familiarity to recognition memory is a function of test format when using similar foils. *The Quarterly Journal of Experimental Psychology*, *62,* 1198-1215. doi: 10.1080/17470210802391599

Murayama, K., Sakaki, M., Yan, V. X., & Smith, G. M. (2014). Type I error inflation in the traditional by-participant analysis to metamemory accuracy: A generalized mixed-effects model perspective. *Journal of Experimental Psychology: Learning Memory and Cognition, 40*, 1287-1306. doi: 10.1037/a0036914

National Research Council. (2014). I*dentifying the Culprit: Assessing Eyewitness Identification.* Washington, DC: The National Academies Press. doi: 10.17226/18891

Neuschatz, J. S., Benoit, G. E., & Payne, D. G. (2003). Effective warnings in the Deese-Roediger-McDermott false-memory paradigm: the role of identifiability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29,* 35. doi: 10.1037/0278-7393.29.1.35

Neuschatz, J. S., Payne, D. G., Lampinen, J. M., & Toglia, M. P. (2001). Assessing the effectiveness of warnings and the phenomenological characteristics of false memories. *Memory*, *9*, 53-71. doi: 10.1037/0278-7393.29.1.35

Odegard, T. N., & Lampinen, T. N. (2006). Memory editing: Knowledge, criteria, and alignment. *Memory*, *14*, 777-787. doi: 10.1080/09658210600648589

Palmer, M. A., & Brewer, N. (2012). Sequential lineup presentation promotes less-biased criterion setting but does not improve discriminability. *Law and Human Behavior*, *36*, 247-255. doi: 10.1037/h0093923

Palmer, M. A., Brewer, N., McKinnon, A. C., & Weber, N. (2010). Phenomenological reports diagnose accuracy of eyewitness identification decisions. *Acta Psychologica*, *133*, 137-145. doi: 10.1016/j.actpsy.2009.11.002

Perfect, T. J., & Weber, N. (2012). How should witnesses regulate the accuracy of their identification decisions: One step forward, two steps back? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1810-1818. doi: 10.1037/a0028461

Pozzulo, J. D., & Balfour, J. (2006). Children's and adults' eyewitness identification accuracy when a culprit changes his appearance: Comparing simultaneous and elimination lineup procedures. *Legal and Criminological Psychology*, *11*, 25-34. doi: 10.1348/135532505x52626

Pozzulo, J. D., Crescini, C., & Lemieux, J. M. (2008). Are accurate witnesses more likely to make absolute judgments? *International Journal of Law and Psychiatry*, *31*, 495-501. doi: 10.1016/j.ijlp.2008.09.006

Pozzulo, J. D., Dempsey, J., Corey, S., Girardi, A., Lawandi, A., & Aston, C. (2008). Can a lineup procedure designed for child witnesses work for adults? Comparing simultaneous, sequential, and elimination lineup procedures. *Journal of Applied Social Psychology, 38*, 2195-2209. doi: 10.1111/j.1559-1816.2008.00387.x

Pozzulo, J. D., & Lindsay, R. C. L. (1999). Elimination lineups: An improved identification procedure for child eyewitnesses. *Journal of Applied Psychology*, *84*, 167-176. doi: 10.1037/0021-9010.84.2.167

Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modelling with crossed random effects and with binomial data. *Journal of Memory and Language, 59, 413-425.* doi: 10.1016/j.jml.2008.02.002

R Core Team. (2014). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological review*, *85*, 59-108. doi: 10.1037/0033-295x.85.2.59

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*, 347-356. doi: 10.1111/1467-9280.00067

Ratcliff, R., & Starns, J. J. (2009). Modeling confidence and response time in recognition memory. *Psychological Review*, *116*, 59. doi: 10.1037/a0014086

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and individual Differences, 7*(1), 1-75. doi: 10.1016/1041-6080(95)90031-4

Rotello, C. M., & Heit, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory & Cognition*, *28*, 907-922. doi: 10.3758/bf03209339

Rotello, C. M., Macmillan, N. A., & Van Tassel, G. (2000). Recall-to-reject in recognition: Evidence from ROC curves. *Journal of Memory and Language, 43,* 67-88. doi: 10.1006/jmla.1999.2701

Rotello, C. M., & Zeng, M. (2008). Analysis of RT distributions in the remember—know paradigm. *Psychonomic Bulletin & Review*, *15*, 825-832. doi: 10.3758/pbr.15.4.825

Sacher, M., Taconnat, L., Souchay, C., & Isingrini, M. (2009). Divided attention at encoding: Effect on feeling-of-knowing. *Consciousness and Cognition*, *18*, 754-761. doi: 10.1016/j.concog.2009.04.001

Sauer, J. D., Brewer, N., & Weber, N. (2008). Multiple confidence estimates as indices of eyewitness memory. *Journal of Experimental Psychology: General*, 137, 528-547. doi: 10.1037/a0012712

Sauerland, M., & Sporer, S. L. (2009). Fast and confident: Postdicting eyewitness identification accuracy in a field study. *Journal of Experimental Psychology: Applied*, *15*, 46-62. doi: 10.1037/a0014560

Schreiber, T. A., & Sergent, S. D. (1998). The role of commitment in producing misinformation effects in eyewitness memory. *Psychonomic Bulletin & Review*, *5*(3), 443-448. doi: 10.3758/bf03208819

Schooler, J. W., Gerhard, D., & Loftus, E. F. (1986). Qualities of the unreal. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 171-181. doi: 10.1037/0278-7393.12.2.171

Souchay, C., Moulin, C. J., Clarys, D., Taconnat, L., & Isingrini, M. (2007). Diminished episodic memory awareness in older adults: Evidence from feeling-of-knowing and recollection. *Consciousness and Cognition*, *16*, 769-784. doi: 10.1016/j.concog.2006.11.002

Sporer, S. L. (1992). Post-dicting eyewitness accuracy: Confidence, decision-times and person descriptions of choosers and non-choosers. *European Journal of Social Psychology*, *22*, 157-180. doi: 10.1002/ejsp.2420220205

Sporer, S. L. (1993). Eyewitness identification accuracy, confidence, and decision times in simultaneous and sequential lineups. *Journal of Applied Psychology*, *78*, 22-33. doi: 10.1037/0021-9010.78.1.22

Starns, J. J., Lane, S. M., Alonzo, J. D., & Roussel, C. C. (2007). Metamnemonic control over the discriminability of memory evidence: A signal detection analysis of warning effects in the associative list paradigm. *Journal of Memory and Language*, *56,* 592-607. doi: 10.1016/j.jml.2006.08.013

Starns, J. J., Ratcliff, R., & McKoon, G. (2012). Evaluating the unequal-variance and dual-process explanations of zROC slopes with response time data and the diffusion model. *Cognitive Psychology*, *64,* 1-34. doi: 10.1016/j.cogpsych.2011.10.002

Tousignant, C., Bodner, G. E., & Arnold, M. M. (2015). Effects of context on recollection and familiarity experiences are task dependent. *Consciousness and cognition, 33*, 78-89. doi: 10.1016/j.concog.2014.11.011

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie Canadienne*, *26*, 1. doi: 10.1037/h0080017

Tulving, E. (1989). Memory: Performance, knowledge, and experience. *European Journal of Cognitive Psychology, 1*, 3-26. doi: 10.1080/09541448908403069

Voskuilen, C., & Ratcliff, R. (2016). Modeling confidence and response time in associative recognition. *Journal of Memory and Language*, *86*, 60-96. doi: 10.1016/j.jml.2015.09.006

Wagenmakers, E. J., Van Der Maas, H. L., & Grasman, R. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review*, *14*, 3-22. doi: 10.3758/bf03194023

Weber, N., & Brewer, N. (2004). Confidence–accuracy calibration in absolute and relative face recognition judgments. *Journal of Experimental Psychology: Applied, 10,* 156–172. doi: 10.1037/1076-898x.10.3.156

Weber, N., & Brewer, N. (2006). Positive versus negative face recognition decisions: Confidence, accuracy and response latency. *Applied Cognitive Psychology, 20,* 17–31. doi: 10.1002/acp.1166

Weber, N., Brewer, N., Wells, G. L., Semmler, C., & Keast, A. (2004). Eyewitness identification accuracy and response latency: The unruly 10-12-second rule. *Journal of Experimental Psychology: Applied*, *10*(3), 139. doi: 10.1037/1076-898x.10.3.139

Weber, N., & Perfect, T. J. (2012). Improving eyewitness identification accuracy by screening out those who say they don't know. *Law and Human Behavior*, *36*, 28-36. doi: 10.1037/h0093976

Weber, N., & Varga, M. (2012). Can a modified lineup procedure improve the usefulness of confidence? *Journal of Applied Research in Memory and Cognition*, *1*, 152-157. doi: 10.1016/j.jarmac.2012.06.007

Wells, G. L. (1984). The Psychology of Lineup Identifications1. *Journal of Applied Social Psychology*, *14*, 89-103. doi: 10.1111/j.1559-1816.1984.tb02223.x

Wells, G. L. (1993). What do we know about eyewitness identification? *American Psychologist*, *48*, 553-571. doi: 10.1037//0003-066x.48.5.553

Wells, G. L. (2008). Field experiments on eyewitness identification: Towards a better understanding of pitfalls and prospects. *Law and Human Behavior*, *32*, 6-10. doi: 10.1007/s10979-007-9098-4

Wells, G. L., & Luus, C. E. (1990). Police lineups as experiments: Social methodology as a framework for properly conducted lineups. *Personality and Social Psychology Bulletin*, *16*, 106-117. doi: 10.1177/0146167290161008

Wells, G. L., & Seelau, E. P. (1995). Eyewitness identification: Psychological research and legal policy on lineups. *Psychology, Public Policy, and Law*,*1*, 765-791. doi: 10.1037/1076-8971.1.4.765

Wells, G. L., & Turtle, J. W. (1986). Eyewitness identification: The importance of lineup models. *Psychological Bulletin*, *99*, 320-329. doi: 10.1037/0033-2909.99.3.320

Wells, G. L., Steblay, N. K., & Dysart, J. E. (2015). Double-blind photo lineups using actual eyewitnesses: An experimental test of a sequential versus simultaneous lineup procedure. *Law and Human Behavior*, *39*, 1.doi: 10.1037/lhb0000096

Westfall, J., Judd, C. M., & Kenny, D. A. (2015). Replicating studies in which samples of participants respond to samples of stimuli. *Perspectives on Psychological Science*, *10*, 390-399. doi: 10.1177/1745691614564879

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, *114*, 152-176. doi: 10.1037/0033-295x.114.1.152

Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review*, *117*, 1025-1054. doi: 10.1037/a0020874

Wixted, J. T., & Mickes, L. (2014). A signal-detection-based diagnostic-feature-detection model of eyewitness identification. *Psychological Review*, *121*, 262-276. doi: 10.1037/a0035940

Wright, D.B., & London, K. (2009). Multilevel modelling: Beyond the basic applications. *British Journal of Mathematical and Statistical Psychology*, 62, 439-456. doi: 10.1348/000711008X327632

Wright, D. B., & McDaid, A. T. (1996). Comparing system and estimator variables using data from real lineups. *Applied Cognitive Psychology,* 10, 75-84. doi: 10.1002/(SICI)1099-0720(199602)10:1%3C75::AID-ACP364%3E3.0.CO;2-E

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*, 441-517. doi: 10.1006/jmla.2002.2864

CHAPTER 4

Only Recollect: Improving Identification Accuracy by Reframing Eyewitness Metacognition and Decision Structure[1]

Abstract

We tested metacognitive instructions to improve eyewitness recollection and sensitivity in simultaneous lineups. In four mini-lineup experiments, participants made standard or modified two-step simultaneous lineup decisions. Two-step participants either selected a best-match before indicating culprit presence (best-match-first); or, decided on culprit presence before identifying the culprit (presence-first). We provided witnesses with information about lineup composition, possible errors, and the phenomenology of accurate memories (Experiments 1 & 2). Results indicated a potential effect of greater discriminability when instructions were given with the presence-first procedure, but with more conservative response bias. For a *presence-recollection lineup* combining only phenomenological information with the presence-first decision structure, greater discriminability was observed than for a standard procedure without instructions, without significantly different bias (Experiment 3). Phenomenological instructions had no effect with either standard (Experiment 3) or best-match-first procedures (Experiment 4). Modified decision structure and enhanced metacognitive control of memory decisions can improve eyewitness identification accuracy.

[1] Under review at *Journal of Memory and Language,* as Guerin, N., Weber, N., & Horry, R. (2016). Only Recollect: Improving Identification Accuracy by Reframing Eyewitness Metacognition and Decision Structure. As first author, I designed and conducted the experiments and analyses, interpreted results, and wrote the manuscript. Nathan Weber, as my PhD supervisor, advised on the experimental and analytical design, and contributed to the interpretation of results and manuscript editing. Ruth Horry, as Adjunct Supervisor, contributed to the interpretation of results and the manuscript.

Introduction

What is the purpose of eyewitness identification tasks? To index eyewitness memory for an offender? To experimentally test a police hypothesis about the guilt of their suspect (e.g. Wells & Luus, 1990)? To “get the guy”? Ostensibly, the objective purpose of eyewitness identification tasks is to diagnose suspect guilt or innocence. The value of an identification procedure, then, rests on how well an eyewitness’ decision makes this diagnosis for a suspect. However, when the task is a test of human recognition, perceived purpose can have a critical influence on the outcome. People’s beliefs and expectations of the task shape the effectiveness of their decision and memory processes (e.g., Johnson, Hashtroudi, & Lindsay, 1993; Malmberg & Xu, 2007; Odegard & Lampinen, 2006). Importantly, the identification test might not direct decision-making toward maximal accuracy within the constraints of memory, independent of contextual factors and eyewitnesses’ own preconceptions. To that extent, the reliability of diagnoses will remain uncertain. Eyewitness identification performance has consistently been shown to vary with the influence on witness expectations of instructions, decision context, suggestion, and other factors extrinsic to memory for the offender and the basic task itself (e.g., Brewer & Palmer, 2010; Clark, Howell, & Davey, 2005; Cutler & Penrod, 1995). Research suggests that many eyewitnesses, as lay people who happen to witness a crime, bring to the task a socially shared belief that the purpose of the identification task is to help the police “get the guy” (e.g., Malpass & Devine, 1981; Wells, 1984, 1993), and value identification evidence highly (Semmler, Brewer, & Douglas, 2011). These beliefs lead eyewitnesses to be overly willing to choose a lineup member as the offender, and to neglect important qualities of retrieved memories that indicate a positive identification should not be made (e.g., Wells & Lindsay, 1980). For an innocent suspect who is a plausible match to the offender, the risk of false identification is substantial.

We propose that when lineup members are presented simultaneously, witnesses' desire to choose from the lineup feeds into best-match-first decision processes that are cued by the structure of multiple-alternative choice tasks themselves (N. Guerin, Weber, & Horry, 2016). These tasks focus attention on positive evidence for match to find the best choice, or in the case of identification tasks, the best-match for memory of the offender (Wells & Lindsay, 1980). Emphasis on positive global match evidence directs attention away from potentially available negative recollective evidence against choice, or of mismatch (e.g., Malmberg & Xu, 2007). Therefore, two prime targets for procedural modifications that could potentially improve the accuracy and reliability of identification performance are 1) eyewitnesses' knowledge and beliefs about identification tasks; and 2) structural task features that cue error-prone decision-making strategies. In the four experiments reported here, we investigated the independent and combined effects of two procedural innovations on eyewitnesses' ability to differentiate guilty from innocent suspects: 1) metacognitive instructions designed to increase the accuracy of eyewitnesses' task expectations; and 2) a restructured, presence-first simultaneous procedure designed to disrupt best-match-first decision-making and facilitate more effective decision processes.

In Experiments 1-3 we tested the efficacy of the presence-first procedure alone and in combination with metacognitive instructions, for improving mock eyewitnesses' use of recollection. In Experiments 3 and 4, we tested metacognitive instructions with both a standard, and a modified best-match-first simultaneous procedure. Thus, we could evaluate the effect of instructions on the retrieval of recollection or post-retrieval weighting of familiarity and recollection, independent of the presence-first decision structure. We discuss our rationale for the presence-first restructure of the simultaneous format and the content of metacognitive instructions in turn, before outlining predictions.

A Best-A Match-First Dual-Process Account of Simultaneous Lineups

Eyewitnesses frequently show a bias toward making a positive identification from a police lineup, regardless of the presence of the culprit (e.g., Horry, Memon, Wright, & Milne, 2011; Malpass & Devine, 1981; Wells, 1984, 1993). This predisposition to choose leads to unacceptably high rates of false identification (Brewer & Palmer, 2010; Brewer & Wells, 2011). Dual-process theories of memory propose that both global familiarity and more effortfully retrieved recollective detail can inform recognition decisions (e.g., Malmberg, 2008; Wixted & Mickes, 2010; Yonelinas, 2002). Our best-match-first dual-process account of simultaneous lineup decisions proposes that, in part, false identification occurs because of a neglect of recollection (N. Guerin et al., 2016; see also Gronlund, 2005; Meissner, Tredoux, Parker, & MacLin, 2005). In presenting several lineup members for evaluation, simultaneous lineups encourage eyewitnesses to take a best-match-first approach and make two component recognition decisions to come to a final identification decision. First, eyewitnesses focus on finding the comparative best-match for the culprit, and only then decide whether the best-match is the culprit. To find the best-match, eyewitnesses can focus on positive evidence of global match, without considering recollective detail whose retrieval and monitoring can be more effortful. We propose that best-match-first decision-making by eyewitnesses results in an over-reliance on the global familiarity of lineup members to make their recognition decisions. This focus on relative familiarity leads witnesses to neglect the importance of recollected details of the culprit's appearance. Specifically, recollected details will reinforce evidence of positive match with a guilty suspect, but provide a critical indication of mismatch with innocent lineup members who look similar to the culprit, and so seem familiar (i.e., *recall-to-reject*, Rotello & Heit, 2000; see also *recollection* rejection, Brainerd, Reyna, Wright, & Mojardin, 2003). Moreover, once established for a first recognition task, familiarity-based processes are likely to be applied to subsequent tasks (e.g., Alban & Kelley, 2012). Thus, potentially available recollective evidence of mismatch is neglected in eyewitnesses' evaluation of offender presence (e.g., Malmberg & Xu, 2007). Therefore, greater use of recollection to offset the misleading familiarity of innocent lineup members should reduce false identification without a significant loss of offender identifications (N. Guerin et al., 2016; see also Palmer, Brewer, McKinnon, & Weber, 2010; Mickes, 2015).

A Presence-First Decision Structure

How can this focus on familiarity be addressed? Wells' (1984) idea of not presenting all lineup members simultaneously is one possible method to counter best-match-first decision-making in response to a lineup. However, there is no evidence that simply presenting lineup members one at a time, or in smaller groups, would discourage a focus on positive match in favour of greater use of recollection. The balance of research evidence suggests that sequential presentation encourages more conservative choosing, but without reliably improving overall accuracy (e.g., Clark, 2012; Palmer & Brewer, 2012; Wells, Steblay, & Dysart, 2015). Simultaneous presentation offers the potential latent benefit of maximising memory retrieval and allowing usefully diagnostic comparisons between lineup members (e.g., Goodsell, Gronlund, & Carlson, 2010). For example, attentional demands are high with sequential presentation, potentially including retrospective comparison between successive lineup members and ongoing adjustment of evidence weighting (e.g., Goodsell et al., 2010). Thus, attentional resources might be drawn away from effortful recollection and monitoring processes (Watson, Bunting, Poole, & Conway, 2005). Therefore, we retained simultaneous presentation of lineup members.

Instead, in the presence-first procedure we dampened cues to familiarity-based responding by separating the two component recognition decisions (best-match from offender-presence) and reversing their typical order (N. Guerin et al., 2016). In the presence-first procedure, participants gave identification responses to simultaneous lineups in two steps: first, they indicated only whether or not the offender was present in the lineup (*Present/Not Present)*; and second, they pointed out the offender, if indicated to be present. We designed the presence-first structure to disrupt both witness preconceptions about lineups, and structural decision cues to familiarity-based best-match-first selection and overall neglect of recollection. Instead, participants would be encouraged to evaluate both positive and negative evidence of match, based on both familiarity and recollected detail retrieved in response to lineup members. Evaluations of offender presence (made for the whole lineup, rather than solely for the best-match candidate) would thus have a richer memorial basis more likely to support accurate decision-making.

In previous research, greater discriminability was found for the presence-first procedure than for standard simultaneous lineups (N. Guerin et al., 2016, Experiment 2). As in the present studies, a mini-lineup paradigm was used. Participants viewed a series of target photos and made identification decisions for four-person mini-lineups. Separating and reversing the order of best-match and offender-presence judgments effectively countered best-match-first processes and facilitated the use of recollected evidence of mismatch. However, this promising effect was fully evident only when naïve participants (i.e., with no previous experience making identification judgments) did not know when they made their first offender-presence decision, that they might later be asked to point out the offender. The efficacy of the presence-first structure was constrained when participants knew they might be asked to identify specific offenders, likely heightening the salience of pre-existing beliefs about identification tasks (N. Guerin et al., 2016, Experiment 1).

Further evidence comes from tests of a modified two-step best-match-first procedure in which component recognition decisions were explicitly separated, but not re-ordered: participants first indicated a best-match selection from a simultaneous lineup (regardless of whether they thought the best-match was the offender), and then indicated separately whether the best-match was thought to be the offender (N. Guerin et al., 2016). Identification performance was not significantly improved, as compared with standard simultaneous lineup decisions. Decision separation alone did not enhance the perceived need for recollection in the second task, and there was no evidence that participants would spontaneously re-calibrate to the demands of the second decision, after familiarity-based responding was established in the first (see also, Alban & Kelley, 2012).

Overall, results suggested that a best-match-first approach was difficult to disrupt, once taken. Moreover, when ingrained expectations about the purpose of the identification task were galvanised, they undermined the efficacy of presence-first cues to avert best-match-first processes. Evidently, stronger cues than those provided by the presence-first procedure alone are needed to facilitate better use of memorial information in identification tasks, against eyewitness inclinations. Providing metacognitive information about the diagnostic qualities of accurate memories offered a means to improve eyewitnesses' use of recollection, without jettisoning the potential latent advantages of simultaneous presentation.

Eyewitness Metacognition and Task Instructions

In recognition tests, participants must differentiate previously-seen stimuli from others that have not been seen before. Performance is affected by participants' metacognitive knowledge of both memory processes and the task itself: the perceived purpose of the task, and how it should be tackled (e.g., Cook et al., 2005; Lane, Roussel, Villa, & Morita, 2007; Malmberg & Xu, 2007; McDermott & Roediger, 1998). People's metacognitions vary in accuracy, and in their appropriateness for specific task contexts (e.g., Benjamin, Bjork, & Schwartz, 1998). For each recognition test, sources of metacognitive knowledge and task orientation act within the task structure to influence people's approach to decision-making (e.g., S. A. Guerin, Robbins, Gilmore, & Schacter, 2012). Both retrieval (e.g., Gallo, 2004; Herron & Rugg, 2003), and post-retrieval monitoring processes (e.g., Huff & Bodner, 2013) are shaped by the decision-making approach taken. Insofar as metacognitive knowledge does not accurately reflect task context and the limits of remembering, these processes might decrease the accuracy of recognition decisions (e.g., Lane et al., 2007; Malmberg & Xu, 2007). For example, underestimating the base rate of suspect guilt in police lineups might promote an inappropriately liberal criterion for positive identification, and so increase false identifications (e.g., Brewer & Wells, 2006; Wells & Seelau, 1995). Or, high similarity between a suspect and lineup fillers may lead to lower eyewitness confidence in a positive identification decision than if suspect-filler similarity was lower (e.g., Charman, Wells, & Joy, 2011; Fitzgerald, Price, Oriet, & Charman, 2013). If eyewitnesses' metacognitive understandings could be more accurately aligned with the intrinsic task purpose, underlying memory processes, and effective monitoring strategies, their identification decision-making might improve within the limits of the quality of memory for an offender.

The most direct approach to improving eyewitness decision-making would be to tell them how by providing task-related or metacognitive information before the identification test. However, people are not always able to use metacognitive information to improve their use of underlying memory and decision processes (e.g., Gallo, 2004). In the case of identification tasks, general warnings that the culprit may or may not be in the lineup (i.e., *unbiased instructions*), or to make careful decisions and avoid errors, tend to push eyewitnesses to take a more conservative approach, without promoting better use of memorial evidence (e.g., Clark, 2005; Malpass & Devine, 1981). However, there is not a clear and consistent pattern of evidence for this effect of warnings across lineup contexts (e.g., Clark, 2005). Diverse mechanisms might contribute to these effects, depending on warning content, and witness and situation factors. Thus, metacognitive instructions with demonstrated efficacy in other recognition tasks, might also usefully improve discriminability in identification tasks.

Instructing people about 1) common errors in specific memory tasks, and 2) the use of available metamemorial cues to improve item and event recognition, has been shown to improve use of recollective evidence to avoid false recognition decisions in a variety of recognition and recall tasks (e.g., Gallo, Roediger, & McDermott, 2001; Lampinen, Odegard, & Neuschatz, 2004; Lane, Roussel, Starns, Villa, & Alonzo, 2008; Mather, Henkel, & Johnson, 1997). However, to our knowledge, these kinds of instructions have not been tested with identification tasks. We proposed that metacognitive instructions would improve eyewitnesses' use of recollection in a simultaneous lineup task, but only if we could also offset structural cues that might independently initiate familiarity-based best-match-first decision-making. That is, by providing instructions with a presence-first procedure, rather than a standard or modified best-match-first simultaneous procedure.

Categories of metacognitive information. One focus of research on false recognition (e.g., Yonelinas, 2002) and false memory (e.g., Johnson et al., 1993; Roediger, 1996) has been strategic memory editing processes that can be used at or following retrieval to avoid these kinds of errors (e.g., Koriat & Goldsmith, 1996; Lampinen et al., 2004; Lindsay & Johnson, 1989). Several theoretical orientations (e.g., dual-process recognition memory theories: Malmberg, 2008; Yonelinas, 2002; the source monitoring framework: Johnson et al., 1993; fuzzy-trace theory: Brainerd & Reyna, 2002; Reyna & Brainerd, 1995) have informed the design of metacognitive instructions or warnings to improve participants' use of recollectible memorial information in both recognition and recall tasks (e.g., Blank & Launay, 2014; Chambers & Zaragoza, 2001; Lane et al., 2007; Lane et al., 2008). Two metamemorial mechanisms have been targeted: directly increasing recall or recollection in response to test stimuli; and, drawing attention to diagnostic features of retrieved memorial evidence. Overall, two broad categories of metacognitive instructions have been tested: 1) task-related; and 2) phenomenological. For example, in a misinformation paradigm, task-related instructions encourage attention to source evidence, by alerting participants to the possibility of exposure to inaccurate post-event misinformation (e.g., Johnson et al., 1993). For associative tasks, task-related instructions about the inclusion of foils at test that are misleadingly similar to studied items highlight the usefulness of effortfully target recollection (e.g., Neuschatz, Benoit, & Payne, 2003; Starns, Lane, Alonzo, & Roussel, 2007; Westerberg & Marsolek, 2003). In contrast, phenomenological instructions inform participants about qualities of the remembering experience that index the accuracy of retrieval (e.g., Lane et al., 2007; Lane et al., 2008; Mather et al., 1997). Both task-related and phenomenological instructions are potentially useful in the identification context.

Task-related metacognitive instructions. Regardless of pre-existing beliefs about the purpose of a task, participants often lack important knowledge about how memory works or the specific features of memory that will be important for accuracy (e.g., Koriat & Goldsmith, 1996). For example, in the DRM paradigm (Roediger & McDermott, 1995), researchers have tested the idea that, in part, memory errors occur because participants do anticipate false associative memory to be a risk of the task (e.g., Gallo, Roediger, & McDermott, 2001; Neuschatz, Payne, Lampinen, & Toglia, 2001). That is, participants do not realise the need to encode and retrieve specific recollective detail of studied words, to offset the misleading familiarity of related words. When informed of the risk, false recognition can be reduced (e.g., Watson, McDermott & Balota, 2004). DRM warnings are most reliably effective when given before encoding (e.g., Gallo, Roberts, & Seamon, 1997; Neuschatz et al., 2001, cf. McCabe & Smith, 2002), but encoding-stage mechanism of memory improvement are not useful for eyewitness memory. There is some evidence of an effect when given prior to retrieval (e.g., Anastasi, Rhodes, & Burns, 2000; Gallo et al., 2001), indicating the contribution of multiple mechanisms to false recognition in the DRM and other paradigms (e.g., Gallo, Bell, Beier, & Schacter, 2006; Gallo, 2010). Further, a retrieval-stage effect of post-study warnings is more consistently evident in source monitoring, misinformation, and associative recognition paradigms (e.g., Blank & Launay, 2014; Lane et al., 2007; Starns et al., 2007). In applied source monitoring tasks (e.g., using a misinformation paradigm), participants can use pre-retrieval task-related instructions to distinguish between accurately recognised and misinformed test items (e.g., Echterhoff, Hirst, & Hussy, 2005; Wyler & Oswald, 2016) insofar as source recollection is available (e.g., Dodson, Holland, & Shimamura, 1998; Horry, Colton, & Williamson, 2014). For associative recognition, pre-retrieval warnings about high target-foil similarity, or information about the type of memorial information needed to avoid false recognition, effectively improved participants' ability to differentiate targets from highly-similar unstudied foils (e.g., Lampinen et al., 2004; Starns et al., 2007).

We argue that fair identification lineups (when fillers were chosen to match the description of the offender) can be likened to item and associative tasks with substantial target-foil similarity. Specifically, they share a characteristically high degree of target-foil similarity whose neglect contributes to false recognition. Therefore, instructions that help participants to avoid false recognition in these more basic tasks are likely to help eyewitnesses avoid false identification in lineup tasks. In the main, as noted, tests of task-related instructions in identification tasks have been limited to investigations of the effects of unbiased instructions (e.g., Malpass & Devine, 1981) that predominantly rein in unduly liberal identification criteria, without improving discriminability. However, when Palmer, Brewer and Weber (2010) gave mock eyewitnesses performance feedback after an initial lineup decision, discriminability in a subsequent lineup decision about the same culprit was affected, depending on the accuracy of the first decision. This finding demonstrates that participants can adjust their use of available memory in response to extrinsic information. However, as few warnings or metacognitive instructions have been tested with eyewitness tasks, the effect of pre-retrieval interventions tested more extensively in other recognition formats is unknown.

A key difference between the unbiased instructions given to witnesses and the task-related instructions provided to participants in source monitoring and associative recognition tests lies in their degree of specificity. Unbiased instructions caution generally against an assumption of offender presence, without providing detailed information about task features, or alerting eyewitnesses to known foibles of identification decision-making. In contrast, source monitoring and associative task instructions inform participants about critical design features, associated errors, error mechanisms, and even researchers' reasons for investigation (e.g., Lampinen et al., 2004). Blank and Launay (2014) proposed a similar specificity distinction in their taxonomy of misinformation paradigm warnings. Unfortunately, due to substantial variation in test conditions, and small number of studies per cell, their meta-analysis did not support strong conclusions about the moderating effect of specificity on the effect of warnings in the misinformation paradigm. However, we argue that converging evidence suggests that specific information about potential errors, and the reasons people are prone to these errors with simultaneous lineups, might allow eyewitnesses to preferentially avoid false identification, rather than only adopting a more conservative criterion for positive identification overall. Indeed, Blank and Launay found a moderating effect of "enlightenment", whereby pre-retrieval debriefing information increased the efficacy of other task-related instructions. That is, providing information about a task and investigators' research goals enhanced the effect of instructions in increasing discriminability by highlighting the importance of effortful recollection, or appropriate weighting of familiarity and recollection. By extension, task-related instructions likely to improve eyewitness identification performance might include information about: 1) methods of filler-selection; 2) high target-foil similarity; 3) the implications of high target-foil similarity for false identification; and 4) the avoidance of identification errors by focussing on recollection, rather than global familiarity alone.

Phenomenological metacognitive instructions. A feature of studies testing task-related instructions is that task-specific content is also frequently accompanied by information about the experience of remembering (e.g., Lane et al., 2007; Lane et al., 2008; Neuschatz et al., 2001). This information guides participants to attend to phenomenological aspects of retrieved memories that index accuracy, including relative vividness, clarity, and detail (e.g., Mather et al., 1997; Norman & Schacter, 1997). Pre-retrieval instructions featuring only task-related content have been shown to effectively enhance discriminability (Lampinen et al., 2004; Starns et al., 2007). However, results from a number of studies featuring high target-foil similarity suggest that the efficacy of task-specific content might sometimes rely on, or work in concert with, enhanced phenomenological awareness, or vice versa (Lane et al., 2007; Lane et al., 2008).

The memory characteristics described by phenomenological instructions are importantly distinct from those implemented in *Remember-Know* (*RK*) ratings tasks (Gardiner & Richardson-Klavehn, 2000). RK ratings require participants to make a binary categorisation of memory experience into that of a global-familiarity-based sense of *knowing*, or a recollection-informed sense of *remembering*, corresponding to the relative contributions of the two objective memory retrieval processes (Tulving, 1985, 1989). While participants' use of these ratings varies with context (e.g., Bodner & Lindsay, 2003; Tousignant, Bodner, & Arnold, 2015), the conscious experience of remembering along this distinction has been shown to categorically differ, corresponding to two underlying continuous memory processes of familiarity and recollection, as Tulving (1985) proposed (Selmeczy & Dobbins, 2014; see also Dodson et al., 1998; Ingram, Mickes, & Wixted, 2012). However, the provision of binary RK information itself before test does not help participants to improve their recognition accuracy (e.g., Anastasi, Rhodes, & Burns, 2000; Mather et al., 1997). In contrast, instructions that draw participants' attention to other specific characteristics of remembering that vary along a continuum and can help to differentiate memories likely to be more accurate (informed by recollection) from those less likely to be accurate (as based primarily on potentially misleading familiarity) are useful in reducing false recognition (e.g., Lane et al., 2007; Lane et al., 2008; Mather et al., 1997; Schooler, Gerhard, & Loftus, 1996). These characterisations of memory quality correspond to experiences of remembering with varying and partial contributions of recollection and familiarity, and differentiate effectively between true and false recognition (e.g., Johnson, Foley, Suengas, & Raye, 1988; Lampinen, Odegard, & Bullington, 2003; Mather et al., 1997; Neuschatz et al., 2001; Norman & Schacter, 1997), even when false recognition is subjectively compelling (Payne, Elie, Blackwell, & Neuschatz, 1996; Schooler et al., 1996).

Phenomenological instructions given prior to retrieval that describe the characteristics of accurate remembering have been shown to improve recognition accuracy independently, or when combined with task-specific information (e.g., Mather et al., 1997; Schooler et al., 1996; Starns et al., 2007). Their efficacy extends to paradigms for which task-related instructions have proven ineffective or unreliable (e.g., Lane et al., 2007; Lane et al., 2008; Mather et al., 1997). However, results are variable (e.g. Neuschatz et al., 2001). Most importantly, neither detailed task-related metacognitive instructions that spotlight the need for recollection, nor phenomenological instructions to improve metamemorial knowledge and monitoring strategies, have yet been tested in eyewitness identification tasks, to our knowledge. Overall, multiple mechanisms contribute to the occurrence of false recognition and its reduction in non-identification tasks (e.g., Dodson & Schacter, 2002; Gallo et al., 2006; Gallo, 2010; Lampinen, Neuschatz, & Payne, 1997). The same is likely true for the more complex, applied test of recognition in eyewitness identification procedures. Therefore, in the first instance, we tested the combined effect of both specific task-related information and phenomenological information in a single set of instructions, to maximise their likely efficacy (e.g., Lane et al., 2008).

The Experiments

In four face recognition mini-lineup experiments, we tested the effectiveness of instructions about 1) the nature of lineup tasks and common familiarity-based errors (*task-related instructions*); and, 2) the diagnostic phenomenology of accurate recollection (*phenomenological instructions*) in standard and modified simultaneous lineup tasks. Participants studied a series of target face photos and made identification decisions about a four-person simultaneous mini-lineup for each target. The mini-lineup paradigm allowed collection of multiple data points per participant to evaluate response patterns across participants and for multiple stimuli. Given the novelty of testing task-related and phenomenological instructions in a lineup paradigm, it was important to gather substantial data to evaluate these regularities and increase generalisability before contemplating testing in a more realistic identification paradigm (cf., Gronlund, Mickes, Clark, & Wixted, 2015; Lane & Meissner, 2008; Sauer et al., 2008; Weber et al., 2004). More realistic paradigms that use only a limited number of mock crime stimulus videos (often only one), require a large sample to power conclusions that have potentially limited generalisability, depending on the stimuli used. Hence, the mini-lineup paradigm represents a more controlled and informative approach for initial tests of the memory and decision processes we aimed to investigate.

Participants completed one of three simultaneous lineup types: standard simultaneous; presence-first; or best-match-first. Standard simultaneous participants either made a positive identification, or indicated that the offender was not present in the lineup. Presence-first participants judged first whether or not the offender was present in the lineup, before identifying the offender (if present). Best-match-first participants first selected the best-match to memory for the offender, then indicated whether the best-match was the offender. Each lineup type was tested with and without experimental instructions.

We had three major research questions. First, would the previously observed (N. Guerin et al., 2016, Experiment 2) effect of greater discriminability with a presence-first decision-structure than for a standard simultaneous procedure (both without instructions) be replicated with non-naïve participants? We expected that non-naive participants' preconceptions about the purpose of identification tasks would not be highly salient with the presence-first procedure when they did not know they might need to pick someone from the lineup, as for naïve participants in the previous experiment (N. Guerin et al., 2016, Experiment 1). Therefore, we expected greater discriminability for the presence-first procedure than a standard simultaneous procedure.

Second, would providing participants with metacognitive instructions with the presence-first lineup structure allow mock witnesses to better differentiate guilty from innocent suspects than witnesses following a standard simultaneous procedure? We predicted that, when combined, a presence-first decision structure and metacognitive instructions would produce greater discriminability than that shown for a standard simultaneous procedure.

Third, would the best-match-first decision-making approach characteristic of standard and modified best-match-first procedures render metacognitive instructions ineffective? We predicted that eyewitnesses following structural cues to best-match-first decision processes would be less able to make use of metacognitive instructions than with a presence-first structure. Therefore, we predicted that any effect of greater discriminability with instructions in the standard or modified best-match-first procedures would be smaller than that observed with the presence-first procedure.

Experiment 1

Our first aim was to replicate the finding of greater discriminability with the presence-first lineup than a standard simultaneous procedure (N. Guerin et al., 2016). Witnesses in the original experiment were novices without any prior participation in eyewitness identification paradigm experiments, whereas in the present experiments they were not. Second, we aimed to test the combined effect on discriminability of 1) a presence-first decision structure; and 2) metacognitive instructions featuring task-related and phenomenological information in a *presence-first-instructions* lineup condition. We predicted greater discriminability for the presence-first-instructions lineup than with the standard simultaneous procedure.

Method

Except where indicated, all experiments followed the same method.

Participants. A priori, we aimed to collect data from 30 participants per lineup condition. Estimates were based on achieving sufficient power to test hypotheses in mixed-effects models (Gelman & Hill, 2007). All participants in Experiments 1-4 were undergraduate students from Flinders University who participated for payment or course credit. In Experiment 1, 93 students (65 female, age: M = 23.30, SD = 6.56 years) participated. All collected data were analysed.

Materials. We used a collection of mini-lineups constructed by Weber and Brewer (2004, 2006) comprising 40 sets of five description-matched colour photos. Photos showed a front view of the head and neck with a neutral facial expression and background (200 × 200 pixels) and were presented on an 18” monitor (resolution: 1920 × 1080 pixels). In each set, we randomly designated the same target, target-replacement (for target-absent mini-lineups), and three foils, for all participants. Face sets were divided into two blocks of 20. In each block, half of the trials were target-present (target plus foils) and half were target-absent (target-replacement plus foils). All mini-lineups were presented in a row of four, with response buttons below, all horizontally centred. To enable participants to look for a specific target in each mini-lineup (Weber & Brewer, 2006), a text cue (name or occupation, by block) was presented with each target (study) and mini-lineup (test), centred horizontally above photos. Photo order in each mini-lineup, trial order in each block, and cue and target-presence assignment, were randomly determined for each participant. Randomisation of name cues was constrained by the gender of the photo subject.

Procedure. After giving informed consent, participants completed the whole experiment on PC in individual cubicles using a mouse to click on-screen buttons in response to written prompts. They completed two blocks of trials (after four practise trials), with block order, and assignment of name or occupation cues to blocks, counterbalanced across participants. Each block included a study phase, a visual distracter task, and a test phase. In the study phase, target photos were presented sequentially, in the centre of the screen. Each cue was presented for 1000 ms before presentation of the target, and throughout the target's exposure (an additional 1000 ms). Following each study phase, participants worked on a spatial memory task for 3 min. In each test phase, participants completed 20 mini-lineup trials. Participants were told that each mini-lineup may or may not include a target (i.e., unbiased instructions, Malpass & Devine, 1981). The participant's response and latency were recorded. We manipulated lineup condition on three levels (standard; presence-first; presence-first-instructions), between-subjects.

In the *standard* lineup condition, participants provided a single response. They identified a lineup member as the studied target (by clicking on the photo), or clicked the *Not Present* button.

In the *presence-first* lineup condition (following N. Guerin et al., 2016, *Experiment 2*), participants gave two identification responses in a modified two-step procedure. The two identification decisions were blocked. Participants first completed two blocks of trials making only their first decision for each mini-lineup. In this first decision, they indicated whether the studied face was *Present* or *Not Present* in the mini-lineup. Before making this first decision, participants were not told that they would later be asked to make a second decision for each mini-lineup. After all trials were completed, participants were presented in series with lineups for which they had made a positive presence decision (*Present),* and asked to click on the face identified as the target.

In the *presence-first-instructions* lineup condition, participants followed the same procedure as those in the presence-first condition, but with two important variations in instructions provided before each test phase. First, we wanted to approximate conditions for real eyewitnesses asked to view a suspect in a lineup. Specifically, even if initially asked to decide only on the presence of the offender, a real-world witness would anticipate being asked to indicate the offender, if present, in some way. Therefore, before making their first decisions (*Present, Not Present*), participants were told that they would later be asked to make a second target-identification decision for each mini-lineup for which they had responded that the target was *Present.* Second, they were given instructions that included information about 1) the task structure and likely errors in lineup tasks (*task-related instructions)* and 2) aspects of the phenomenology of remembering that are diagnostic of accuracy (*phenomenological instructions)*.[2] In task-related instructions, participants were told that lineups were description-matched: "*Each group of photographs in this recognition test is made up of faces that match the same description as the studied face in gender, face shape, hair colour and style, eye colour, skin colour and complexion, and age. However, some faces in the group will look more similar to the studied face than others."* They were further told that previous research had shown that "*Because some of the faces in the group are very similar to the studied face, people sometimes mistakenly decide that the studied face is*

[2] Instructions for all experiments are included in full in online supplemental materials.

present when all four photographs show faces that have not been studied. When people focus on the similarity between test faces and the studied face, they pay less attention to evidence that the test faces are not the studied face." In phenomenological instructions, we told participants that more clear, vivid memories, with recollected detail, are diagnostic of likely accuracy: "*Research has shown that accurate decisions about the studied face being present are likely to be accompanied by more vivid and clear memories of studying the face than when an unstudied test face is inaccurately and mistakenly recognised. These can include memories of how the studied face looked, memories of how the name/occupation and the studied face are related, and other thoughts and feelings that were experienced when the face was studied."* Both task-related and phenomenological instructions were repeated when we asked participants to use this information to improve the accuracy of their identification decisions: "*We would like you to use this information to accurately decide when a studied face is present in the test group of photographs and when it is not. We would like you to avoid mistakenly and inaccurately recognising a test face that is similar to the studied face, but was not studied. When you think that the face you studied with the name/occupation is present, think about how clear, vivid and detailed your memory is. Can you remember the thoughts and feelings you experienced when you studied the face? If you have a clear and vivid memory of the studied face, you should indicate that the studied face is Present. If you think the test photograph is similar to your memory for the studied face, but your memory is not clear, vivid or detailed, you should respond Not Present".* We reminded participants of these final instructions at the halfway point of each test block.

Following the final test phase, all participants completed a manipulation check questionnaire that focussed on the content of task instructions and participants' own decision strategies. Three open-ended items, 1 option-ranking item, and 17 items requiring responses on 7-point Likert scales addressed: 1) participants' knowledge about the content of the task-related and phenomenological instructions provided; and 2) participants' self-reported identification strategies. See online supplement for the complete questionnaire.[3]

Results & Discussion

Overview. We used an alpha level of .05 for all inferential analyses and focused on trial-by-trial performance for identification decision response bias (choosing) and discriminability (e.g., De Carlo, 1989; Wright & London, 2009). For confirmatory hypothesis testing, we evaluated logistic mixed-effects regression models that included all fixed factors of interest for key pairwise comparisons (Barr, Levy, Scheepers, & Tily, 2013; Gelman & Hill, 2007; Gelman, Hill & Yajima, 2012). Models included random effects, allowing evaluation of fixed factors after accounting for random variability within and between non-independent observations for multiple stimuli from our participant sample (e.g., Baayen, Davidson, & Bates, 2008; Judd, Westfall, & Kenny, 2012; Murayama Sakaki, Yan & Smith, 2014). To avoid over-parameterising models, we included random effects according to importance in the improvement of model fit, without undue reduction in power to test predictions. Models that predicted identification decision type (lineup rejection; positive identification) typically included random intercept by participant and stimulus, and random slopes by participant and stimulus for within-subjects and within-stimulus manipulations, respectively (random effects for all reported models are presented with supplemental materials). Data were analysed using the lme4 package (Bates, Maechler, Bolker, & Walker, 2014) for R (R Core Team, 2014).

Logistic regression models predicting a binary categorical outcome variable give the log-odds for observing a given outcome. Log odds *v*alues are symmetrical around 0: values greater than 0 indicate the predicted outcome (e.g., a positive identification decision) is more likely; and values lower than 0 indicate the alternative outcome (e.g., a lineup rejection) is

[3] Supplemental materials for this chapter are presented in Appendix 2

more likely. Coefficients of individual binary fixed factors in a model represent the log odds ratio (*lnOR)* for the two factor levels. We examined the coefficients in maximal models to evaluate the significance of the effects of factors on the outcome variable. Effects were taken to be significant if the 95% confidence interval (*CI*) around the coefficient (*lnOR* ± [1.96 * SE_b]) excluded 0. We also report odds ratios (OR) for differences between experimental conditions to indicate effect size. Aggregate descriptive statistics are presented as an index of identification performance.

When a witness identifies a known-innocent filler lineup member from an identification lineup, this is a known error. Therefore, filler identifications need not be considered for evaluating applied questions about the reliability of a lineup procedure as a test of suspect guilt. However, patterns of filler identifications can also indicate differences in memory or decision processes (Clark & Godfrey, 2009; Wells, 2008) and have important theoretical implications. Therefore, we report analyses with filler identifications included in the dataset, and without.

Decision type and accuracy (choosing and discriminability). For all experiments, we compared the likelihood of choosing (positive identifications of suspects and fillers) pairwise between lineup procedures. We also examined the effect of lineup on the accuracy of lineup decisions, using a measure of comparative discriminability. With filler identifications included, discriminability indexed the accuracy of target *presence* decisions, regardless of whether the lineup member identified was the suspect (innocent or guilty) or a filler. When filler identifications (from both target-absent and target-present lineups) were excluded, discriminability indexed the accuracy of positive identifications (i.e., witnesses' ability to discriminate guilty from innocent suspects). Patterns of effects were similar with or without filler identifications and we discuss these patterns together, with specific reference to filler identification patterns when pertinent.

All predictions for pairwise lineup comparisons were tested in models with decision type (lineup rejection or positive identification) as the outcome variable. Lineup, target presence (target-absent = 0, target-present = 1), and their interaction were fixed factors. The effect of lineup indexed any difference in witnesses' tendency to choose (response bias), while the effect of target presence provided an index of the accuracy of decisions, independent of bias (discriminability). The interaction term indicated how much discriminability varied between lineup procedures (De Carlo, 1989; Wright & London, 2009). Identical analyses were conducted for each experiment, and effects are most easily interpreted with reference to the plotted model.

Presence-first decision structure. To test replication of the finding of greater discriminability with a presence-first procedure than a standard procedure (N. Guerin et al., 2016) we compared the standard with the presence-first condition (i.e., without metacognitive instructions) with filler identifications included in analyses. Neither lineup, nor the lineup × target presence interaction term significantly predicted decision type (lineup rejection; positive identification) (see Tables 1 & 2 for all models relating to this experiment). Thus there was no evidence of a significant difference in choosing or discriminability, respectively (Figure 1, Panel A). Figure 1 (Panel B) shows that when filler identifications were excluded, response bias was lower for the presence-first procedure (significant effect of lineup), with no significant advantage in discriminability. Presence-first witnesses were thus less likely to identify both innocent and guilty suspects, but without the predicted increase in discriminability that would be commensurate with significantly greater use of recollection. Therefore, the presence-first advantage was not replicated with non-novice witnesses.

Table 1

Model Fit Statistics for Mixed-Effects Models Predicting Decision Type (Positive Identification; Lineup Rejection) by Lineup Type in Experiment 1

Filler Identifications	Comparison	χ^2	*df*	*p*
Included	Standard vs Presence-First (PF)	72.64	3	< .001
	Standard vs PF Instructions (PFI)	95.41	3	< .001
	PF vs PFI	84.39	3	< .001
Excluded	Standard vs PF	57.05	3	< .001
	Standard vs PFI	80.99	3	< .001
	PF vs PFI	60.49	3	< .001

Number of observations with fillers included (fillers excluded): Standard vs PF = 2400 (2220), *n* = 60 (60); Standard vs PFI = 2520 (2384), *n* = 63 (63); PF vs PFI = 2520 (2344), *n* = 63 (63).

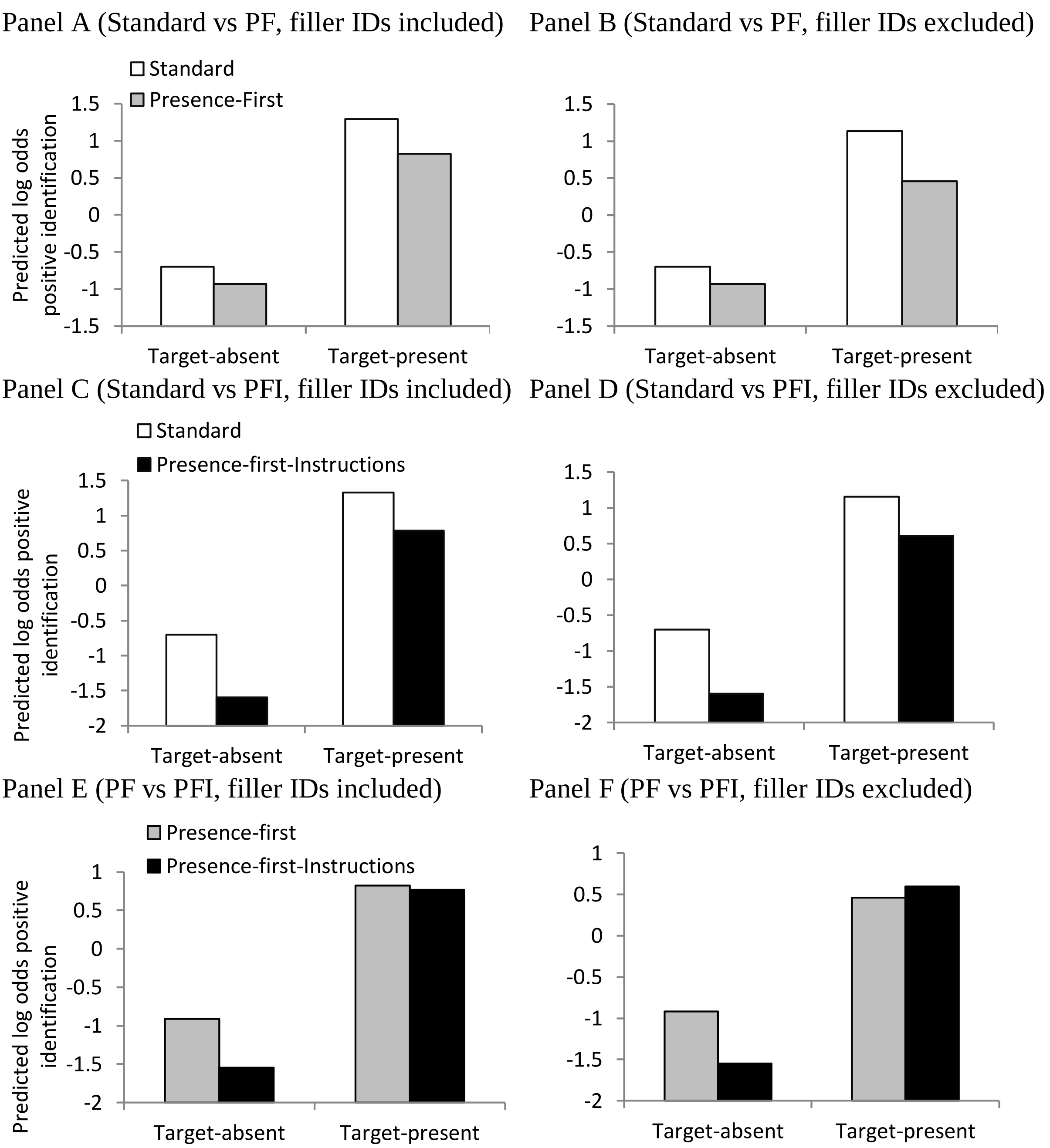


Figure 1. Plots of predicted log odds of a positive identification decision in Experiment 1 by lineup type and target presence. Values greater than 0 indicate a positive identification from the mini-lineup is more likely, and values lower than 0 indicate a lineup rejection is more likely. Coefficient standard errors, indicative of the variability of the estimates, are displayed in Table 2.

Table 2

Fixed Effect Coefficients for Mixed-Effects Model Predicting Decision Type (Lineup Rejection; Positive Identification) by Lineup Type in Experiment 1

		Lineup Comparisons								
		#Standard vs PF			#Standard vs PFI			#PF vs PFI		
Filler Identifications	Fixed Effect	*b*	SE_b	95% CI_b	*b*	SE_b	95% CI_b	*b*	SE_b	95% CI_b
Included	Intercept	-0.70			-0.70			-0.91		
	Lineup	-0.23	0.30	-0.82, 0.36	-0.89*	0.31	-1.49, -0.28	-0.63*	0.27	-1.16, -1.03
	TP	1.88 *	0.25	1.51, 2.47	2.03*	0.25	1.55, 2.52	1.74*	0.24	1.26, 2.22
	Lineup × TP	-0.24	0.35	-0.93, 0.45	0.34	0.35	-0.34, 1.04	0.58	0.34	-0.10, 1.25
Excluded	Intercept	-0.70			-0.70			-0.92		
	Lineup	-0.23*	0.30	-0.81, 0.35	-0.90*	0.31	-1.11, -0.28	-0.63*	0.27	-1.16, -1.02
	TP	1.83*	0.27	1.30, 2.36	1.86*	0.27	1.31, 2.39	1.38*	0.27	0.84, 1.91
	Lineup × TP	-0.44	0.38	-1.19, 0.31	0.35	0.39	-0.42, 1.11	0.77*	0.38	0.02, 1.52

Note: * indicates significant predictors, # indicates the reference lineup condition. Number of observations, fillers included (fillers excluded): Standard vs PF = 2400 (2220), *n* = 60 (60); Standard vs PFI = 2520 (2384), *n* = 63 (63); PF vs PFI = 2520 (2344), *n* = 63 (63).

As anticipated, a potential explanation for this finding is that non-naïve participants in this experiment, who had previously taken part in lineup experiment(s), brought different expectations to the task than the novices in the original experiments. In both procedures, witnesses made their first and second identification judgments for the two-step presence-first procedure in separate test blocks. Importantly, before providing responses for their first (presence) decision, they were not told that they would later be asked to provide a second (offender selection) response for mini-lineups indicated to include a target. We proposed that in the original experiment this prevented naïve participants from adopting a best-match-first approach for their first presence decisions. However, in the current experiment, witnesses' prior experience with a standard lineup procedure might have led them to adopt a best-match-first approach to decision-making overall. As we have outlined, ample evidence from basic (e.g., Cook et al., 2005; Malmberg & Xu, 2007) and applied (e.g., Lane et al., 2007) recognition tasks has demonstrated that task expectations are likely to critically influence the use of available memorial information in this way.

This possibility highlights the importance of considering eyewitness expectations for the conduct of real police identification procedures, and for the development of better procedures. The effectiveness of structural or other changes to lineup procedure to improve eyewitness decision-making are likely to be limited by witnesses' prior knowledge and expectations of procedural aims. Witnesses will, of course, generally anticipate that the aim of a lineup is to identify an offender, if present. Or, at least, to help police diagnose the guilt or innocence of their suspect. Therefore, any procedure that relies on the absence of these expectations will be of little practical use, even if effective in the laboratory.

Metacognitive instructions. In this experiment, our main aim was to test the effectiveness of providing task-related and phenomenological instructions about the use of memory in identification tasks with a presence-first procedure. More specifically, we proposed that instructions, with the presence-first decision structure, would offset any encouragement of best-match-first decision-making by witnesses' expectations about lineup tasks. The pattern of responses to manipulation check questionnaire items was generally consistent with participants having attended to key elements of metacognitive instructions, when provided (descriptive statistics and pairwise comparisons are reported in the online supplement). We predicted that instructions would enable witnesses to make better use of memory with the presence-first procedure (as shown by greater discriminability than for a standard procedure), even with the knowledge that they might later be asked to pick an offender from the lineup. However, comparison of standard and presence-first-instructions procedures (see Table 2, and Figure 1, Panels C & D) revealed significantly lower response bias for the presence-first-instructions procedure (significant effect of lineup), with no significant difference in discriminability (the lineup × target presence interaction term was not a significant predictor). Thus, providing instructions with the presence-first procedure appears to have made participants more reluctant to choose from the lineup (more conservative response bias), without enabling them to better differentiate between innocent and guilty suspects.

Inspection of these two models (standard vs presence-first; standard vs presence-first-instructions) revealed that the effect of lineup on response bias was larger for presence-first-instructions, than presence-first (without instructions). This difference in effect size was carried by a lower likelihood of false identifications for the presence-first instructions than the presence-first procedure (i.e., in the direction of greater discriminability). To further explore the effect of instructions, we directly compared the presence-first procedures with and without experimental instructions in a third model (Tables 1 & 2). Panels E & F of Figure 1 clearly illustrate significantly greater discriminability when the presence-first procedure was conducted with instructions, than without. Further, there was no significant difference between the presence-first procedures in participants' likelihood of accurately identifying a guilty suspect. Thus, the provision of task-related and phenomenological instructions allowed participants in a presence-first procedure to more successfully avoid false identifications than when participants followed the same procedure without these instructions. Greater discriminability was achieved without any significant difference in the likelihood of accurately identifying a guilty suspect (a *hit*).

In sum, the presence-first procedure, either with or without instructions, led to more conservative response bias, without better discriminability, when compared with a standard simultaneous procedure. However, providing metacognitive instructions to witnesses with the presence-first decision structure encouraged better use of memory to avoid false identification decisions than in the presence-first procedure without instructions, with no significant effect on the likelihood of accurately identifying an offender. This finding of greater discriminability, driven by a lower likelihood of false identification, is particularly notable, given the already lower response bias in the presence-first procedures, compared with the standard lineup. That is, even with a significantly lower likelihood of false identification in the presence-first procedure than for standard lineups, metacognitive instructions lowered the risk for innocent suspects further.

Experiment 2

Robust evidence from the eyewitness identification literature demonstrates that unbiased instructions (that the offender may or may not be in the lineup) and strong warnings against false identification produce conservative responding (e.g., Malpass & Devine, 1981). It is possible that the task-related instruction elements in Experiment 1 had disproportionately highlighted the possible error of false identification, along the same lines as these warnings. If so, the warning effect might have overshadowed witnesses' ability to use phenomenological instructions elements. As a result, while discriminability was better with instructions than for the presence-first procedure alone, the effect of instructions was not sufficient to produce significantly greater discriminability than in the standard procedure. Therefore, in Experiment 2, we greatly reduced the emphasis on the avoidance of false identifications, instead focussing strongly on the usefulness of phenomenology.

We also sought to test the effect of instructions when given in a standard simultaneous procedure. We proposed that the standard procedure's encouragement of best-match-first decision-making would interfere with the effect of metacognitive instructions, rendering them less effective than when given with the presence-first procedure. If, instead, the standard procedure did not prevent eyewitnesses from using metacognitive instructions, we expected to see significantly better discriminability with instructions than without.

We adopted a mixed design in Experiment 2 to allow: 1) a further between-subjects test of the presence-first as compared with the standard procedure; and 2) within-subjects tests of the effect of instructions with both standard and presence-first procedures. All participants first completed 40 test trials in either the standard or presence-first procedure, before then completing a further 40 test trials with the same procedure, with the addition of metacognitive instructions. Testing across a variety of stimuli is crucial for increasing the generalizability of results for these important applied questions, and is particularly important when testing a novel procedure (Brewer, Weber, Wootton, & Lindsay, 2012; Gronlund et al., 2015). We put together an additional collection of mini-lineup sets to use with the original collection to broaden the number and diversity of tested stimuli. Including these stimuli also created a stronger test of the presence-first procedure and metacognitive instructions in that target appearance varied between study and test (Hancock, Bruce, & Burton, 2000).

Method

Participants. In Experiment 2, we tested 64 participants (47 female, age: M = 21.97, SD = 6.11 years).

Materials. We used both the original collection of mini-lineups constructed by Weber and Brewer (2004, 2006) as used in Experiment 1, and a second collection of 40 sets that we constructed using the same method, with one key difference. In the second collection of mini-lineups, photos of designated targets were varied from study to test. As in the original set, target photos at test showed a front view of the head and neck with a neutral facial expression. However, target photos at study showed either a front view of the head and neck with a different facial expression (smiling), or a three-quarter view of the head and neck with the same facial expression. Study-test variation in face targets has been demonstrated to make recognition more challenging (e.g., Hancock et al., 2000), providing a stronger test of modified procedures, and more closely approximating encoding-test variation for eyewitnesses.

Procedure. After giving informed consent, each participant completed 80 mini-lineup trials in four blocks of 20. The manipulation of lineup type followed that in Experiment 1, with the following key differences.

Lineup. Lineup condition was manipulated on two levels between-subjects (standard; presence-first) and instructions were manipulated on two levels within-subjects (none; instructions). The block order of mini-lineup collections was counterbalanced within lineup and instructions conditions. The *standard* and *presence-first* lineup conditions were identical to those in Experiment 1, but for the additional within-subjects manipulation of instructions.

Instructions. Instructions were varied on two levels within-subjects (*none*; *instructions*). In the first two blocks of the experiment, participants followed the same procedure as for their respective lineup conditions in Experiment 1. Then, before the test phases in two further blocks of mini-lineup trials, all participants were given the task-related and phenomenological information provided in the *instructions* condition in Experiment 1, with the following important change: *task-related* references to the structure of lineup tasks and similarity-based identification errors were removed. *Phenomenological* instructions were unchanged. However, when we asked participants to use this information to make accurate identification decisions, they were still instructed that: "*If you think the test photograph is similar to your memory for the studied face, but your memory is not clear, vivid or detailed, you should response Not Present.*" That is, one explicit instruction to avoid similarity-based positive identification decisions remained.

All participants completed the same manipulation check questionnaire as for Experiment 1.

Results & Discussion

We first compared response bias and discriminability for the standard and presence-first procedures *without* instructions, to test replication of the presence-first advantage found previously (N. Guerin et al., 2016). No significant differences were observed (see Tables 3 & 4, Figure 2, Panels A & B). However, the magnitude of the coefficients for the effect of lineup was not trivial, and their CIs largely included values less than zero. This suggested meaningfully lower response bias with the presence-first than the standard procedure, consistent with Experiment 1. Therefore, we again failed to replicate the presence-first advantage previously observed with novice eyewitnesses (N. Guerin et al., 2016, *Experiment 2*), with participants who had previously participated in experimental tests of standard simultaneous lineups.

The pattern of responses to manipulation check questionnaire items was generally consistent with participants having attended to key elements of metacognitive instructions, when provided (descriptive statistics and pairwise comparisons are reported in the online supplement). In separate models, we compared the performance of witnesses given metacognitive instructions in the presence-first procedure with 1) that of witnesses in the standard procedure (without instructions); and 2) their own performance in initial presence-first test blocks performed without the instructions. For both comparisons, lineup type significantly predicted the likelihood that a witness would make a positive identification from the lineup (see Tables 3 & 4). Specifically, witnesses in the presence-first procedure with metacognitive instructions were significantly less likely to choose from the lineup than witnesses in the standard procedure (without instructions), and less likely than they themselves had been in the presence-first procedure without instructions. The plotted model used to compare the standard procedure (without instructions) and the presence-first procedure with instructions (Figure 2, Panels C & D), illustrates that the significant effect on choosing was not accompanied by any significant difference in discriminability. Contrary to expectations, the use of metacognitive instructions with the presence-first procedure did not lead witnesses to make significantly better use of memory to avoid false identifications than in a standard lineup. Rather, it encouraged more conservative decision-making.

In contrast, and following the pattern of results observed in Experiment 1, the provision of metacognitive instructions with the presence-first procedure meaningfully improved discriminability over participants' own earlier performance with the presence-first procedure without instructions (see Figure 2, Panels E & F). While the interaction term (lineup × target presence) was not a significant predictor of decision type, the coefficient was substantial and its CIs were biased strongly toward values falling below 0, indicating a meaningful effect. Again, the key observation from an applied perspective, is that when given instructions, presence-first witnesses' ability to differentiate guilty from innocent suspects was meaningfully better, following Cumming's (2013) approach to data interpretation, though this effect was not statistically significant. Notably, this pattern of responding was evident even when comparison of the presence-first procedure with and without instructions was within-subjects. Instructions thus had a meaningful effect, even though participants had to override the less optimal decision-making patterns established during test blocks completed without instructions. If presence-first participants had not practiced and established a more strongly familiarity-based approach to decision-making before being provided with the metacognitive instructions, instructions might have had a bigger impact. If so, our results likely underestimate the potential applied impact of metacognitive instructions. However, practice effects were also a potential threat as the provision of instructions could not sensibly be counterbalanced. If practice effects had driven the change, we would have expected to see an equivalent effect of instructions for both standard and presence-first procedures, against our predictions.

In a model comparing witnesses' performance in the standard procedure with and without the metacognitive instructions (Tables 2 & 4, and Figure 2, Panels G & H), instructions significantly reduced response bias. When filler identifications were included, there was no significant difference in discriminability. However, when filler identifications were not included, instructions reduced the likelihood that a witness would accurately identify a guilty suspect significantly more than they reduced the likelihood of false identification. That is, rather than improving discriminability, instructions reduced witnesses' ability to discriminate guilty from innocent suspects. Given the markedly lower choosing also observed following instructions, this effect on discriminability might have resulted from a more restricted scope for potential reductions in false identifications than for hits. Regardless, metacognitive instructions clearly did not significantly improve discriminability when given with a standard procedure, after witnesses had made standard lineup decisions without instructions. Rather, witnesses were substantially more reluctant to pick a lineup member as the offender after receiving instructions, regardless of target presence. As accurate identifications were reduced more than filler identifications from target-present lineups, it seems unlikely that this effect was based on greater consideration of recollection, or due to practice effects.

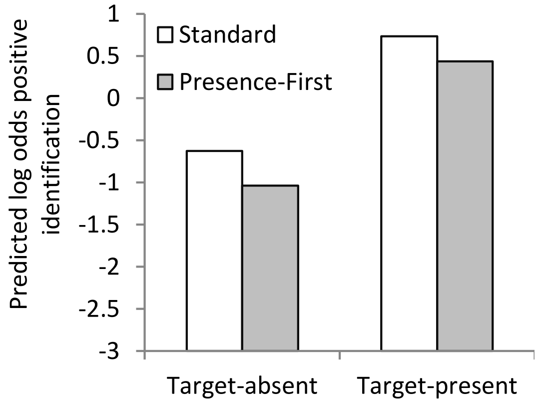
Table 3

Model Fit Statistics for Mixed-Effects Models Predicting Decision Type (Positive Identification; Lineup Rejection) by Lineup Type in Experiment 2

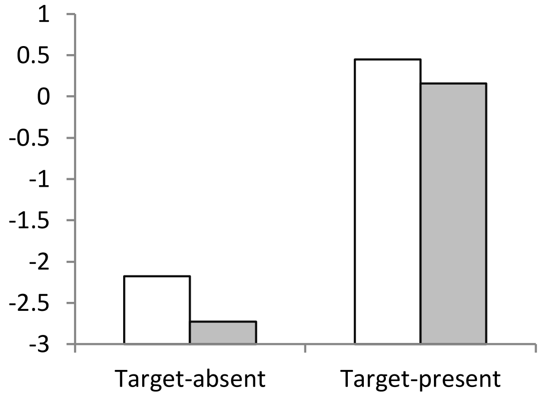
Filler Identifications	Comparison	χ^2	*df*	*p*
Included	Standard vs Presence-First (PF)	76.99	3	< .001
	Standard vs PFI	85.95	3	< .001
	Standard vs Standard Instructions (SI)	36.57	3	< .001
	PF vs PFI	56.98	3	< .001
Excluded	Standard vs PF	122.71	3	< .001
	Standard vs PFI	130.05	3	< .001
	Standard vs SI	58.75	3	< .001
	PF vs PFI	78.23	3	< .001

Number of observations, fillers included (fillers excluded): Standard vs PF = 2560 (2086), n = 64 (64); Standard vs PFI = 2560 (2131), n = 64 (64); Standard vs SI = 2560 (1870), n = 32 (32); PF vs PFI = 2560 (2179), n = 32 (32).

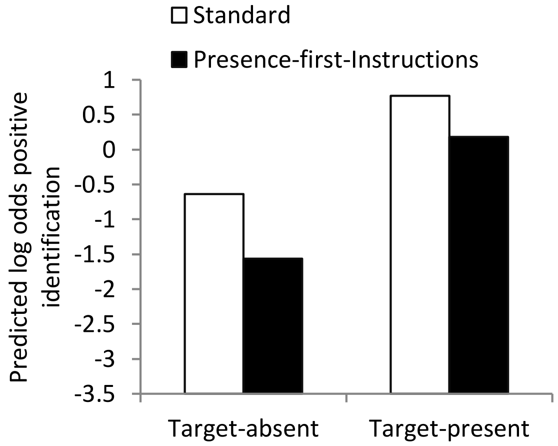
Panel A (Standard vs PF, filler IDs included)



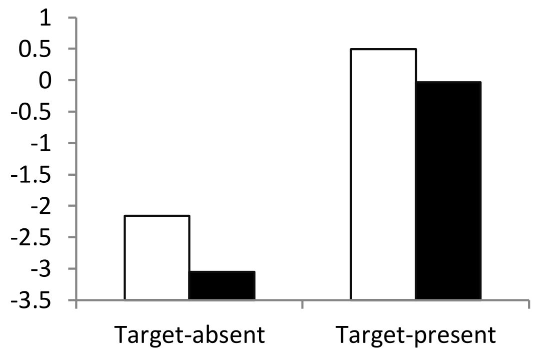
Panel B (Standard vs PF, filler IDs excluded)



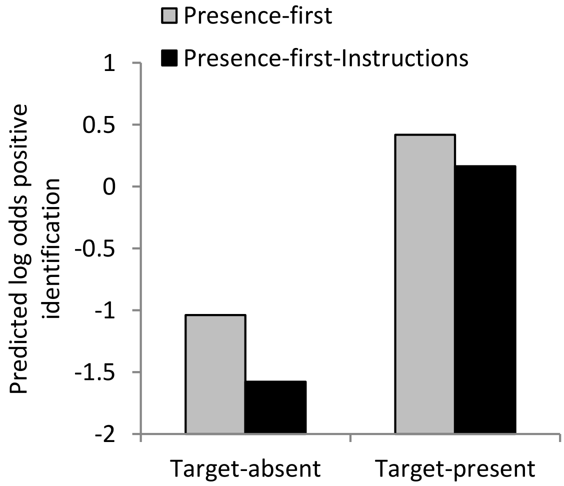
Panel C (Standard vs PFI, filler IDs included)



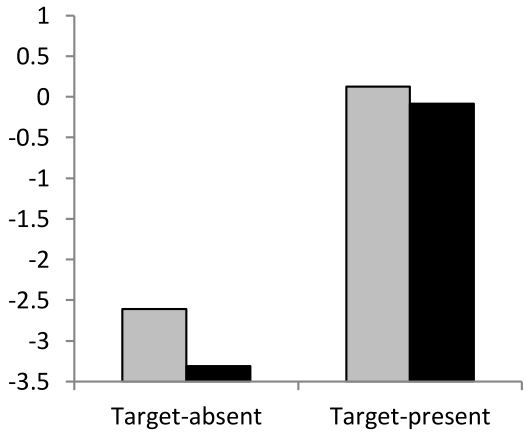
Panel D (Standard vs PFI, filler IDs excluded)



Panel E (PF vs PFI, filler IDs included)



Panel F (PF vs PFI, filler IDs excluded)



Panel G (Standard vs SI, filler IDs included) Panel H (Standard vs SI, filler IDs excluded)

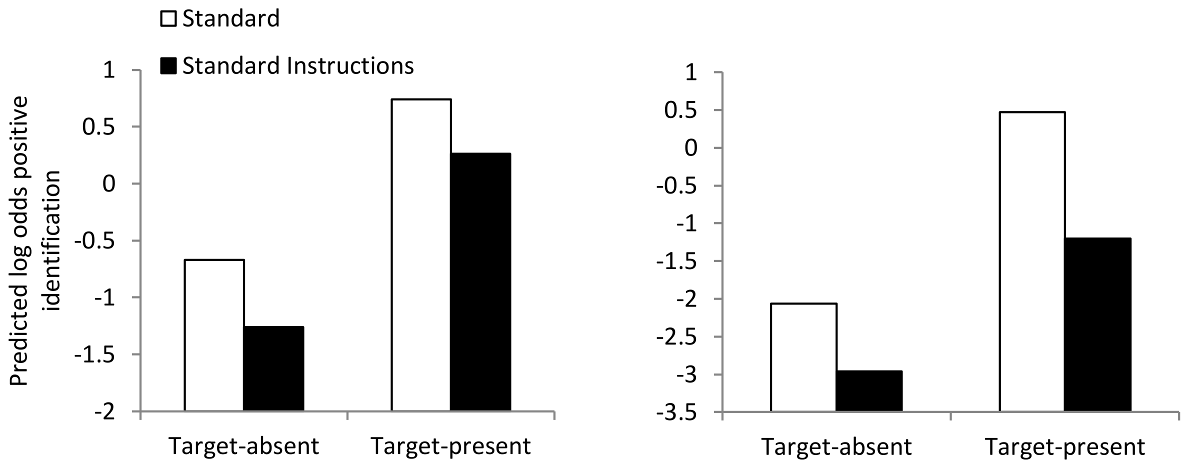


Figure 2. Plots of predicted log odds of a positive identification decision in Experiment 2 by lineup type and target presence. Values greater than 0 indicate a positive identification from the mini-lineup is more likely, and values lower than 0 indicate a lineup rejection is more likely. Coefficient standard errors, indicative of the variability of the estimates, are displayed in Table 4.

Table 4

Fixed Effect Coefficients for Mixed-Effects Models Predicting Decision Type (Lineup Rejection; Positive Identification) by Lineup Type in Experiment 2

		Lineup Comparisons											
		#Standard vs PF			#Standard vs PFI			#Standard vs SI			#PF vs PFI		
Filler IDs	Fixed Effect	*b*	SE_b	95% CI_b	*b*	SE_b	95% CI_b	*b*	SE_b	95% CI_b	*b*	SE_b	95% CI_b
Included	Intercept	-0.62			-0.64			-0.67			-1.04		
	Lineup	-0.41	0.23	-0.87, .04	-0.93*	0.27	-1.02, -0.26	-0.59*	0.21	-1.00, -0.19	-0.54*	0.16	-0.85, -0.24
	TP	1.36*	0.17	1.02, 1.69	1.41*	0.19	1.04, 1.78	1.41*	0.20	1.02, 1.80	1.46*	0.17	1.35, 1.78
	Lineup × TP	0.12	0.23	-0.34, 0.58	0.34	0.26	-0.17, 0.85	-0.11	0.23	-0.35, 0.58	0.29	0.19	-0.08, 0.65
Excluded	Intercept	-2.18			-2.16			-2.07			-2.61		
	Lineup	-0.55	0.31	-1.14, 0.05	-0.89*	0.33	-1.54, -0.24	-0.90*	0.25	-1.38, -0.40	-0.70*	0.29	-1.26, -0.14
	TP	2.63*	0.23	2.18, 3.08	2.65*	0.22	2.21, 3.09	2.54*	0.23	2.09, 2.99	2.74*	0.23	2.29, 3.19
	Lineup × TP	0.26	0.30	-0.32, 0.85	0.37	0.31	-0.24, 0.98	-0.78*	-.30	-1.36, -0.20	0.49	0.29	-0.08, 1.07

Note: * indicates significant predictors, # indicates the reference lineup condition. Number of observations, fillers included (fillers excluded): Standard vs PF = 2560 (2086), *n* = 64 (64); Standard vs PFI = 2560 (2131), *n* = 64 (64); Standard vs SI = 2560 (1870), *n* = 32 (32); PF vs PFI = 2560 (2179), *n* = 32 (32).

In sum, while metacognitive instructions reduced response bias in the standard procedure, they were not effective in improving discriminability. As regards our main research question, the pattern of effects across standard, presence-first, and presence-first-instructions procedures, replicated that observed between-subjects in Experiment 1, when instructions had included more substantial task-related elements. Importantly, these effects were clear in responses made to a more varied and difficult stimulus set that more strongly tested predicted differences between procedures.

Experiment 3

As in Experiment 1, the improvement in discriminability observed for the presence-first procedure with metacognitive instructions in Experiment 2 was meaningful, but not statistically significant. As the design provided adequate power to detect a non-trivial effect (Gelman & Hill, 2007; Hallahan & Rosenthal, 1996) we again revisited the content of instructions. Following the reasoning that guided revisions for Experiment 2, we removed all remaining references to the shared similarity of lineup members and did not advise participants to use instructions to avoid false identifications. Instructions thus included only phenomenological elements. We varied lineup type between-subjects on three levels: standard; standard-instructions (standard procedure with revised instructions); and *presence-recollection* (presence-first procedure with the revised instructions). Varying lineup between-subjects would allow us to test two key possibilities. First, that response bias was lower with instructions in Experiment 2 due to instructions test blocks being completed by participants after two full blocks without instructions, rather than to the instructions themselves. Second, that the impact of revised metacognitive instructions would be greater when mock eyewitnesses had not previously established a more heavily familiarity-based decision-making approach to test lineups.

Method

Participants. Experiment 3 included 96 participants (79 female, age: *M* = 23.27, *SD* = 8.22 years).

Materials and Procedure. We used both collections of mini-lineups. After giving informed consent, each participant completed 80 mini-lineup trials in four blocks of 20. The manipulation of lineup condition followed that in Experiment 1, with the following key differences. Lineup condition was manipulated on three levels between-subjects (standard; standard-recollection; presence-recollection). The order of mini-lineup collections was counterbalanced within lineup and instructions blocks.

The *standard* lineup condition was identical to that in Experiment 1, but for the addition of two further blocks of mini-lineup trials.

The *standard-recollection* lineup condition was identical to the standard lineup condition, but for the provision of *phenomenological instructions* before test phases. *Phenomenological instructions* were identical to those provided in Experiment 2, except for the removal of the explicit instruction to avoid similarity-based positive identification decisions (i.e., "*If you think the test photograph is similar to your memory for the studied face, but your memory is not clear, vivid or detailed, you should response Not Present.")*. To sum up, these instructions provided only phenomenological information about the qualities of remembering diagnostic of accuracy: vividness, clarity, and the recollection of specific detail.

The *presence-recollection* lineup condition combined the presence-first procedure with instructions that were identical to those provided in the *standard-instructions* lineup condition.

All participants completed the same manipulation check questionnaire as for Experiment 1.

Results & Discussion

The pattern of responses to manipulation check questionnaire items was generally consistent with participants having attended to key elements of metacognitive instructions, when provided (descriptive statistics and pairwise comparisons are reported in the online supplement). Comparison of the standard and presence-recollection conditions revealed that lineup, target presence and the target presence × lineup interaction term were all significant predictors of decision type (Tables 5 & 6). As the plotted model shows (see Figure 3, Panels A & B), witnesses in the presence-recollection procedure demonstrated significantly better discriminability than those in the standard procedure. Presence-recollection participants were significantly less likely to falsely identify an innocent suspect ($odds_{PR} = 0.33$; $odds_{Standard} = 0.67$), without a significant difference in the likelihood of accurate identifications ($odds_{PR} = 1.70$; $odds_{Standard} = 2.23$). As predicted, witnesses made significantly better use of memorial evidence when we provided metacognitive instructions that included only phenomenological information, with a presence-first decision structure. Compared with a standard procedure, the presence-recollection procedure produced significantly lower rates of false identification decisions, without a significant reduction in hits.

In stark contrast, comparison of standard and standard-instructions conditions revealed no significant differences in identification performance (see Tables 5 & 6). Figure 3 (Panels C & D) shows that providing the revised metacognitive instructions to participants in a standard lineup had no significant effect on either response bias or discriminability as compared with the standard procedure without instructions (for false identifications, $OR_{Standard:SI} = 1.00$; for accurate identifications, $OR_{Standard:SI} = 0.96$).

Table 5

Model Fit Statistics for Mixed-Effects Model Predicting Decision Type (Positive Identification; Lineup Rejection) by Lineup Type in Experiment 3

Filler Identifications	Comparison	χ^2	*df*	*p*
Included	Standard vs Presence-Recollection (PR)	93.94	3	< .001
	Standard vs SI	80.67	3	< .001
Excluded	Standard vs PR	55.07	3	< .001
	Standard vs SI	38.97	3	< .001

Number of observations, fillers included (fillers excluded): Standard vs PR = 5120 (4727), *n* = 64 (64); Standard vs SI = 5120 (4669), *n* = 64 (64)

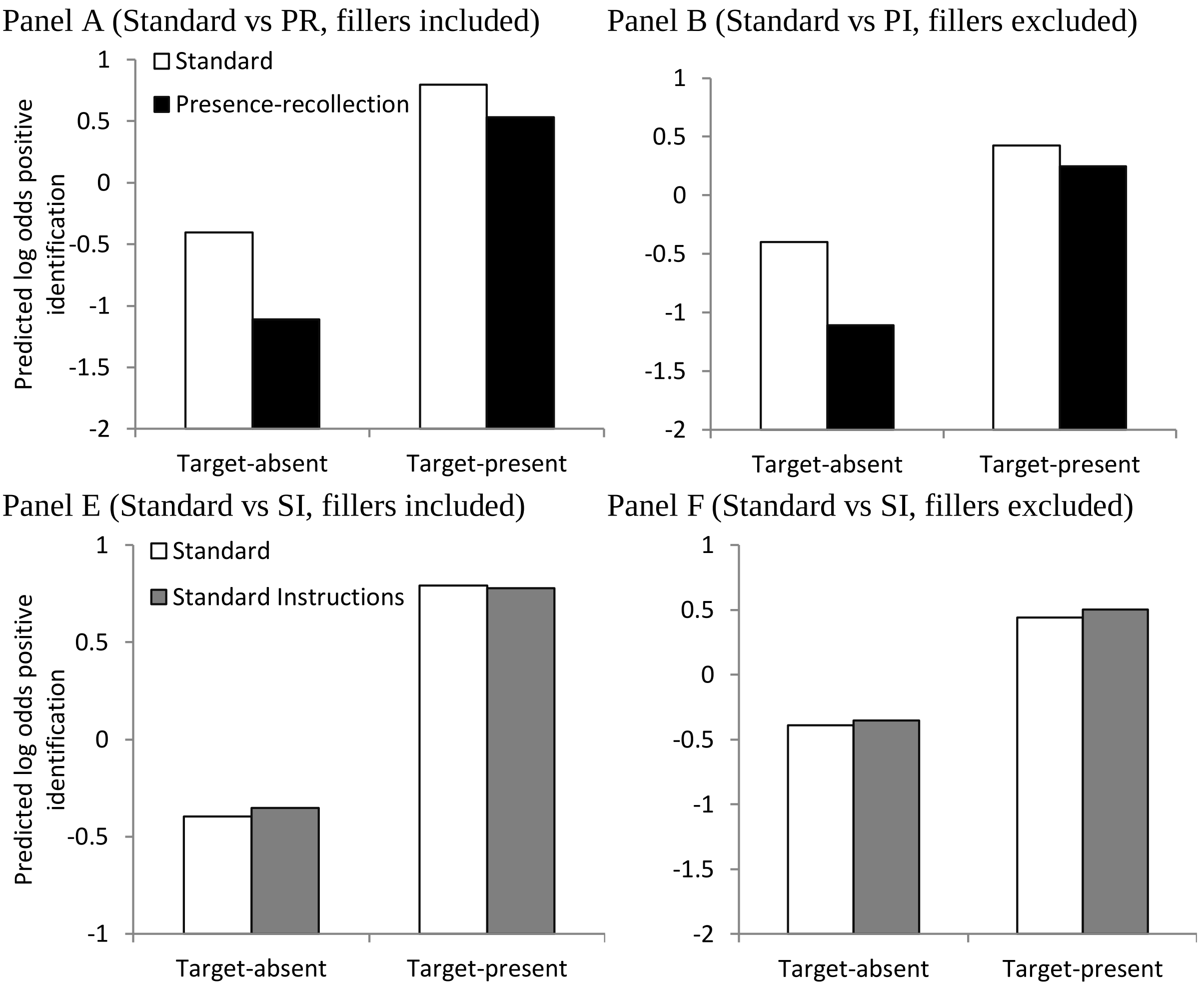


Figure 3. Plots of predicted log odds of a positive identification decision in Experiment 3 by lineup type and target presence. Values greater than 0 indicate a positive identification from the mini-lineup is more likely, and values lower than 0 indicate a lineup rejection is more likely. Coefficient standard errors, indicative of the variability of the estimates, are displayed in Table 6.

Table 6

Fixed Effect Coefficients for Mixed Effects Model Predicting Decision Type (Lineup Rejection; Positive Identification) by Lineup Type in Experiment 3

		Lineup Comparisons					
		#Standard vs PR			#Standard vs SI		
Filler IDs	Fixed Effect	*b*	SE_b	95% CI_b	*b*	SE_b	95% CI_b
Included	Intercept	-0.40			-0.40		
	Lineup	-0.71*	0.21	-1.11, -0.30	0.00	0.20	-0.39, 0.39
	TP	1.20*	0.16	0.89, 1.50	1.19*	0.14	0.91, 1.46
	Lineup × TP	0.44*	0.20	0.06, 0.83	0.04	0.17	-0.29, 0.38
Excluded	Intercept	-0.40			-0.39		
	Lineup	*-0.71	0.21	-1.11, -0.30	-0.00	0.20	-0.39, -0.39
	TP	*0.83	0.18	0.47, 1.18	0.83	0.16	0.52, 1.14
	Lineup × TP	*0.53	0.23	0.08, 0.98	0.06	0.21	-0.34, 0.48

Note: * indicates significant predictors, # indicates the reference lineup condition. Number of observations, fillers included (fillers excluded): Standard vs PR = 5120 (4727), *n* = 64 (64); Standard vs SI = 5120 (4669), *n* = 64 (64).

Summary of Results for Experiments 1-3

Experiments 3 results demonstrated that phenomenological instructions in a presence-first procedure enabled eyewitnesses to distinguish between guilty and innocent suspects significantly more effectively than in a standard simultaneous lineup procedure. These results are consistent with our two-pronged best-match-first dual-process account of memory and decision processes in standard simultaneous lineup tasks (N. Guerin et al., 2016). Specifically, witnesses following a standard simultaneous procedure tend to adopt best-match-first decision-making that encourages an over-reliance on familiarity and neglect of recollection. This account predicts that task structures and procedure that disrupt best-match-first decision-making and encourage the use of recollection, will improve the reliability of eyewitness identification decisions. The *presence-recollection* procedure tested in Experiment 3 is one example of this kind of procedure, and results supported our predictions for greater discriminability than in a standard lineup procedure. Importantly, and contrary to ample evidence from recognition memory paradigms (e.g., Blank & Launay, 2014; Gallo, 2010; Starns et al., 2007), results suggested that task-related elements of instructions that drew attention to the mechanics of the simultaneous lineup tended to worked against the positive effect of enhanced phenomenological awareness, to some degree (e.g., evident in the somewhat smaller effect size for the effect of instructions in Experiments 1 & 2, compared with that in Experiment 3). Specifically, greater metacognitive awareness of lineup composition appears to have hindered the disruption of best-match-first decision-making otherwise evident for phenomenological instructions with the presence-first decision structure.

It is also important to note that results clearly show that neither presence-first structural elements nor phenomenological instructions alone (i.e., when used with a standard simultaneous procedure) are sufficient to effect greater discriminability. Only in combination did these modifications effectively reframe mock witnesses' metacognitive approach to a simultaneous lineup, and counter best-match-first decision-making and neglect of recollection. The presence-first procedure thus discouraged best-match-first decision-making, even when witnesses were aware they may later need to pick an offender from the lineup, but only when witnesses were also cued to attend to the phenomenology of retrieval. Likewise, phenomenological instructions effectively enabled eyewitnesses to make better use of memorial evidence related to the offender, but only when delivered with a presence-first procedure.

A subsidiary point is that the pattern of effects across all three experiments revealed that when metacognitive instructions included task-related elements, witnesses' overall willingness to make a positive identification was affected. When task-related elements were included in instructions for either standard or presence-first procedures, witnesses were discouraged from choosing from the lineup (significantly lower response bias in Experiments 1 and 2). However, when these elements of metacognitive instructions were removed in Experiment 3, response bias was not significantly lower for targets in either the standard procedure (and no significant effect on discriminability) or the presence-first procedure (with greater discriminability driven by significantly less false identification). Despite the serious consequences of false identification, witnesses' tendency to adopt an overly lenient response criterion in lineup tasks is well-documented (e.g., Brewer & Wells, 2006; Malpass & Devine, 1981). Therefore, task-related instructions that reliably reduce eyewitnesses' overall tendency to choose are potentially valuable in themselves, even without better discriminability.

Experiment 4

In Experiments 1-3, we tested metacognitive instructions with standard and modified presence-first procedures. We used Experiment 4 to test two key ideas about the mechanisms contributing to the effectiveness of metacognitive instructions. First, we aimed to evaluate our proposal that metacognitive instructions would not be effective when given to witnesses in procedures that do not structurally disrupt best-match-first processes. Second, we wanted to test the possibility that instructions would effectively improve recollection, even with a best-match-first approach to lineup decisions, if the two component decisions were separated in two explicit response steps (i.e., best-match, then offender-presence). Therefore, we tested phenomenological instructions with a modified best-match-first procedure in which best-match and offender-presence judgments were made in two steps. We compared a standard lineup procedure with the two-step best-match-first procedure, with and without phenomenological instructions, in a fully between-subjects design. If, contrary to expectations, separating the two component decisions allowed participants to use phenomenological instructions to improve the memorial basis of their identification decisions, we expected to see greater discriminability when provided with a best-match-first procedure than for a standard simultaneous procedure (without instructions).

Method

Participants. A further 112 students (87 female, age: M = 23.13, SD = 6.99 years) completed Experiment 4.

Materials and procedure. We used the original collection of 40 mini-lineups in this experiment. The manipulation of lineup condition followed that in Experiment 1, with the following differences. Lineup condition was manipulated on three levels between-subjects (standard; best-match-first; best-match-recollection).

The *standard* lineup condition was identical to that in Experiment 1.

The *best-match-first* lineup condition followed that in previous studies testing lineup decision order (N. Guerin et al., 2016). Like the presence-first procedures, the best-match-first condition also required two steps of decision-making: a *best-match* decision; and a *target presence* decision. However, the two identification decisions were not blocked. That is, participants made two decisions for each lineup, before proceeding to the next lineup. For the first decision (best-match), participants were asked to click on the lineup member that best matched the studied face. Once a best-match was selected, the remaining faces in the mini-lineup disappeared from the screen, leaving the best-match photo on-screen for an additional 1.5 s. For the second decision, participants indicated whether or not their displayed best-match selection was the studied target (*Yes; No)*.

The *best-match-recollection* lineup condition was identical to the *best-match-first* lineup condition, except that before each test phase participants were also given the *phenomenological instructions* used in Experiment 3.

All participants completed the same manipulation check questionnaire as for Experiment 1.

Results & Discussion

We first compared the standard and best-match-first procedures (See Tables 7 & 8; Figure 4, Panels A & B). Neither lineup nor the target presence × lineup interaction term was a significant predictor of decision type, indicating no significant differences between procedures in either response bias or discriminability (for false identifications, $OR_{Standard:BMF} = 1.05$; for hits, $OR_{Standard:BMF} = 1.05$). This finding replicates our previous results (N. Guerin et al., *Experiment 1*), further supporting the conclusion that witnesses follow best-match-first decision-making processes in both the standard and the best-match-first procedure.

Results from pairwise comparison of the standard and best-match-first-instructions conditions supported our expectation that there would be no evidence of an effect of instructions when provided with the best-match-first procedure. The pattern of responses to manipulation check questionnaire items was generally consistent with participants having attended to key elements of metacognitive instructions, when provided (descriptive statistics and pairwise comparisons are reported in the online supplement). The model comparing standard and best-match-first-instructions procedures (see Tables 7 & 8; Figure 4, Panels C & D) similarly showed no significant effect of lineup on response bias or discriminability (for false identifications, $OR_{Standard:BMFI} = 0.90$; for hits, $OR_{Standard: BMFI} = 1.44$). Likewise, the model comparing best-match-first procedures with and without instructions showed no significant difference in response bias or discriminability (for false identifications, $OR_{BMF:BMFI} = 0.87$; for hits, $OR_{BMF:BMFI} = 0.98$) with instructions (Tables 7 & 8; Figure 4, Panels E & F). Effect size indices suggest a small detriment to discriminability, if anything, in the best-match-first-instructions condition. In sum, no significant differences in choosing or discriminability were found between any of the procedures (see Table 9 for aggregate descriptive statistics). Results provide no evidence that splitting component decisions in a modified best-match-first procedure would allow participants to make use of phenomenological instructions to improve identification performance. In conclusion, when provided with procedures that cue best-match-first decision-making, there was no evidence that metacognitive instructions affect either choosing or discriminability.

Table 7

Model Fit Statistics for Mixed-Effects Model Predicting Decision Type (Positive Identification; Lineup Rejection) by Lineup Type in Experiment 4

Filler Identifications	Comparison	χ^2	*df*	*p*
Included	Standard vs BMF	73.24	3	< .001
	Standard vs BMFI	55.9	3	< .001
	BMF vs BMFI	52.58	3	< .001
Excluded	Standard vs BMF	89.74	3	< .001
	Standard vs BMFI	76.64	3	< .001
	BMF vs BMFI	83.59	3	< .001

Number of observations, fillers included (fillers excluded): Standard vs BMF = 2220 (1809), *n* = 78 (78); Standard vs BMFI = 2260 (1831), *n* = 79 (79); BMF vs BMFI = 2680 (2204), *n* = 67 (67).

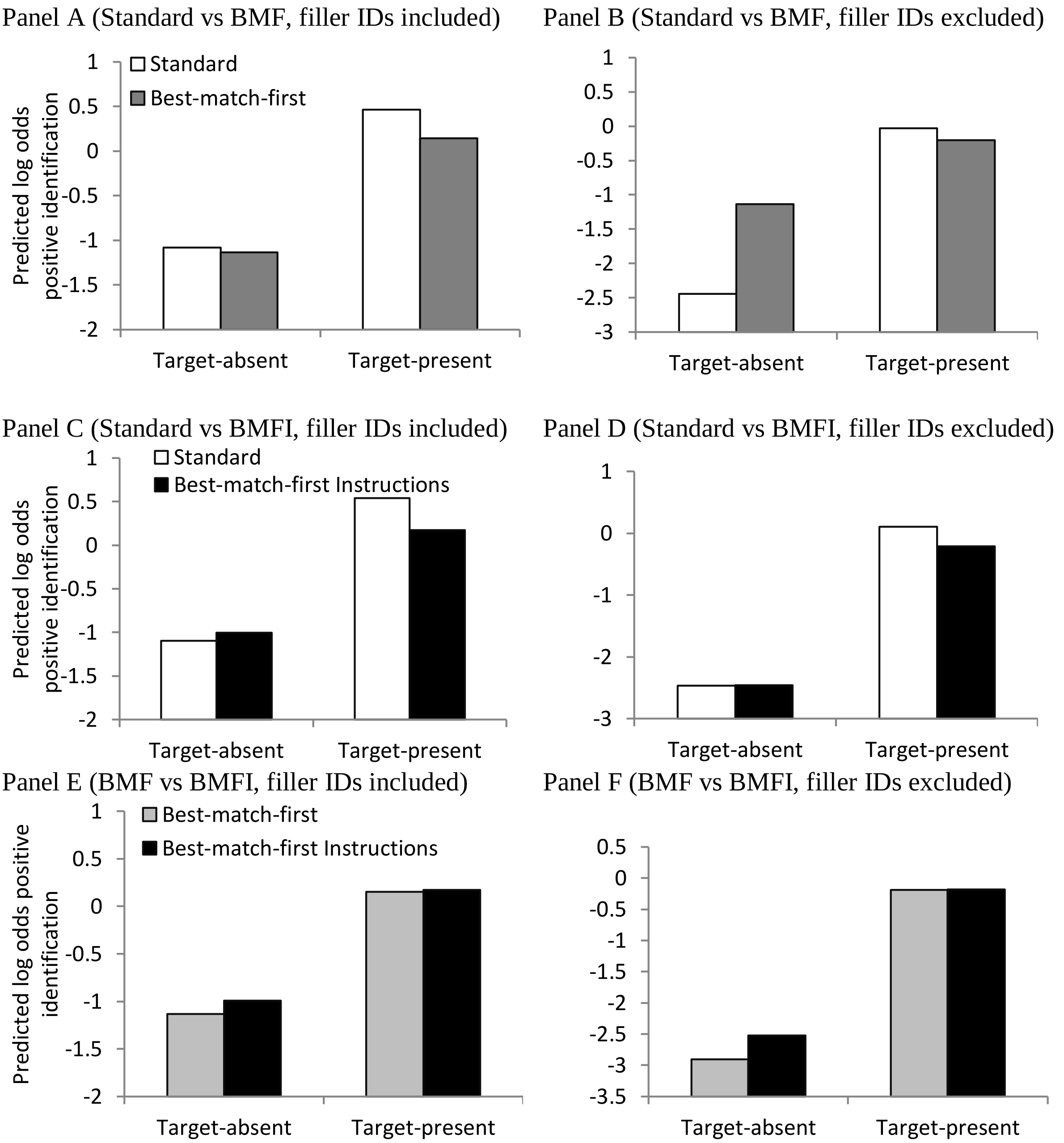


Figure 4. Plots of predicted log odds of a positive identification decision in Experiment 4 by lineup type and target presence. Values greater than 0 indicate a positive identification from the mini-lineup is more likely, and values lower than 0 indicate a lineup rejection is more likely. Coefficient standard errors, indicative of the variability of the estimates, are displayed in Table 8.

Table 8

Fixed Effect Coefficients for Mixed Effects Model Predicting Decision Type (Lineup Rejection; Positive Identification) by Lineup Type in Experiment 4

			Lineup Comparisons								
			#Standard vs BMFI			#BMF vs BMFI			#Standard vs BMF		
Filler IDs	Effect		*b*	SE_b	95% CI_b	*b*	SE_b	95% CI_b	*B*	SE_b	95% CI_b
Included	Fixed	Intercept	-1.10			-1.13			-1.08		
		Lineup	0.10	0.27	-0.43, 0.63	0.14	0.28	-0.41, 0.69	-0.05	0.27	-0.57, 0.46
		TP	1.64*	0.23	1.19, 2.09	1.28*	0.20	0.90, 1.67	1.55*	0.19	1.17, 1.92
		Lineup × TP	-0.47	0.31	-1.07, 0.14	-0.12	0.28	-0.66, 0.42	-0.27	0.27	-0.79, 0.25
Excluded	Fixed	Intercept	-2.46			-2.91			-2.44		
		Lineup	0.01	0.33	-0.64, 0.65	0.39	0.35	-0.30, 1.07	-0.55	0.38	-1.29, 0.19
		TP	2.57*	0.34	1.91, 3.22	2.72*	0.30	2.14, 3.30	2.41*	0.30	1.82, 3.00
		Lineup × TP	-0.32	0.44	-1.19, 0.54	-0.38	0.39	-1.16, 0.39	0.38	0.42	-0.44, 1.20

Note: * indicates significant predictors, # indicates the reference lineup condition. Number of observations, fillers included (fillers excluded): Standard vs BMF = 2220 (1809), *n* = 78 (78); Standard vs BMFI = 2260 (1831), *n* = 79 (79); BMF vs BMFI = 2680 (2204), *n* = 67 (67).

Table 9

Descriptive Statistics for Identification Responses (Hit rates, Filler ID Rates, Correct Rejection Rates and False ID Rates) by Lineup Type for Experiments 1-4

			Response Proportion							
			Target-present				Target-absent			
Expt	Lineup Type	Statistic	Hits	Filler IDs (TP)	Misses	*n*	Correct Rejections	Filler IDs (TA)	False IDs	*n*
1	Standard	*M*	.65	.12	.23	30	.64	.24	.13	30
		SD	.16	.11	.14		.24	.16	.11	
		95% CI	.64, .66	.11, .13	.22, .24		.62, .65	.22, .25	.12, .14	
	PF	*M*	.50	.18	.32	30	.69	.19	.12	30
		SD	.19	.13	.13		.20	.15	.09	
		95% CI	.48, .52	.17, .19	.31, .33		.67, .70	.18, .21	.11, .13	
	PFI	*M*	.57	.10	.33	33	.79	.13	.08	33
		SD	.19	.11	.17		.16	.14	.07	
		95% CI	.56, .59	.09, .11	.31, .34		.78, .80	.12, .14	.08, .09	

2	Standard	*M*	.51	.15	.34	32	.63	.26	.11	32
		SD	.19	.10	.18		.21	.17	.09	
		95% CI	.50, .53	.14, .16	.32, .35		.61, .65	.24, .27	.10, .12	
	SI	*M*	.13	.43	.44	32	.70	.24	.06	32
		SD	.09	.21	.24		.25	.22	.06	
		95% CI	.12, .14	.41, .45	.42, .46		.68, .72	.22, .26	.05, .06	
	PF	*M*	.47	.13	.40	32	.71	.21	.08	32
		SD	.15	.09	.16		.15	.13	.07	
		95% CI	.46, .48	.12, .13	.39, .41		.70, .73	.20, .22	.07, .08	
	PFI	*M*	.45	.10	.45	32	.79	.16	.05	32
		SD	.19	.11	.21		.17	.14	.05	
		95% CI	.43, .46	.09, .11	.44, .47		.77, .80	.15, .17	.05, .06	
3	Standard	*M*	.49	.18	.32	32	.59	.30	.12	32
		SD	.15	.12	.16		.17	.14	.07	

		95% CI	.48, .50	.18, .19	.32, .33		.58, .59	.29, .33	.12, .12	
	PFI	*M*	.50	.12	.37	32	.72	.19	.09	32
		SD	.14	.09	.14		.16	.11	.06	
		95% CI	.50, .51	.12, .13	.37, .38		.71, .73	.18, .19	.09, .09	
	SI	*M*	.52	.17	.31	32	.58	.29	.12	32
		SD	.13	.14	.11		.17	.14	.08	
		95% CI	.51, .53	.16, .18	.31, .32		.58, .59	.29, .33	.12, .13	
4	Standard	*M*	.43	.16	.40	45	.73	.20	.07	45
		SD	.17	.14	.15		.20	.16	.07	
		95% CI	.42, .44	.15, .17	.39, .41		.72, .75	.19, .21	.07, .08	
	BMF	*M*	.39	.13	.48	33	.71	.21	.07	33
		SD	.17	.13	.19		.20	.15	.07	
		95% CI	.38, .40	.12, .14	.46, .49		.70, .73	.20, .22	.07, .08	
	BMFI	*M*	.40	.14	.46	34	.70	.23	.08	34
		SD	.19	.15	.19		.20	.17	.08	

95% CI	.39, .41	.13, .15	.45, .48		.68, .71	.21, .24	.07, .09

General Discussion

The potential benefits of theoretically-grounded investigation of factors affecting eyewitnesses' memory retrieval and decision-making during an identification test go beyond increasing understanding of the limits to identification performance, to refining our efforts curtailing those limits. Here, guided by basic recognition research, we focused on the key effects of task structure and eyewitness metamemory. Our best-match-first dual-process framework posits that eyewitnesses' predisposition to choosing from a simultaneous lineup triggers an over-reliance on familiarity-based memory, and a neglect of potentially disconfirming recollective detail. Across four experiments, we found evidence of the efficacy of two procedural interventions to improve memory and decision processes in simultaneous lineups, when used in conjunction. Specifically, 1) metacognitive instructions designed to direct eyewitness attention to qualities of retrieved memories that would indicate their likely accuracy (i.e., by indexing the contribution of recollection); and 2) a presence-first decision structure designed to disrupt best-match-first decision-making. Almost every previous demonstration of better identification performance with adult eyewitnesses has been linked to a conservative criterion shift (Clark, 2012). The modified presence-recollection procedure improved discriminability itself, rather than simply reducing the likelihood of false identification by reducing the overall likelihood of witnesses making a positive identification.

Our results provide the first demonstration of the efficacy of metacognitive intervention in an identification lineup task. Consistent with demonstrations from other memory tasks (e.g., Blank & Launay, 2014; Gallo, 2010; Lane et al., 2007; Starns et al., 2007), participants used the metamemorial information provided to more effectively differentiate between guilty and innocent suspects. This research supports the proposition that inaccurate, or insufficient, metacognitive understanding of how memory works contributes to identification errors, just as it contributes to errors in other recognition tasks (e.g., Lampinen & Odegard, 2004; Mather et al., 1997; Starns et al., 2007). Two aspects of our findings are particularly important for identification tasks. First, we showed that eyewitness accuracy could be enhanced with simple metacognitive instructions. Second, and most crucially, discriminability was significantly improved, rather than witnesses simply becoming more conservative and adopting a stricter criterion for positively identifying a lineup member as the offender.

However, significantly greater discriminability was only observed when metacognitive instructions focused exclusively on increasing participants' awareness of the phenomenology of memory accuracy (i.e., relatively greater vividness, clarity, and detail). In Experiments 1 and 2, instructions also included task-related information about the structure and composition of lineups, and how and why errors are frequently made in these tasks. With the inclusion of task-related elements, patterns of choosing were more conservative regardless of decision structure (i.e., in both standard and modified simultaneous procedures), but without significantly greater discriminability. Also, metacognitive instructions produced greater discriminability only when provided with a presence-first decision structure. When phenomenological instructions were provided with standard simultaneous lineups, or the modified best-match-first procedure, no significant or meaningfully-sized improvements in identification performance were evident.

These key findings point to an important conclusion about eyewitness decision-making for simultaneous lineups: best-match-first decision-making is characteristic, and entails an over-reliance on global familiarity that is not easy to over-ride. First, when task-related information directed eyewitnesses' attention to lineup mechanics and highlighted potential errors, they were less able to make effective use of phenomenological information. It seems likely that enhanced knowledge and salience of suspect-centric methods of lineup composition directed participants' attention to finding the suspect among similar foils, and then (more cautiously) evaluating the likely guilt of the suspect. In other words, task-related instructions facilitated the very best-match-first decision-making approach we aimed to disrupt. If so, their cueing of familiarity-based decision-making would have been in direct opposition to the action of phenomenological instructions intended to increase retrieval and weighting of recollection. Non-naïve participants had previous experience with standard laboratory lineup tasks (i.e., characterised by best-match-first decision-making), and were likely to carry typical preconceptions about the purpose of identification tasks (i.e., to identify a likely-guilty suspect). In retrospect, then, it is not surprising that best-match-first decision-making was less effectively disrupted by the conflicting metacognitive instructions provided in the first two experiments.

Second, participants were not able to use phenomenological information to better evaluate match in standard or modified best-match-first procedures (Experiments 3 and 4) when best-match-first decision-making was not disrupted. Specifically, mock eyewitnesses did not better evaluate potentially misleading familiarity-based evidence of positive match between innocent lineup members and memory for the offender, to successfully avoid false identifications. In sum, metacognitive instructions for more effective use of available memory were only effective in an identification lineup task when best-match-first decision-making was not facilitated by structural cues. This finding thus also supports the idea that best-match-first decision-making works against effective use of recollection.

Two possible mechanisms could explain the ineffectiveness of phenomenological instructions when combined with best-match-first decision-making. First, if participants were unable to attend to all of the information in instructions when they included both task-related and phenomenological elements, task-related elements that were consistent with pre-existing understandings might have received preferential attention (e.g., Kahnemann, 2011). Second, even if attended to, participants might have been unable to apply phenomenological information if best-match-first decision processes prevented adequate retrieval of recollection. That is, if the effect of phenomenological instructions relied on enhanced recall-to-reject (Rotello & Heit, 2000) via post-retrieval monitoring of critical recollection, application to familiarity-based memories might have been fruitless.

Our results do not allow us to specify the retrieval and post-retrieval mechanisms operative in the presence-recollection procedure. However, this question has important implications for the usefulness of instructions under varying witnessing and test conditions (cf. Clark & Godfrey, 2009). Using continuous dual-process theory (e.g., Wixted & Mickes, 2010), we argued that phenomenological information would allow eyewitnesses to more accurately evaluate familiarity and recollection retrieved within a presence-first decision structure. The presence-first decision structure would facilitate recall-to-reject, and phenomenological instructions might: 1) further cue more effortful recollection; and/or 2) increase accuracy for monitoring the match of lineup members. On this view, the effectiveness of the presence-recollection procedure would depend on the specific post-retrieval monitoring processes that were affected. For example, if enhanced monitoring relied on the presence of recollective detail (e.g., S. A. Guerin et al., 2012), its efficacy would be limited when encoding or retention conditions were poor. Alternatively, monitoring mechanisms that do not rely to the same extent on substantial recollection might be enhanced. For example, use of the distinctiveness heuristic (Schacter, Israel & Racine, 1999) to compare retrieved information with that which might be expected to be cued by the presence of the offender, might be equally effective in the absence or presence of recollection. We argue that just as multiple mechanisms are likely to contribute to recognition or identification errors (e.g., Clark & Godfrey, 2009; Gallo, 2010; Odegard & Lampinen, 2006), multiple mechanisms might potentially contribute to their redress. Accordingly, the mechanisms triggered by the presence-recollection procedure are likely to vary across offenders and lineups, and even between participants. Such questions are clearly an important focus for future investigation.

In addition, we found evidence that a presence-first decision structure (without phenomenological instructions) would not effectively divert non-naïve participants from best-match-first decision processes. Thus, the greater discriminability previously observed for a presence-first procedure with naïve participants was not replicated (Guerin et al., 2016). We conclude that a presence-first decision structure alone cannot effectively counter established biases to familiarity-based best-match-first decision-making. However, when combined with phenomenological information that highlighted the importance of weighing recollective qualities of memory more heavily than global familiarity, the presence-first structure allowed participants to use this metamemorial information. Participants could better discriminate between diagnostic (recollection-based) and misleading (familiarity-based) evidence of positive match.

By extension, this finding suggests that sequential presentation of lineup members would be unlikely to disrupt eyewitnesses' bias to familiarity-based identification responses. Choosing from a sequential lineup is frequently more conservative than for a simultaneous lineup, when eyewitnesses are less able to evaluate the degree of match to memory for the offender of a lineup member in the context of other potential matches, but without greater discriminability (Clark & Godfrey, 2009). We found no evidence that chunking the identification task was sufficient to encourage greater recollection, even when non-standard identification responses were required. Therefore, phenomenological instructions might also strengthen the memorial basis of eyewitness responses in sequential procedures when either binary yes/no decisions or match confidence ratings (e.g., Brewer et al., 2012) are given for each lineup member.

In practice, results for the presence-recollection lineup have immediate applied implications. With two straightforward procedural tweaks, the presence-recollection lineup procedure offers the possibility of a reduction in false identification, without a concomitant loss of accurate identification. However, testing across a range of witnessing and test variables is needed. In particular, the apparent strength of witness preconceptions about the purpose of lineup tasks in the mini-lineup paradigm suggests that participants' pre-existing metacognitions might be more highly activated and salient in a more realistic identification paradigm. Therefore, it will be important to rigorously test the presence-recollection lineup, or any novel procedure informed by these findings, in a more realistic identification paradigm across diverse suspect, encoding and retention conditions to evaluate generalisability. Similarly, external contextual factors that might cue best-match-first decision-making, such as administrator or co-witness bias, should also be a prime focus of tests of the reliability of the presence-recollection procedure. A novel identification procedure that relies on metacognitive factors for its efficacy will be of limited value if rendered ineffective by common counteracting influences on metamemorial processes.

Our results also have important theoretical implications, both for the development of eyewitness identification theory, and for the more general literature on recognition memory. In terms of advancing theoretical models of identification tasks, findings are consistent with the two ideas underlying our best-match-first dual-process framework for the analysis of simultaneous lineup performance: that eyewitnesses tend to take a best-match-first approach to simultaneous lineups, and that this leads to a neglect of critical recollection. More importantly, these results demonstrate that structural and metacognitive interventions in lineup tasks can improve the memorial basis of eyewitnesses' identification decisions. Of particular note, the greater discriminability observed in the presence-recollection procedure was observed without a significant reduction in the likelihood of a hit. These effects demonstrate that eyewitnesses can make better use of available memory, despite ingrained preconceptions about the task and metamemorial beliefs that typically encourage serious identification errors. As noted, this kind of improvement might also be possible in other lineup formats that require a different set of responses to lineup members, and thus entail different decision-making strategies (e.g., sequential lineup procedures, or confidence rating procedures, see Brewer et al., 2012; Sauer et al., 2008). Thus, while the presence-recollection lineup procedure has potential applied utility in itself, our findings highlight the potential value of further investigating decision and memory processes, and the role of metacognition, in other eyewitness identification tasks. Overall, this research reaffirms the value of extending theoretical understandings derived from basic laboratory tests to more complex memory tasks with serious consequences in applied contexts.

In terms of theoretical research into recognition and metamemory more broadly, our results also contribute to the literature investigating the operation of basic processes across a variety of laboratory and applied tasks. First, our findings go beyond previous findings about the efficacy of metacognitive instructions to reduce false recognition in demonstrating that more, and more accurate, information is not always better (cf. Lane et al., 2008). For example, mock eyewitnesses in our experiments were more conservative choosers when given specific and accurate information about lineup construction and likely errors, leading to lower rates of false identification. However, while patterns of discriminability were in the predicted direction, there was no significant improvement of discriminability. In contrast, simple phenomenological instructions were clearly effective (i.e., when provided without task-related elements). Blank and Launay (2014) suggested the effectiveness of task-related instructions would increase with their degree of specificity. However, our results indicate that seemingly neutral, specific task-related information (e.g., about the method used to select filler lineup members) can limit, or even counter, the potential benefits of greater task understanding. A critical factor is likely to be the extent to which task-related information inadvertently promotes pre-existing biases to error-prone decision-making mechanisms, rather than disrupting them.

Second, we have demonstrated that the effect of metacognitive instructions will be moderated by the type of responses participants are asked to provide. For recognition of a previously-seen stimulus, the effect of phenomenological instructions on eyewitness accuracy depended on the questions asked. In our experiments, greater accuracy with instructions was gained simply by asking witnesses to indicate first only whether or not the offender was present in a lineup, even when there was no change in how the test lineup was presented. Clearly, different questions trigger critically different decision-making processes. But further, the accessibility of memorial evidence to conscious metamemorial monitoring and control also depends on the decision-making approach adopted. Potentially, then, recognition performance in other tasks might similarly benefit by changing the recognition responses people are required to give.

Conclusions. Results were consistent with our best-match-first dual-process account of simultaneous lineup decision-making. We effectively reduced false identification and increased discriminability with a novel presence-recollection lineup procedure that: 1) enhanced metacognitive awareness of the diagnostic phenomenology of accurate memories; 2) disrupted eyewitnesses' typical over-reliance on global familiarity. The effectiveness of the presence-recollection lineup demonstrates the usefulness of insights from basic recognition memory research for explaining eyewitness memory use in identification tasks, and its potential improvement (e.g., Blank & Launay, 2014; Gallo et al., 2010; Johnson et al., 1993; Odegard & Lampinen, 2006). In addition, we demonstrated that providing more accurate metacognitive information could nevertheless trigger potentially error-prone decision-making processes. Thus, the effectiveness of metacognitive instructions depended on how their content related to participants' pre-existing understandings of the task. Finally, we demonstrated that neither a presence-first structure nor phenomenological instructions independently effected the greater discriminability that was evident when these two lineup innovations were implemented together. Hence, structure and instructions must work in concert, if eyewitness identification performance is to be improved by enhancing the accuracy of eyewitness metacognition.

References

Alban, M. W., & Kelley, C. M. (2012). Variations in constrained retrieval. *Memory & Cognition*, *40*, 681-692. doi: 10.3758/s13421-012-0185-5

Anastasi, J. S., Rhodes, M. G., & Burns, M. C. (2000). Distinguishing between memory illusions and actual memories using phenomenological measurements and explicit warnings. *The American Journal of Psychology*, *113*, 1-26. doi: 10.2307/1423458

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390-412. doi: 10.1016/j.jml.2007.12.005

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structures for confirmatory hypothesis testing: Keeping it maximal. *Journal of Memory and Language*, 68, 255-278. doi: 10.1016/j.jml.2012.11.001

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2014). _lme4: Linear mixed-effects models using Eigen and S4_. Retrieved from http://CRAN.R-project.org/package=lme4

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: when retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, *127*, 55-69. doi: 10.1037/0096-3445.127.1.55

Blank, H., & Launay, C. (2014). How to protect eyewitness memory against the misinformation effect: A meta-analysis of post-warning studies. *Journal of Applied Research in Memory and Cognition*, *3*, 77-88. doi: 10.1016/j.jarmac.2014.03.005

Bodner, G. E., & Lindsay, D. S. (2003). Remembering and knowing in context. *Journal of Memory and Language*, *48*, 563-580. doi: 10.1016/s0749-596x(02)00502-8

Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science*, *11*, 164-169. doi: 10.1111/1467-8721.00192

Brainerd, C. J., Reyna, V. F., Wright, R., & Mojardin, A. H. (2003). Recollection rejection: false-memory editing in children and adults. *Psychological Review*, *110*, 762-784. doi: 10.1037/0033-295x.110.4.762

Brewer, N., & Palmer, M. A. (2010). Eyewitness identification tests. *Legal and Criminological Psychology*, *15*, 77-96. doi: 10.1348/135532509x414765

Brewer, N., Weber, N., Wootton, D., & Lindsay, D. S. (2012). Identifying the bad guy in a lineup using confidence judgments under deadline pressure. *Psychological Science*, *23*, 1208-1214. doi: 10.1177/0956797612441217

Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied*, *12*, 11-30. doi: 10.1037/1076-898X.12.1.11

Brewer, N., & Wells, G. L. (2011). Eyewitness identification. *Current Directions in Psychological Science*, *20*, 24-27. doi: 10.1177/0963721410389169

Chambers, K. L., & Zaragoza, M. S. (2001). Intended and unintended effects of explicit warnings on eyewitness suggestibility: Evidence from source identification tests. *Memory & Cognition*, *29*, 1120-1129. doi: 10.3758/bf03206381

Charman, S. D., Wells, G. L., & Joy, S. W. (2011). The dud effect: Adding highly dissimilar fillers increases confidence in lineup identifications. *Law and Human Behavior*, *35*, 479-500. doi: 10.1007/s10979-010-9261-1

Clark, S. E. (2005). A re-examination of the effects of biased lineup instructions in eyewitness identification. *Law and Human Behavior*, *29*, 575-604. doi: 10.1007/s10979-005-7121-1

Clark, S. E. (2012). Costs and benefits of eyewitness identification reform psychological science and public policy. *Perspectives on Psychological Science*, *7*, 238-259. doi: 10.1177/1745691612439584

Clark, S. E., Howell, R. T., & Davey, S. L. (2008). Regularities in eyewitness identification. *Law and Human Behavior*, *32*, 187-218. doi: 10.1007/s10979-006-9082-4

Cook, G. I., Marsh, R. L., & Hicks, J. L. (2005). Revisiting the role of recollection in item versus forced-choice recognition memory. *Psychonomic Bulletin & Review*, *12*, 720-725. doi: 10.3758/bf03196763

Cumming, G. (2013). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York: Routledge.

Cutler, B. L., & Penrod, S. D. (1995). *Mistaken identification: The eyewitness, psychology and the law*. Cambridge University Press.

De Carlo, L. T. (1998). Signal detection theory and generalized linear models. *Psychological Methods*, *3*, 186-205. doi: 10.1037//1082-989x.3.2.186

Dodson, C. S., Holland, P. W., & Shimamura, A. P. (1998). On the recollection of specific- and partial-source information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 1121-1136. doi: 10.1037/0278-7393.24.5.1121

Dodson, C. S., & Schacter, D. L. (2002). When false recognition meets metacognition: The distinctiveness heuristic. *Journal of Memory and Language*, *46*, 782-803. doi: 10.1037/e501882009-058

Echterhoff, G., Hirst, W., & Hussy, W. (2005). How eyewitnesses resist misinformation: Social postwarnings and the monitoring of memory characteristics. *Memory & Cognition*, *33*, 770-782. doi: 10.3758/bf03193073

Fitzgerald, R. J., Price, H. L., Oriet, C., & Charman, S. D. (2013). The effect of suspect-filler similarity on eyewitness identification decisions: A meta-analysis. *Psychology, Public Policy, and Law*, *19*, 151-164. doi: 10.1037/a0030618

Gallo, D. A. (2004). Using recall to reduce false recognition: diagnostic and disqualifying monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 120-128. doi: 10.1037/0278-7393.30.1.120

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition*, *38*, 833-848. doi: 10.3758/mc.38.7.833

Gallo, D. A., Bell, D. M., Beier, J. S., & Schacter, D. L. (2006). Two types of recollection-based monitoring in younger and older adults: Recall-to-reject and the distinctiveness heuristic. *Memory*, *14*, 730-741. doi: 10.1080/09658210600648506

Gallo, D. A., Roberts, M. J., & Seamon, J. G. (1997). Remembering words not presented in lists: Can we avoid creating false memories? *Psychonomic Bulletin & Review*, *4*, 271-276. doi: 10.3758/bf03209405

Gallo, D. A., Roediger, H. L., & McDermott, K. B. (2001). Associative false recognition occurs without strategic criterion shifts. *Psychonomic Bulletin & Review*, *8*, 579-586. doi: 10.3758/bf03196194

Gardiner, J. M., & Richardson-Klavehn, A. (2000). Remembering and knowing. In E. Tulving & F. I. M. Craik (Eds.), The Oxford handbook of memory (pp. 229-244). London: Oxford University Press. doi: 10.1002/acp.822.

Gelman, A., & Hill, J. (2007). *Data Analysis using Regression and Multilevel/Hierarchical Models.* New York: Cambridge University Press.

Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness*, *5*(2), 189-211. doi: 10.1080/19345747.2011.618213

Goodsell, C. A., Gronlund, S. D., & Carlson, C. A. (2010). Exploring the sequential lineup advantage using WITNESS. *Law and Human Behavior*, *34*, 445-459. doi: 10.1007/s10979-009-9215-7

Gronlund, S. D. (2005). Sequential lineup advantage: Contributions of distinctiveness and recollection. *Applied Cognitive Psychology*, 19, 23-37. doi: 10.1002/acp.1047

Gronlund, S. D., Mickes, L., Wixted, J. T., & Clark, S. E. (2015). Conducting an eyewitness lineup: How the research got it wrong. In B. H. Ross (Ed.), The psychology of learning and motivation (Volume 63, pp. 1– 43). Waltham, MA: Academic Press. doi: 10.1016/bs.plm.2015.03003

Guerin, N., Weber, N., & Horry, R. (2016). Split Decision: An Account of Eyewitness Decision-Making in Simultaneous Lineups and a Novel Presence-First Identification Procedure. *Under review*

Guerin, S. A., Robbins, C. A., Gilmore, A. W., & Schacter, D. L. (2012). Retrieval failure contributes to gist-based false recognition. *Journal of Memory and Language*, *66*, 68-78. doi: 10.1016/j.jml.2011.07.002

Hallahan, M., & Rosenthal, R. (1996). Statistical power: Concepts, procedures, and applications. *Behaviour research and therapy*, *34*, 489-499. doi: 10.1016/0005-7967(95)00082-8

Hancock, P. J., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences*, *4*(9), 330-337. doi: 10.1016/s1364-6613(00)01519-9

Herron, J. E., & Rugg, M. D. (2003). Strategic influences on recollection in the exclusion task: Electrophysiological evidence. *Psychonomic Bulletin & Review, 10*, 703-710. doi: 10.3758/bf03196535

Horry, R., Colton, L. M., & Williamson, P. (2014). Confidence–accuracy resolution in the misinformation paradigm is influenced by the availability of source cues. *Acta Psychologica, 151*, 164-173. doi: 10.1016/j.actpsy.2014.06.006

Horry, R., Memon, A., Wright, D. B., & Milne, R. (2011). Predictors of eyewitness identification decisions from video lineups in England: a field study. *Law and Human Behavior, 36*, 257-265. doi: 1007/s10979-011-9279-z

Huff, M. J., & Bodner, G. E. (2013). When does memory monitoring succeed versus fail? Comparing item-specific and relational encoding in the DRM paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*, 1246- 1256. doi: 10.1037/a0031338

Ingram, K. M., Mickes, L., & Wixted, J. T. (2012). Recollection can be weak and familiarity can be strong. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 325-339. doi: 10.1037/e520592012-514

Johnson, M. K., Foley, M. A., Suengas, A. G., & Raye, C. L. (1988). Phenomenal characteristics of memories for perceived and imagined autobiographical events. *Journal of Experimental Psychology: General*, *117*, 371-376. doi: 10.1037//0096-3445.117.4.371

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*, 3-28. doi: 10.1037//0033-2909.114.1.3

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology, 103*(1), 54-69. doi: 10.1037/a0028347

Kahnemann, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus, & Giroux.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*, 490-517. doi: 10.1037/0033-295x.103.3.490

Lampinen, J. M., Neuschatz, J. S., & Payne, D. G. (1997). Memory illusions and consciousness: Examining the phenomenology of true and false memories. *Current Psychology, 16,* 181-224. doi: 10.1007/s12144-997-1000-5

Lampinen, J. M., Odegard, T. N., & Bullington, J. L. (2003). Qualities of memories for performed and imagined actions. *Applied Cognitive Psychology*, *17*, 881-893. doi: 10.1002/acp.916

Lampinen, J. M., Odegard, T. N., & Neuschatz, J. S. (2004). Robust recollection rejection in the memory conjunction paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30,* 332-342. doi: 10.1037/0278-7393.30.2.332

Lane, S. M., & Meissner, C. A. (2008). A ‘middle road’ approach to bridging the basic–applied divide in eyewitness identification research. *Applied Cognitive Psychology*, *22*, 779-787. doi: 10.1002/acp.1482

Lane, S. M., Roussel, C. C., Starns, J. J., Villa, D., & Alonzo, J. D. (2008). Providing information about diagnostic features at retrieval reduces false recognition. *Memory*, *16*, 836–851. http://doi: 10.1080/09658210802337340

Lane, S. M., Roussel, C. C., Villa, D., & Morita, S. K. (2007). Features and feedback: Enhancing metamnemonic knowledge at retrieval reduces source-monitoring errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 1131-1142. doi: 10.1037/0278-7393.33.6.1131

Lindsay, D. S., & Johnson, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory & Cognition*, *17*, 349-358. doi: 10.3758/bf03198473

Lindsay, R. C., & Wells, G. L. (1980). What price justice? Exploring the relationship of lineup fairness to identification accuracy. *Law and Human Behavior*, *4*, 303-313. doi: 10.1007/bf01040622

Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology*, *57*, 335-384. doi: 10.1016/j.cogpsych.2008.02.004

Malmberg, K. J., & Xu, J. (2007). On the flexibility and the fallibility of associative memory. *Memory & Cognition*, *35*, 545-556. doi: 10.3758/bf03193293

Malpass, R. S., & Devine, P. G. (1981). Eyewitness identification: Lineup instructions and the absence of the offender. *Journal of Applied Psychology*, *66*, 482-489. doi: 10.1037/0021-9010.66.4.482

Mather, M., Henkel, L. A., & Johnson, M. K. (1997). Evaluating characteristics of false memories: Remember/know judgments and memory characteristics questionnaire compared. *Memory & Cognition*, *25*, 826-837. doi: 10.3758/bf03211327

McCabe, D. P., & Smith, A. D. (2002). The effect of warnings on false memories in young and older adults. *Memory & Cognition*, *30*, 1065-1077. doi: 10.3758/bf03194324

McDermott, K. B., & Roediger, H. L. (1998). Attempting to avoid illusory memories: Robust false recognition of associates persists under conditions of explicit warnings and immediate testing. *Journal of Memory and Language*, *39*, 508-520. doi: 10.1006/jmla.1998.2582

Meissner, C. A., Tredoux, C. G., Parker, J. F., MacLin, O. H. (2005). Eyewitness decisions in simultaneous and sequential lineups: a dual-process signal detection theory analysis. *Memory & Cognition*, 33, 783-792. doi: 10.3758/bf03193074

Mickes, L. (2015). Receiver operating characteristic analysis and confidence-accuracy characteristic analysis in investigations of system variables and estimator variables that affect eyewitness memory. *Journal of Applied Research in Memory and Cognition, 4,* 83-102. doi: 10.1016/j.jarmac.2015.01.003

Murayama, K., Sakaki, M., Yan, V. X., & Smith, G. M. (2014). Type I error inflation in the traditional by-participant analysis to metamemory accuracy: A generalized mixed-effects model perspective. *Journal of Experimental Psychology: Learning Memory and Cognition, 40*(5), 1287-1306. doi: 10.1037/a0036914

Neuschatz, J. S., Benoit, G. E., & Payne, D. G. (2003). Effective warnings in the Deese-Roediger-McDermott false-memory paradigm: the role of identifiability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 35-41. doi: 10.1037/0278-7393.29.1.35

Neuschatz, J. S., Payne, D. G., Lampinen, J. M., & Toglia, M. P. (2001). Assessing the effectiveness of warnings and the phenomenological characteristics of false memories. *Memory*, *9*, 53-71. doi: 10.1037/0278-7393.29.1.35

Norman, K. A., & Schacter, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition*, *25*, 838-848. doi: 10.3758/bf03211328

Odegard, T. N., & Lampinen, T. N. (2006). Memory editing: Knowledge, criteria, and alignment. *Memory, 14,* 777-787. doi: 10.1080/09658210600648589

Palmer, M. A., & Brewer, N. (2012). Sequential lineup presentation promotes less-biased criterion setting but does not improve discriminability. *Law and Human Behavior*, *36*, 247-255. doi: 10.1037/h0093923

Palmer, M. A., Brewer, N., McKinnon, A. C., & Weber, N. (2010). Phenomenological reports diagnose accuracy of eyewitness identification decisions. *Acta Psychologica*, *133*, 137-145. doi: 10.1016/j.actpsy.2009.11.002

Palmer, M. A., Brewer, N., & Weber, N. (2010). Postidentification feedback affects subsequent eyewitness identification performance. *Journal of Experimental Psychology: Applied*, *16*, 387-398. doi: 10.1037/a0021034

Payne, D. G., Elie, C. J., Blackwell, J. M., & Neuschatz, J. S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory and Language*, *35*, 261-285. doi: 10.1006/jmla.1996.0015

Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modelling with crossed random effects and with binomial data. *Journal of Memory and Language, 59, 413-425.* doi: 10.1016/j.jml.2008.02.002

R Core Team. (2014). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and individual Differences*, *7*(1), 1-75. doi: 10.1016/1041-6080(95)90031-4

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 803-814. doi: 10.1037/0278-7393.21.4.803

Rotello, C. M., & Heit, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory & Cognition*, *28*, 907-922. doi: 10.3758/bf03209339

Sauer, J. D., Brewer, N., & Weber, N. (2008). Multiple Confidence Estimates as Indices of Eyewitness Memory. *Journal of Experimental Psychology: General,* 137, 528-547. doi: 10.1037/a0012712

Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language*, *40,* 1-24. doi: 10.1006/jmla.1998.2611

Schooler, J. W., Gerhard, D., & Loftus, E. F. (1986). Qualities of the unreal. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 171-181. doi: 10.1037/0278-7393.12.2.171

Selmeczy, D., & Dobbins, I. G. (2014). Relating the content and confidence of recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 66-85. doi: 10.1037/a0034059

Semmler, C., Brewer, N., & Douglass, A. B. (2011). Jurors believe eyewitnesses. In B. L. Cutler (Ed.), *Conviction of the innocent: Lessons from psychological research* (pp. 185–209). Washington, DC: American Psychological Association. doi: 10.1037/13085-009

Starns, J. J., Lane, S. M., Alonzo, J. D., & Roussel, C. C. (2007). Metamnemonic control over the discriminability of memory evidence: A signal detection analysis of warning effects in the associative list paradigm. *Journal of Memory and Language*, *56*, 592-607. doi: 10.1016/j.jml.2006.08.013

Tousignant, C., Bodner, G. E., & Arnold, M. M. (2015). Effects of context on recollection and familiarity experiences are task dependent. *Consciousness and Cognition*, *33*, 78-89. doi: 10.1016/j.concog.2014.11.011

Tulving, E. (1985). Memory and Consciousness. *Canadian Psychology/Psychologie Canadienne*, *26,* 1-12. doi: 10.1037/h0080017

Tulving, E. (1989). Memory: Performance, Knowledge, and Experience. *European Journal of Cognitive Psychology, 1*, 3-26. doi: 10.1080/09541448908403069

Watson, J. M., Bunting, M. F., Poole, B. J., & Conway, A. R. (2005). Individual differences in susceptibility to false memory in the Deese-Roediger-McDermott paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 76.

Watson, J. M., McDermott, K. B., & Balota, D. A. (2004). Attempting to avoid false memories in the Deese/Roediger—McDermott paradigm: Assessing the combined influence of practice and warnings in young and old adults. *Memory & Cognition, 32,* 135-141. doi: 10.3758/bf03195826

Weber, N., & Brewer, N. (2004). Confidence–accuracy calibration in absolute and relative face recognition judgments. *Journal of Experimental Psychology: Applied, 10,* 156–172. doi: 10.1037/1076-898x.10.3.156

Weber, N., & Brewer, N. (2006). Positive versus negative face recognition decisions: Confidence, accuracy and response latency. *Applied Cognitive Psychology, 20,* 17–31. doi: 10.1002/acp.1166

Wells, G. L. (1984). The Psychology of Lineup Identifications1. *Journal of Applied Social Psychology, 14*, 89-103. doi: 10.1111/j.1559-1816.1984.tb02223.x

Wells, G. L. (1993). What do we know about eyewitness identification? *American Psychologist*, *48,* 553-571. doi: 10.1037//0003-066x.48.5.553

Wells, G. L. (2008). Field experiments on eyewitness identification: Towards a better understanding of pitfalls and prospects. *Law and Human Behavior*, *32*, 6-10. doi: org/10.1007/s10979-007-9098-4

Wells, G. L., & Lindsay, R. C. (1980). On estimating the diagnosticity of eyewitness nonidentifications. *Psychological Bulletin*, *88,* 776-784. doi: 10.1037/0033-2909.88.3.776

Wells, G. L., & Luus, C. E. (1990). Police Lineups as Experiments: Social methodology as a framework for properly conducted lineups. *Personality and Social Psychology Bulletin*, *16*, 106-117. doi: 10.1177/0146167290161008

Wells, G. L., & Seelau, E. P. (1995). Eyewitness identification: Psychological research and legal policy on lineups. *Psychology, Public Policy, and Law,1,* 765-791. doi: 10.1037/1076-8971.1.4.765

Wells, G. L., Steblay, N. K., & Dysart, J. E. (2015). Double-blind photo lineups using actual eyewitnesses: An experimental test of a sequential versus simultaneous lineup procedure. *Law and Human Behavior*, *39*, 1-14. doi: 10.1037/lhb0000096

Westerberg, C. E., & Marsolek, C. J. (2006). Do instructional warnings reduce false recognition? *Applied Cognitive Psychology*, *20*, 97-114. doi: 10.1002/acp.1170

Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review*, *117*, 1025-1054. doi: 10.1037/a0020874

Wright, D.B., & London, K. (2009). Multilevel modelling: Beyond the basic applications. *British Journal of Mathematical and Statistical Psychology,* 62, 439-456. doi: 10.1348/000711008X327632

Wright, D. B., & McDaid, A. T. (1996). Comparing system and estimator variables using data from real lineups. *Applied Cognitive Psychology,* 10, 75-84. doi: 10.1002/(SICI)1099-0720(199602)10:1%3C75::AID-ACP364%3E3.0.CO;2-E

Wyler, H., & Oswald, M. E. (2016). Why misinformation is reported: evidence from a warning and a source-monitoring task. *Memory*, 1-16. *In press*

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*, 441-517. doi: 10.1006/jmla.2002.2864

CHAPTER 5

Do Eyewitnesses Recall-to-Reject? Familiarity and Recollection in Simultaneous Lineup Decisions[1]

Abstract

A driving force behind eyewitness identification research has been the role of mistaken identification evidence in many cases of wrongful conviction. Yet, few attempts have been made to theorise and test underlying memory mechanisms that contribute to false identification. We extended a continuous dual-process model of recognition memory to the simultaneous lineup task to investigate eyewitnesses' use of familiarity and recollection in their lineup decisions. Lineup tasks are akin to basic recognition tests that include highly-similar targets and foils. To avoid false recognition of foils that are misleadingly familiar because of this similarity to a target, recollection of target detail is needed (recall-to-reject). However, in basic recognition tasks, participants sometimes fail to use recall-to-reject when needed, or underestimate the importance of recollection. We proposed that eyewitnesses routinely neglect recollection when they adopt a best-match-first decision-making strategy. In two mini-lineup experiments, we manipulated exposure conditions to vary the availability of familiarity and recollection and investigate eyewitnesses' use of recall-to-reject in simultaneous lineups. Participants relied on familiarity when recollection was relatively weak, and were thus susceptible to false identification. However, available recollection was not neglected. Results indicate that familiarity-based responding is likely in commonplace

[1] Under review at *Journal of Experimental Psychology: Learning, Memory, and Cognition* as Guerin, N., Weber, N., & Horry, R. (2016). Do eyewitnesses recall-to-reject? Familiarity and Recollection in Simultaneous Lineup Decisions. As first author, I designed and conducted the experiments and analyses, interpreted results, and wrote the manuscript. Nathan Weber, as my PhD supervisor, advised on the experimental and analytical design, and contributed to the interpretation of results and manuscript editing. Ruth Horry, as Adjunct Supervisor, contributed to the interpretation of results and the manuscript.

witnessing conditions that limit available recollection, raising the risk of false identification. Therefore, interventions for the improvement of identification accuracy should focus on enhancing eyewitnesses' metacognitive awareness of the absence of recollection, and maximising the use of available recollection.

Introduction

It has become an acknowledged truism that mistaken eyewitness identifications are an example of real-world false recognition that carries the most serious of consequences. Yet identification evidence continues to be influential and held in high regard by police (Wogalter, Malpass, & McQuiston, 2004) and juries (Garrett, 2011; Semmler, Brewer, & Douglas, 2011). The Innocence Project (2016) and other organisations challenging wrongful conviction, report that for over 70% of exonerees, misidentification evidence was a major contributor to conviction. In lab studies, a robust effect of false identification is apparent (Cutler & Penrod, 1995; Steblay, Dysart, & Wells, 2011). Rare studies of police lineup evidence demonstrate that eyewitnesses identify known-innocent filler lineup members at considerable rates (e.g., 26%, Horry, Memon, Wright & Milne, 2012; 36.2%, Wells, Steblay, & Dysart, 2015). The base rate of suspect guilt in police lineups cannot be known, but these high rates of filler identification indicate a substantial risk of false identification for innocent suspects.

While the risks of misidentification are well-documented, little research has investigated the memory mechanisms likely to be responsible for these errors (Brewer & Wells, 2011; Gronlund, Mickes, Wixted, & Clark, 2015; Lane & Meissner, 2008). Continuous dual-process models of recognition (e.g., Wixted, 2007; Wixted & Mickes, 2010) suggest one likely contributory mechanism. These models propose that varying amounts of familiarity and recollection can contribute to recognition. In basic laboratory tasks, greater false recognition of foils that closely resemble studied targets is observed when the availability of recollection is more restricted (for reviews, see Malmberg, 2008; Yonelinas, 2002). These errors result from the substantial but misleading familiarity and global match of similar foils to targets. Moreover, people frequently neglect the importance of recollected detail for avoiding false recognition of highly-similar foils (e.g., Malmberg & Xu, 2007). We propose that eyewitnesses routinely neglect recollection. Specifically, eyewitnesses rely too much on the degree of positive global match between lineup members and memory for the offender, and neglect potentially disconfirming recollected detail, rendering false identification more likely (N. Guerin, Weber, & Horry, 2016a).

False Recognition in Simultaneous Lineups

In simultaneous lineups, several lineup members are presented live, or in a photo or video array. The lineup includes a suspect (guilty or innocent) and several known-innocent fillers, selected for their resemblance to the suspect's appearance or a description of the offender (Luus & Wells, 1991). As in old/new recognition tasks, eyewitnesses must distinguish between guilty (old) and innocent (new) suspects. But, peculiar to lineup tasks, they must also differentiate the offender, when present, from several highly-similar fillers. Three possible errors may result: a missed identification of a guilty suspect; a false identification of an innocent suspect; or, a filler identification, from a target-present or a target-absent lineup. In genuine police lineups, filler identifications represent a missed opportunity to garner more direct evidence of guilt or innocence, but do not carry as serious consequences as other possible errors. As a result, their importance has often been overlooked in identification research (cf. Clark & Wells, 2008; Wells, 2008; Wells & Olson, 2003). However, filler identifications can index the memorial basis of identification decisions. Both filler and false identifications indicate that detailed, specific recollection has not contributed to the witnesses' decision (though, false subjective recollective experience is possible, e.g., Brainerd, Reyna, Wright, & Mojardin 2001; Lampinen, Neuschatz, & Payne, 1999; Lampinen, Watkins, & Odegard, 2006). Thus, these errors provide clear evidence of responses based on global familiarity.

A dual-process account of false identification. In continuous dual-process models of recognition (e.g., Mickes, Wais, & Wixted, 2009; Wixted & Mickes, 2010) the strengths of familiarity and recollection vary continuously. The amount, clarity, and relevance of recollective detail may vary, and both familiarity and recollection may be weak or strong (e.g., Benjamin, 2001; Mickes, Johnson, & Wixted, 2010). Recollection can either bolster familiarity-based evidence of positive match for a target, or offset the misleading familiarity of similar-looking foils (i.e., recall-to-reject, Rotello & Heit, 2000; or recollection-rejection; Brainerd et al., 2003; Reyna & Brainerd, 1995). Both processes are proposed to be activated together when cued, but strong familiarity may be accompanied by no or weak recollection (Ingram, Mickes, & Wixted, 2012). Therefore, participants may rely on the strength of familiarity alone, when recollection is unavailable (e.g., Gallo, Sullivan, Daffner, Schacter, & Budson, 2004; Hintzmann & Curran 1994). Importantly, the more effortful retrieval of recollection may also be neglected even when available, if its usefulness is not appreciated by the decision-maker (e.g., Dyne, Humphries, Bain & Pike, 1990; Malmberg & Xu, 2007; Starns, Lane, Alonzo, & Roussel, 2007).

For some tasks, familiarity alone can support accurate responding (Hintzmann, 1988). For example, in an *n*-alternative forced-choice (*n*-AFC) task, each test trial comprises a target and *n*-1 foils, with the distribution of target familiarity expected to be higher than the foil distribution. Thus, a familiarity-based response will tend to be accurate (e.g., Cook, Marsh, & Hicks, 2005; Migo, Montaldi, Norman, Quamme, & Mayes, 2009). Similarly, in single item recognition, familiarity will reliably differentiate between targets and non-similar, unstudied foils (Gronlund & Ratcliff, 1989). However, in tasks that call for discrimination between targets and highly-similar foils, familiarity-based responding is susceptible to false recognition of foils that seem familiar by dint of their resemblance to targets. The lineup is such a task.

The complexity and idiosyncrasy of human faces makes estimation of the degree of target-filler similarity in a specific lineup difficult, and levels of similarity between lineup members will vary considerably between lineups (e.g. Clark & Godfrey, 2009; Clark, Howell, & Davey, 2005; Clark & Tunnicliff, 2001; Fitzgerald, Price, Oriet, & Charman, 2013; Tunnicliff & Clark, 2000). Much evidence for the role of recollection in avoiding false recognition comes from associative or plurality discrimination tasks (Malmberg, 2008; Yonelinas, 2002). High target-foil similarity in these tasks is such that foil familiarity increases with increasing target familiarity (Hintzmann & Curran, 1994). However, target-foil similarity may also be substantial in standard tasks, without reaching the level of similarity characteristic of associative recognition tasks or plurality discrimination (Malmberg, 2008; Yonelinas, 2002). Whether target-foil similarity is associative, or just substantial, there is ample evidence that misleading foil familiarity is implicated in false recognition (e.g., Kelley & Wixted, 2001; Nobel & Shiffrin, 2001; Rotello, Macmillan, & Van Tassel, 2000), including memory for faces (e.g., Horry, Wright, & Tredoux, 2010; Marcon, Susa, & Meissner, 2009; Xu & Malmberg, 2007). In simultaneous lineups, both innocent suspects and fillers are likely to be associated with substantial familiarity. Therefore, recollection of the target is needed to offset this misleading familiarity and prevent mistaken identification of fillers and innocent suspects.

Neglect of recollection. Eyewitnesses asked to make an identification decision for a police lineup will generally be aware of the serious ramifications of their responses. Why, then, would they fail to consciously attempt to retrieve the kind of detailed recollection of the offender that might be important for making an accurate identification decision when it is potentially available? We propose two key factors: the structure of the task itself, and the preconceptions that eyewitnesses bring to the task.

Task structure. Our best-match-first dual-process account proposes that eyewitnesses tend to make simultaneous lineup decisions in two steps (N. Guerin et al., 2016a; see also, Clark, 2003; Wells, 1984, 1993). First, eyewitnesses focus on finding the best-matching lineup member to their memorial representation of the offender. This decision can be likened to an *n*-AFC task, which can be performed successfully on the basis of familiarity (Malmberg, 2008; Yonelinas, 2002). Second, eyewitnesses evaluate whether the best-match is the offender (best-match-first decision-making). If task expectations are not re-set, witnesses may not more effortfully retrieve the further recollective evidence needed for this second judgment (e.g., Alban & Kelley, 2012; Cook et al., 2005). Instead, witnesses might commit to their best-match selection (e.g., Deffenbacher, Bornstein, & Penrod, 2006; Goodsell, Neuschatz, and Gronlund, 2009; Schreiber & Sergent, 1998), or set a low criterion for identifying the best-match candidate as the offender. Eyewitnesses are thus likely to base their identification decision overall on the familiarity retrieved to select the best-match, rather than also considering recollection. This approach leaves them susceptible to false identification when the best-match is an innocent lineup member who looks like the offender. Therefore, in the two experiments reported here, we sought to extend previously reported but less direct evidence of reliance on familiarity (N. Guerin et al., 2016a, 2016b) with a direct examination of the relative contribution of familiarity and recollection to simultaneous lineup decisions.

Metacognitive approach. Just as immediate task perceptions influence decision-making (Dyne et al., 1990; Malmberg & Xu, 2007), so do pre-existing metamemorial knowledge and understandings (e.g., Lane, Roussel, Villa, & Morita, 2007; Odegard & Lampinen, 2006; Starns et al., 2008). However, metacognitive understanding does not always correspond to task requirements, resulting in sub-optimal performance (e.g., Blank & Launay, 2014; Gallo, 2010; Lampinen, Odegard & Neuschatz, 2004). For example, a robust predisposition to choose from simultaneous lineups is evident in the identification literature (e.g., Brewer & Palmer, 2010; Cutler & Penrod, 1995; Wells et al., 1998; Wells & Seelau, 1995). Wells (1993) demonstrated that when the offender was removed from a lineup, witnesses tended to positively identify the next-most-plausible lineup member. Charman, Gregory, and Carlucci (2009) demonstrated that prior belief in suspect guilt can inflate the perceived similarity between a photo of a suspect and a facial composite of an offender, biasing the overall evaluation of available evidence. A similar bias applied to lineup members would increase the risk of false identification. Furthermore, unbiased instructions that the offender may or may not be in the lineup reduce eyewitness choosing without improving discriminability, consistent with an original predisposition to adopt a lenient identification criterion (Clark, 2005, 2012; Malpass & Devine, 1981). The tendency to choose might reflect a general preconception that the purpose of an identification task is to help the police apprehend the offender (N. Guerin et al., 2016b). Eyewitnesses' predisposition to choose is then reinforced and facilitated in simultaneous lineups by the best-match-first decision processes encouraged by task structure, as we have outlined (cf. Gronlund, 2005; Meissner, Tredoux, Parker, & MacLin, 2005).

This combination is particularly problematic if eyewitnesses also overestimate the quality of their memory for the offender. Indeed, eyewitness are frequently over-confident about their positive identification decisions (Brewer & Weber, 2008; Palmer, Brewer, Weber, & Nagesh, 2013). Overconfidence is evidence of a lack of alignment between eyewitnesses' metamemorial knowledge and the nature of the task, but might also indicate decision monitoring based on positive familiarity alone, without the counterbalancing contribution of disconfirming recollection (Higham, Perfect, & Bruno, 2009; Starns et al., 2007). Overall, the typical approach to a simultaneous lineup leaves eyewitnesses susceptible to familiarity-based false identification when an innocent suspect is a plausible match for the offender.

The Experiments

In two face-recognition mini-lineup experiments, we aimed to investigate the contributions of familiarity and recollection to simultaneous lineup decisions. Specifically, we aimed to test whether a neglect of recollection would explain observed patterns of false identification. Participants studied photos of faces, and made an identification decision for a simultaneous four-person lineup for each target (Sauer, Brewer, & Weber, 2008; Weber & Brewer, 2004, 2006). Collecting multiple data points per participant allowed regularities of responding to be evaluated across stimulus variation (cf. an identification paradigm, which typically yields one data point per witness).

To disaggregate their effects on identification responses, we separately varied the strengths of both familiarity and recollection using encoding manipulations demonstrated to have differential effects on the two processes (e.g., Malmberg, 2008; Yonelinas, 2002). Massed repetitions of targets during study increase familiarity in face recognition paradigms, without affecting recollection to the same extent (e.g., Mantyla & Cornoldi, 2002; Parkin, Gardiner, & Rosser, 1995). We varied the number of massed repetitions (two, six) of target face photos within-subjects during study to vary target familiarity on two levels (low, high). To vary available recollection between-subjects on two levels (low, high), we used a divided attention manipulation (divided, full). Dividing attention decreases recollection more than familiarity in recognition tasks (e.g., Gardiner & Parkin, 1990; Reinitz, Morrissey, & Demb, 1994), including lineups (e.g., Palmer, Brewer, McKinnon, & Weber, 2010). Thus, in the divided attention condition, when encoded recollection was expected to be relatively low for all stimuli, we could evaluate the effect of differences in the strength of target familiarity on responses to both target-present and target-absent mini-lineups. This response pattern could be compared with that obtained from participants who studied targets with full attention, and were able to encode relatively more recollective detail related to targets.

Overall, we expected a significantly higher likelihood of accurate positive identification of targets with increased target familiarity. Innocent suspects were expected to elicit substantial familiarity due to their description-matched similarity to the target, but specific expectations for patterns of false identification responses with increased target familiarity depended on whether the lineup task was more akin to a standard or an associative task (Malmberg, 2008; Yonelinas, 2002). In the Results section, we discuss the predictions and analytic rationale relevant to this issue. Regardless, our main focus of interest was that increased recollection of the target would tend to disconfirm the misleadingly positive familiarity associated with similar-looking fillers and innocent suspects. Therefore, if recollection was not neglected, the likelihood of false identification should be lower when more recollection was available. Thus, we predicted that if eyewitnesses used recollection to avoid false identification decisions, discriminability would be greater with full than divided attention, with a lower likelihood of false identification. In contrast, if recollection was neglected, we expected an equivalent effect of repetitions, regardless of attention condition.

Method

Except where indicated, both experiments followed the same method.

Participants and Design

We used a 2 (attention: divided; undivided) × 2 (repetitions: two; six) x 2 (target presence: target-present; target-absent) mixed experimental design was used, with attention varied between-subjects, and repetitions and target-presence manipulated within-subjects. For each participant, mini-lineups were randomly assigned within repetitions to target-present or target-absent status.

We aimed to collect data from 80 participants per experiment, a priori.[2] Ninety-four undergraduate students from Flinders University participated in Experiment 1 for course credit or payment. Two participants were excluded from analyses (one female aged 74; one male reported memory impairment following his participation) leaving a final sample of 92 (37 male, age: M = 22.63, SD = 6.31). In Experiment 2, 100 undergraduate students participated (30 male, age: M = 24.3, SD = 8.05). All collected data in the final samples were analysed.

Materials

For Experiment 1, we used 40 sets of mini-lineup stimuli from Weber and Brewer (2004, 2006). Each set included a randomly designated target, a description-matched target-replacement, and three description-matched fillers. Each photo showed a front view of the head and neck with neutral facial expression and background (200 × 200 pixels). Mini-lineup sets were divided into two blocks of 20. In each block, half of the lineups were target-present and half target-absent. Stimuli were presented on an 18” monitor (resolution: 1920 × 1080 pixels). All mini-lineups were presented in a row of four, with response buttons below. To allow participants to look for a specific target in each mini-lineup (as a witness would look for a specific offender), rather than any studied face (Weber & Brewer, 2006), a text cue (name or occupation, by block) was presented with each target (study) and mini-lineup (test). The cue was centred above photos. Photo order in each mini-lineup, trial order in each block, and cue and target-presence assignment were randomly determined for each participant.

Procedure

Participants gave informed consent, and completed the experiment on PC in individual cubicles using a mouse to click on-screen buttons in response to written prompts. They

[2] Guided by power estimates to test hypotheses with mixed-effects models (Gelman & Hill, 2007),

completed two blocks of trials (after four practise trials), with block order, and assignment of name or occupation cues to blocks, counterbalanced. Each block included a study phase, a visual distractor task, and a test phase. In the study phase, target photos were presented sequentially, in the centre of the screen. Each cue was presented for 1000 ms before presentation of the target, and throughout all repetitions of that target (500 ms per repetition with an inter-stimulus interval of 1000 ms). Following each study phase, participants worked on a spatial memory task for 3 min. In each test phase, participants were given unbiased instructions that each mini-lineup may or may not include a target (Malpass & Devine, 1981) and completed 20 mini-lineup trials. Participants provided an identification response by identifying a lineup member as the studied target (by clicking on the photo), or clicking the *Not Present* button. In Experiment 1, after each identification response, participants indicated their confidence in their decision from 0% to 100% in 10% increments by clicking one of 11 buttons displayed in a single row. During study, r*epetitions* were manipulated within-subjects and *attention* was manipulated between-subjects.

Repetitions. Half of the target photos were presented for study with two massed (consecutive) repetitions and the remaining half with six. Assignment of targets to repetitions condition was randomly determined for each participant. Participants were informed that photos would be presented a variable number of times.

Attention. For all participants, the presentation of targets was accompanied by a sequence of audio tones randomised for pitch (high, low) and intervening interval (1000 or 2000 ms). In the divided attention condition, participants responded to tones with a mouse click (left-button for high, right for low). This was expected to limit available attentional resources for encoding recollective detail related to target stimuli (Yonelinas, 2002). In the undivided condition, participants were told to ignore the soundtrack. The attention manipulation was included in the practice trials.

Experiment 2. Experiment 2 differed in two key ways. First, decision confidence was not measured. Second, we used a different stimulus set. This set comprised 40 mini-lineups with different images of targets used at study and test (N. Guerin et al., 2016b). Target photos at test showed a front view of the head and neck with neutral facial expression and background. In contrast, target photos at study showed either a front view of the head and neck with a different facial expression (smiling), or a three-quarter view of the head and neck with the same facial expression. This study-test variation makes recognition more challenging (e.g., Hancock, Bruce, & Burton, 2000) and more closely approximates the encoding-test variability experienced by eyewitnesses.

Results & Discussion

Overview

We used an alpha level of .05 for all inferential analyses and focussed on trial-by-trial performance in planned pairwise comparisons between conditions of interest (Gelman, Hill & Yajima, 2012). Logistic mixed-effects regression models including all fixed factors of interest were used for confirmatory hypothesis testing (Gelman & Hill, 2007; Judd, Westfall, & Kenny, 2012). Random effects included random intercept by participant and stimulus, and random slopes by participant for within-subjects, and stimulus for within-stimulus, manipulations (Baayen, Davidson, & Bates, 2008; Barr, Levy, Scheepers, & Tily, 2013; Murayama Sakaki, Yan & Smith, 2014; Westfall, Judd, & Kenny, 2015). The main outcome variable was identification decision type (lineup rejection; positive identification) in models used to evaluate the effects of repetitions and attention on choosing (response bias) and discriminability (accuracy, regardless of response bias) (De Carlo, 1998; Wright & London, 2009). We also examined models predicting decision accuracy to assess the moderating effect of attention on the confidence-accuracy relationship.[3]

Logistic regression models predict a binary categorical outcome variable and estimate the log odds for observing a given outcome. Values are symmetrical around zero, with values greater than 0 indicating the predicted outcome (e.g., a positive identification) is more likely, and values lower than 0 indicating the opposite outcome (e.g., a lineup rejection) is more likely. Coefficients of binary fixed factors indicate the log odds ratio (*lnOR)* of the outcome variable between factor levels. We took effects to be significant if the 95% confidence interval (CI) around the coefficient excluded 0. For ease of interpretation, we also report odds ratios (*ORs,* the exponents of lnOR), when discussing differences between experimental cells. Aggregate descriptive statistics are presented as an index of identification performance.

Filler identifications are known errors, not always relevant for applied questions about lineup reliability (except insofar as they are informative about lineup fairness, Wells, 2008). However, filler identification patterns can index memory or decision processes (e.g., the use of familiarity rather than recollection) and have theoretical implications. Therefore, we report analyses both with and without filler identifications included in the dataset.

Accuracy and Decision Type (Choosing and Discriminability)

For both experiments, we compared the likelihood of choosing (positive identifications of suspects and fillers) between levels of attention (divided; full) as moderated by repetitions (two; six). We also examined the effects of attention and repetitions on decision accuracy, using a measure of comparative discriminability. When filler identifications (from both target-absent and target-present lineups) were excluded, our measure of discriminability indexed the accuracy of positive identifications (i.e., witnesses' ability to discriminate guilty

[3] Data were analysed using the lme4 package (Bates, Maechler, Bolker, & Walker, 2014) for R (R Core Team, 2014).

from innocent suspects). When included, our measure of discriminability indexed the accuracy of target *presence* decisions, regardless of whether the lineup member identified was the suspect (innocent or guilty) or a filler. As patterns of effects were similar when filler identifications were included or not, we discuss them together, but with specific reference to patterns of filler identification when pertinent.

Predictions for the effects of attention and repetitions were tested in models with decision type (lineup rejection; positive identification) as the outcome variable. Attention (divided = 0, full = 1), repetitions (two = 0, six = 1), target presence (target-absent = 0, target-present = 1), and their interactions were fixed factors. The coefficient for the fixed factor of attention indexed any difference in witnesses' tendency to choose (response bias) between divided and full-attention conditions. Similarly, the coefficient for the fixed factor of repetitions indexed differences in witnesses' response bias depending on the number of repetitions of the target at study. The coefficient for target presence indexed the accuracy of choosing decisions, independent of bias (discriminability). Variation in discriminability between levels of attention was thus indexed by the attention × target presence interaction term (De Carlo, 1998; Wright & London, 2009). Likewise, variation in discriminability between levels of repetitions, was indexed by the repetitions × target presence interaction term. Finally, difference in the effect of attention on discriminability depending on repetitions condition was indexed by the interaction term for all fixed factors (repetitions × attention × target presence).

Experiment 1. Models used to assess the effects of repetitions and attention on choosing and discriminability are reported in Tables 1 & 2, and plotted in Figure 1 (for random effects for each model, see online supplement).[4] Effects are most easily interpreted

[4] Supplemental materials for this chapter are presented in Appendix 3

with reference to the plotted model. We used these models to check the manipulation of familiarity, before testing hypotheses.

Manipulation checks. As expected if the manipulation of study repetitions had successfully affected target familiarity, positive identifications for target-present lineups (hits) were substantially more likely when targets had been presented six times at study, than when presented only twice. Examination of model coefficients revealed that target presence, and the interaction term for target presence with repetitions, were significant predictors of decision type. We compared this effect of repetitions on target present trials at each level of attention to check that our manipulation of attention had not confounded the manipulation of familiarity. As evident in Figure 1 (Panels A-D), for target-present lineups there was no significant or meaningfully-sized difference in the effect of repetitions on target familiarity between level of attention (when filler identifications were included: for divided attention, $OR_{2\ Repetitions:6\ Repetitions} = 0.54$; for full attention, $OR_{2\ Repetitions:6\ Repetitions} = 0.61$; when filler identifications were not included: for divided attention, $OR_{2\ Repetitions:6\ Repetitions} = 0.53$; for full attention, $OR_{2\ Repetitions:6\ Repetitions} = 0.61$). Thus, we can confidently conclude that the manipulation of repetitions produced two levels of familiarity, unconfounded by our manipulation of attention.

Consistent with participants in the divided attention condition having attended to the tone monitoring task, the proportion of tones for which divided attention participants made a correct response before the commencement of the next tone was significantly and substantially above chance, $t = 16.63$, $df = 48$, $p < .001$ ($M = .69$, $SD = .08$). There was no way to evaluate the effect of the tones on participants in the full-attention condition, separate from hypothesis testing.

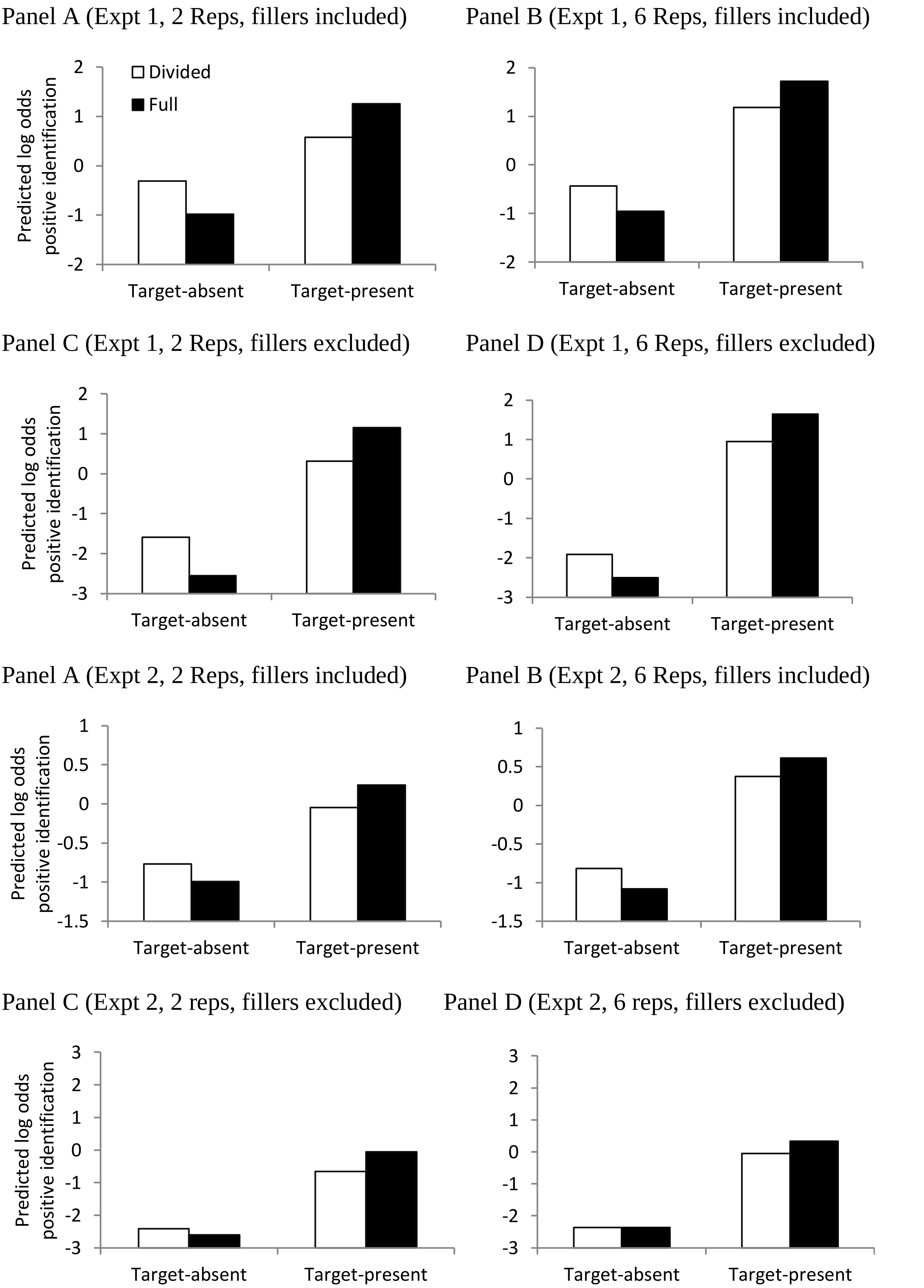


Figure 1. Plots of predicted log odds of a positive identification decision in Experiments 1 & 2 by repetitions and attention. Values greater than 0 indicate a positive identification from the

mini-lineup is more likely, and values lower than 0 indicate a lineup rejection is more likely. Coefficient standard errors, indicative of the variability of the estimates, are displayed in Table 2.

Table 1

Model Fit Statistics for Mixed-Effects Models Predicting Decision Type (Positive Identification; Lineup Rejection) by Attention and Repetitions in Experiments 1 & 2

Filler Identifications	Experiment	χ^2	*df*	*p*
Included	1	97.35	3	< .001
	2	79.26	3	< .001
Excluded	1	110.09	3	< .001
	2	110.43	3	< .001

Number of observations, filler identifications included (excluded): *Experiment 1* = 3680 (2946), n = 92 (92); *Experiment 2* = 4000 (3172), n = 100 (100).

Table 2

Fixed Effect Coefficients for Mixed Effects Models Predicting Decision Type (Lineup Rejection; Positive Identification) by Repetitions and Attention in Experiments 1 & 2

		Expt 1			Expt 2		
Filler IDs	Fixed Effects	*b*	SE_b	95% CI_b	*b*	SE_b	95% CI_b
Included	Intercept	-00.31			-0.77		
	Reps	-0.14	-.14	-0.42, .0.14	-0.05	0.15	-0.34, 0.24
	Attn	-0.68*	0.28	-1.22, -0.13	-0.22	0.28	-0.78, 0.34
	TP	0.89*	0.23	0.43, 1.33	0.72*	0.17	0.38, 1.06
	Reps × TP	0.75*	0.21	0.35, 1.15	0.47*	0.20	0.08, 0.86
	Attn × TP	1.35*	0.31	0.74, 1.97	0.51*	0.25	0.03, 1.00
	Reps × Attn	0.13	0.22	-0.30, 0.56	-0.04	0.22	-0.47, 0.38
	Reps × Attn × TP	-0.25	0.32	-0.87, 0.37	-0.00	0.29	-0.57, 0.56
Excluded	Intercept	-1.59			-2.41		
	Reps	-0.32	0.22	-0.75, 1.10	0.04	0.25	-0.45, 0.54
	Attn	-0.97*	0.34	-1.63, -0.30	-0.19	0.34	-0.85, 0.47
	TP	1.90*	0.30	1.33, 2.47	1.75*	0.26	1.23, 2.27
	Reps × TP	0.96*	0.27	0.43, 1.49	0.57	0.30	-0.01, 1.14
	Attn × TP	1.81*	0.41	1.00, 2.61	0.79*	0.35	0.11, 1.48
	Reps × Attn	0.37	0.35	-0.31, 1.06	0.19	0.37	-0.53, 0.91
	Reps × Attn × TP	-0.51	0.43	-1.35, 0.32	-0.41	0.42	-1.24, 0.42

Note: * indicates significant predictors. Number of observations, filler identification included (excluded): *Experiment 1* = 3680 (2946), *n* = 92 (92); *Experiment 2* = 4000 (3172), *n* = 100 (100).

Target-foil similarity. In associative recognition tasks, the familiarity of foils that are highly similar to targets increases significantly with target familiarity (e.g., Malmberg, 2008). In contrast, the familiarity of foils in standard recognition tasks with relatively lower target-foil similarity is unaffected by target familiarity, even when target-foil similarity is substantial. Our interpretation of the effect of familiarity and the moderating effect of recollection, depended on whether the mini-lineup task was an example of standard or associative recognition. Description-matched lineup members share substantial similarity. However, human faces are extraordinarily diverse, even within the limits of a modal description (e.g., Brigham, Meissner, & Wasserman, 1999; Malpass & Lindsay, 1999; Tredoux, 2002). Therefore, we evaluated whether the level of target-foil similarity within lineups created a standard or associative task.

We found no significant or meaningfully-sized effect of repetitions on the likelihood of false identification (Table 2). There was no significant effect of repetitions and no significant difference in the effect of repetitions between levels of attention (the target presence × repetitions × attention interaction term was not a significant predictor). Figure 1 (Panels A-D) illustrates that false identification decisions were no more likely when the target had been presented six times at study than two, at either level of attention (when filler identifications were included: for divided attention, $OR_{2\ Repetitions:6\ Repetitions} = 1.15$; for full attention, $OR_{2\ Repetitions:6\ Repetitions} = 1.01$; when filler identifications were not included: for divided attention, $OR_{2\ Repetitions:6\ Repetitions} = 1.38$; for full attention, $OR_{2\ Repetitions:6\ Repetitions} = 0.95$). Thus, target familiarity had no significant effect on the likelihood that witnesses would make a false identification, providing no evidence of an associative effect of target-foil similarity on the familiarity of target-replacements. Effect sizes were negligible, and often in the opposite direction than expected for associative recognition, consistent with a standard pattern of recognition, albeit with substantial target-foil similarity.[5]

Neglect of recollection. We predicted a neglect of recollection would be evident in witnesses' ability to discriminate between guilty and innocent suspects. Counter to predictions, we found a clear and substantial effect of attention on discriminability (i.e., significant target presence × attention, and target presence × repetitions interactions, see Figure 1, Panels A-D). Participants in the full-attention condition were more likely to accurately identify a target from a target-present lineup and less likely to falsely identify an innocent suspect from a target-absent lineup, than those in the divided-attention condition. Thus, full attention allowed participants to better identify studied targets and to discount the misleading familiarity of innocent suspects to avoid familiarity-based false identification. This is consistent with the use of recall-to-reject (Rotello & Heit, 2000; also known as recollection rejection, e.g., Lampinen et al., 2004), a distinctiveness heuristic (Gallo, Bell, Beier & Schacter, 2006), or other recollection-based monitoring processes.

Confidence-accuracy. Results were consistent with participants in the full-attention condition having made better use of recollection than those in the divided-attention condition. If so, a significantly stronger positive relationship between confidence and accuracy with full than divided attention should be evident (Higham et al., 2009; Starns et al., 2007). Therefore, we evaluated models predicting decision accuracy (inaccurate; accurate) from confidence. As confidence is generally a better predictor of accuracy for positive than negative recognition decisions (Brewer & Wells, 2006; Weber & Brewer, 2004, 2006; Wixted, Mickes, Clark, Gronlund, & Roediger, 2015), we included decision type (positive identification, lineup rejection) as a fixed effect.

---

[5]These results cannot be generalised to other lineups, in which target-foil similarity may be lower or higher and must be evaluated case-by-case.

As expected, decision type (lineup rejection; positive identification), decision confidence, and their interaction, were significant predictors of accuracy when filler identifications were included (see Tables 3 & 4). Confidence was significantly positively related to accuracy for both positive identifications and lineup rejections, with a stronger relationship for positive identifications. The attention × confidence interaction term indexed variability in the confidence-accuracy relationship depending on whether participants had studied targets with divided or full attention. As predicted, the interaction term was a significant predictor, showing the confidence-accuracy relationship to be stronger for full than divided attention, for both positive identifications and lineup rejections (see Figure 2).

Table 3

Model Fit Statistics for Mixed-Effects Models Predicting Decision Accuracy by Attention, Decision Type, and Confidence in Experiment 1

Filler Identifications	χ^2	*df*	*p*
Included	522.84	7	< .001
Excluded	2055.00	7	< .001

No. of observations, filler identifications included (excluded): *Experiment 1* = 3680 (2946), *n* = 92 (92).

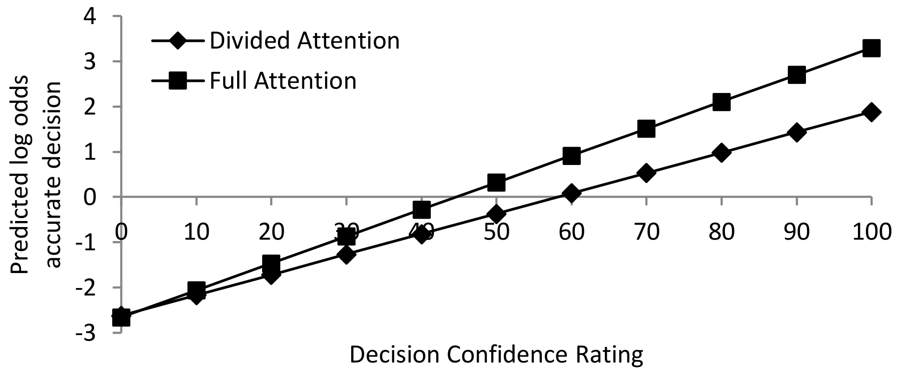
Table 4

Fixed Effect Coefficients for Mixed Effects Model Predicting Decision Accuracy by Attention, Decision Type and Confidence in Experiment 1

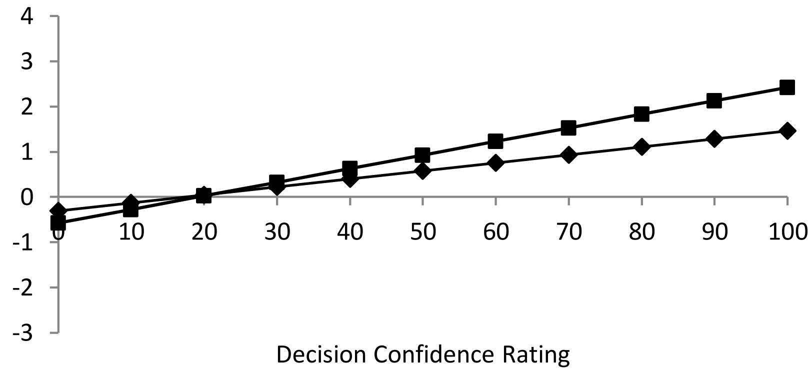
	Filler Identifications					
	Included			Excluded		
Fixed Effect	*b*	SE_b	95% CI_b	*b*	SE_b	95% CI_b
Intercept	-0.30			-0.88		
Confidence (Conf)	0.18*	0.03	0.11, 0.24	0.08*	0.03	0.01, 0.14
Attn	-0.27	0.34	-0.93, 0.39	-0.31	0.33	-0.95, 0.33
Decision Type (Dec)	-2.32*	0.25	-2.80, -1.83	0.36	0.30	-0.23, 0.96
Attn × Dec	0.23	0.40	-0.55, 1.01	0.83	0.51	-0.17, 1.83
Attn × Conf	0.12*	0.05	0.02, 0.23	0.08	0.06	-0.03, 0.19
Conf × Dec	0.27*	0.04	0.19, 0.36	0.27*	0.05	0.17, 0.38
Conf × Attn × Dec	-0.02	0.07	-0.15, 0.11	-0.06	0.09	-0.23, 0.10

Note: * indicates significant predictors. Number of observations, filler identifications included (excluded): 3680 (2946), n = 92 (92).

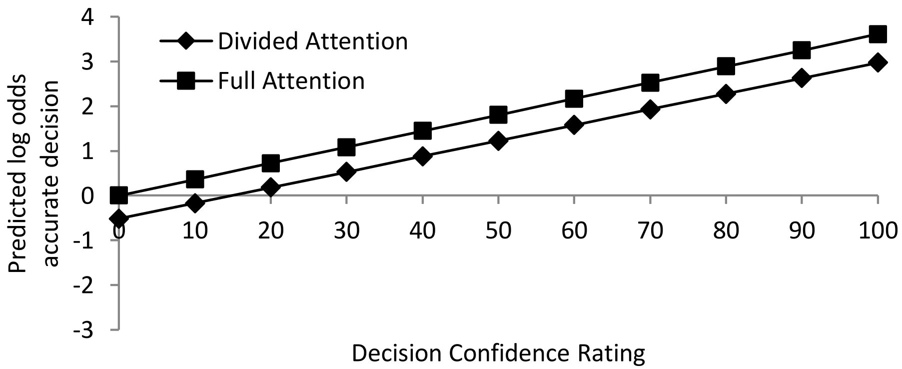
Panel A (Expt 1, Positive Identifications, fillers included)



Panel B (Expt 1, Lineup Rejections, fillers included)



Panel C (Expt 1, Positive Identifications, fillers excluded)



Panel D (Expt 1, Lineup Rejections, fillers excluded)

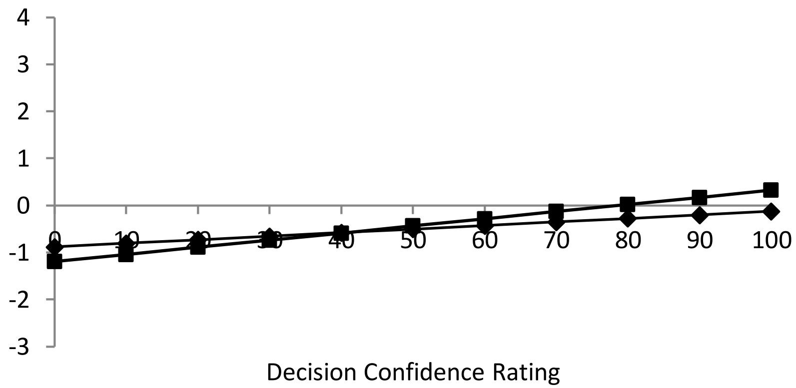


Figure 2. Plots of predicted log odds of an accurate identification decision in *Experiment 1* by attention, decision type and confidence. Values greater than 0 indicate a positive identification from the mini-lineup is more likely, and values lower than 0 indicate a lineup rejection is more likely. Coefficient standard errors, indicative of the variability of the estimates, are displayed in Table 4.

When filler identifications were excluded, the pattern of results was in the same direction, but less marked. This finding is consistent with a greater proportion of filler identifications than target identifications being familiarity-based. Removing these likely familiarity-based responses from the dataset left a more similar proportion of recollection-based responses in each condition, and thus reduced the contrast between divided-attention (i.e., relatively less recollection-based) and full-attention (i.e., relatively more recollection-based) conditions. While the interaction term for attention × confidence was not a significant predictor, the span of CIs extended largely above zero (see Table 4) and substantially overlapped the CIs observed when filler identifications were included, suggesting a meaningful relationship. Attention was not a significant predictor. With familiarity-based filler identifications removed from analyses, the effect of recollection on the confidence-accuracy relationship was less pronounced. As we would predict when a proportion of the familiarity-based incorrect responses that caused the difference in the confidence-accuracy relationship were excluded, and responses were based more predominantly on recollection, this corresponded to a smaller accuracy difference, as shown in Figure 2. Overall, results are consistent with participants having made appropriate use of available recollective evidence to make identification decisions in the full attention condition, thus using better evidence than those in the divided-attention condition.

Experiment 2. We aimed to replicate results from *Experiment 1* with a greater range of stimuli, and to rule out the possibility that making a confidence judgment following each identification decision had affected decision-making on subsequent trials. Specifically, participants might have made appropriate use of recollection as a result of metacognitive evaluation of their own decision processes.[6] Therefore, the design of *Experiment 2* followed

[6] Analyses of participants' responses to only the first mini-lineup trial did not demonstrate a neglect of recollection, but greater power was needed to rule out this possibility.

Experiment 1, with two key differences: we used different mini-lineup stimuli, and did not measure decision confidence. If *Experiment 1* results replicated, we expected participants in the full-attention condition to use available recollection to demonstrate better discriminability than those with divided attention.

Manipulation checks. As in *Experiment 1*, results were consistent with the manipulation of study repetitions having successfully affected familiarity. Positive identifications for target-present lineups were significantly more likely when targets had been presented six times at study than only two (See Tables 1 & 2), when filler identifications were included in analyses. There was also no evidence of a confounding effect of attention (target presence and the interaction term for target presence with repetitions were significant predictors of decision type, but the repetitions × attention interaction term was not). When filler identifications were excluded, the target presence × repetitions interaction term was not a significant predictor, but the relationship was meaningful.[7] In line with predictions, hits from target-present lineups were meaningfully more likely when targets had been presented six times during study than when they were presented only twice. The effect of repetitions on hits did not differ between levels of attention (when filler identifications were included: for divided attention, $OR_{2\ Repetitions:6\ Repetitions} = 0.66$; for full attention, $OR_{2\ Repetitions:6\ Repetitions} = 0.68$; when filler identifications were not included: for divided attention, $OR_{2\ Repetitions:6\ Repetitions} = 0.59$; for full attention, $OR_{2\ Repetitions:6\ Repetitions} = 0.73$). Further, in the tone monitoring task in the divided-attention condition, the proportion of tones for which a correct response was made, was significantly and substantially above chance, $t = 16.83$, $df = 51$, $p < .001$ ($M = .71$, $SD = .09$), consistent with the tones having been monitored.

[7] The range of values within the CI for this coefficient fell largely above zero. The smaller effect when filler identifications were excluded from analyses is likely to have resulted from participants having made proportionately fewer familiarity-based identifications of the target than familiarity-based identifications of fillers from target-present lineups.

Target-foil similarity. It was possible that the level of target-foil similarity for the mini-lineup stimuli would differ from that of stimuli used in *Experiment 1*. Figure 1 (Panels E-H) illustrates that false identifications were no more likely when the target had been presented six than two times at study for those in the divided-attention condition. While target presence and the target presence × repetitions interaction term significantly predicted a positive identification decision, there was no significant effect of repetitions (Table 2). The interaction term for all fixed factors (target presence, repetitions, and attention) was not a significant or meaningfully-sized predictor, indicating that this effect did not differ with attention (when filler identifications were included: for divided attention, $OR_{2\ Repetitions:6\ Repetitions} = 1.05$; for full attention, $OR_{2\ Repetitions:6\ Repetitions} = 1.09$; when filler identifications were not included: for divided attention, $OR_{2\ Repetitions:6\ Repetitions} = 1.04$; for full attention, $OR_{2\ Repetitions:6\ Repetitions} = 1.15$). Thus, there was no evidence of an associative pattern of recognition with these stimuli.

Neglect of recollection. Consistent with *Experiment 1*, and counter to predictions for a neglect of recollection, target presence, and its interaction terms with attention and repetitions, each significantly predicted decision type (See Tables 1 & 2; Figure 1, Panels E-H). Discriminability was significantly better with full than divided attention, at each level of repetitions. The smaller effect size than observed for *Experiment 1* reflected a higher level of recognition difficulty with these stimuli, due to study-test target variation (see also aggregate measures in Table 5).

As in *Experiment 1*, differences in the observed likelihood of both accurate and false identifications contributed to the significant effect of target presence. Specifically, participants in the full-attention condition more likely to make a positive identification from a target-present lineup (when filler identifications were or were not included in analyses), and less likely to make a false positive identification from a target-absent lineup than those in the divided-attention condition. When filler identifications were excluded, greater discriminability with full than divided attention was driven to a larger extent by additional accurate identifications of targets. False identification rates were less affected by the greater availability of attention. This is likely to have been due to lower choosing overall for the more challenging lineup stimuli used in *Experiment 2* (revealed by the lower intercept value for *Experiment 2* than *Experiment 1* models).

Overall, we can confidently rule out the possibility that better use of recollection demonstrated by participants with full rather than divided attention in *Experiment 1* was caused by the additional metacognitive confidence judgment made for each test lineup. Results replicate, with different stimuli, the finding that participants in a simultaneous lineup task used available recollection to improve identification decisions for both target-present and target-absent lineups.

Table 5

Proportion of Identification Responses (Hit rates, Filler ID Rates, Correct Rejection Rates and False ID Rates) by Attention and Repetitions for Experiments 1 & 2

				Response Proportion							
				Target-present				Target-absent			
Expt	Attn	Reps	Statistic	Hits	Filler IDs (TP)	Misses	*n*	Correct Rejections	Filler IDs (TA)	False IDs	*n*
1	Divided	2	*M*	.49	.14	.38	49	.55	.31	.14	49
			SD	.25	.14	.25		.25	.20	.13	
			95% CI	.46, .51	.12, .15	.36, .40		.53, .57	.29, .32	.13, .15	
		6	*M*	.57	.17	.26		.59	.30	.11	
			SD	.22	.16	.20		.24	.22	.09	
			95% CI	.55, .59	.16, .18	.25, .28		.57, .61	.28, .32	.11, .12	
	Full	2	*M*	.63	.13	.24	43	.68	.23	.10	43
			SD	.22	.18	.18		.23	.20	.10	
			95% CI	.61, .65	.11, .14	.23, .26		.65, .70	.21, .25	.09, .10	

		6	*M*	.72	.09	.18		.67	.22	.11	
			SD	.20	.14	.17		.24	.19	.10	
			95% CI	.70, .74	.08, .11	.17, .20		0.65, .69	0.20, .24	0.10, .12	
2	Divided	2	*M*	.29	.20	.51	52	.64	.28	.08	52
			SD	.27	.28	.28		.28	.26	.27	
			95% CI	.27, .32	.18, .22	.48, .53		.62, .66	.26, .31	.05, .10	
		6	*M*	.40	.18	.42		.65	.27	.08	
			SD	.27	.28	.29		.31	.19	.08	
			95% CI	.37, .42	.16, .21	.39, .44		.62, .67	.25, .29	.07, .09	
	Full	2	*M*	.43	.12	.45	48	.69	.24	.06	48
			SD	.29	.29	.30		.30	.28	.28	
			95% CI	.41, .46	.10, .50	.42, .47		.66, .72	.22, .27	.04, .09	
		6	*M*	.50	.13	.37		.70	.21	.08	
			SD	.29	.29	.30		.33	.17	.10	
			95% CI	.47, .52	.11, .16	.34, .40		.67, .73	.20, .23	.07, .09	

General Discussion

We aimed to test whether neglect of recollection contributed to false eyewitness identification decisions in simultaneous lineups. In two mini-lineup experiments, participants clearly made use of recollection when available, to counter the misleading familiarity of fillers and innocent suspects that were highly-similar to guilty suspects, and avoid false positive identification decisions. Just as clearly, participants were also willing to make error-prone familiarity-based positive identification decisions, when recollection was relatively weak. The overall pattern of identification responses indicated that participants were generally oriented to retrieve available recollection and adopted a recall-to-reject strategy (Rotello & Heit, 2000; see also Brainerd et al., 2001) or post-retrieval memory editing mechanism(s) that relied on the presence of recollection. Results are less consistent with use of the distinctiveness heuristic (Schacter, Israel & Racine, 1999, cf. S. A. Guerin, Robbins, Gilmore, & Schacter, 2012) that can increase accuracy in the absence of substantial recollection. These findings provide the first direct evidence of the separate contributions of familiarity and recollection to simultaneous lineup decisions (cf., Gronlund, 2005; N. Guerin et al., 2016a, 2016b; Meissner et al., 2005; Mickes, 2015; Palmer, Brewer, Weber & McKinnon, 2010).

The first major implication of these results is clear when they are set alongside our previous demonstration of reduced false identification with modifications to simultaneous lineup procedure and instructions designed to enhance eyewitnesses' use of recollection (N. Guerin et al., 2016b). Considered together, these results suggest three related conclusions. First, that in the full attention conditions in the current experiments, mock eyewitnesses used recollection in some degree to avoid making false identifications from standard simultaneous lineups. But, second, as we have shown that discriminability can be improved by interventions that facilitate the use of recollection, participants did not spontaneously use the full potential of recall-to-reject to increase the accuracy of their lineup decisions. Therefore, these findings are consistent with partial neglect of recollection in simultaneous lineups, or a failure to accurately weight the importance of recollective evidence, as predicted by our best-match-first dual-process framework (N. Guerin et al., 2016a). Finally, the two sets of results provide important evidence that eyewitnesses' use of recall-to-reject or other post-retrieval mechanisms could be improved by targeting task structure and metacognitive control, even when some spontaneous use of recollection was evident. It should be noted that asking participants to rate identification decision confidence had no effect on patterns of discriminability, even over multiple trials. Thus, there was no evidence that this simple, non-specific cue to metacognitive reflection would improve eyewitnesses' use of memory.[8] This finding is also consistent with evidence that decision confidence measured immediately after a response is based on the same memorial information as the recognition decision itself (e.g., van Zandt, 2000; Wixted, 2007), rather than extrinsic factors.

A second major implication of the greater discriminability observed with greater available recollection in the current experiments is that recall-to-reject (Rotello & Heit, 2000) was used, either as part of an overall orientation to retrieval for all test stimuli, or via retrieval monitoring processes that prompted more effortful recollection retrieval when ambiguous evidence of match was initially retrieved (e.g., Gray & Gallo, 2015; Jacoby et al., 1999). Any of several retrieval and post-retrieval monitoring mechanisms have been implicated in variable use of recollection in recognition (e.g., Gallo, 2010; Jacoby, Kelley, & McElree, 1999; Johnson, Hashtroudi, & Lindsay, 1993; Koriat & Goldsmith, 1996; Odegard & Lampinen, 2006; Schacter et al., 1999). Which of these mechanisms operate in specific tasks is influenced by immediate task-induced constraints and expectations (e.g., Cook et al., 2005;

---

[8] Though, there is some evidence that a salient *don't-know* response option (Weber & Perfect, 2012), free-report (Perfect & Weber, 2012) or a wildcard lineup member (Zajac & Karageorge, 2009) may enhance eyewitness use of memorial information.

Gray & Gallo, 2015, S. A. Guerin et al., 2012; Malmberg & Xu, 2007), participants' metamemorial understandings and preconceptions (e.g., Lane et al., 2007; Lane, Roussel, Starns, Villa, & Alonzo, 2008; Mather, Henkel, & Johnson, 1997), and their controlled use of different decision-making strategies (e.g., Blank & Launay, 2014; Jacoby et al., 1999; Starns et al., 2008). For the simultaneous lineup task investigated here, the higher rate of false identification when relatively less recollection was objectively available provides evidence against an effective contribution of the distinctiveness heuristic (Schacter et al., 1999). Schacter and colleagues (1999) proposed that when participants use the distinctiveness heuristic following retrieval, they are alerted to the possibility that high familiarity may be misleading by the relative absence of retrieved recollective detail. That is, participants compare retrieved evidence with that expected for a studied item. We observed higher false identification with less available recollection, suggesting that the distinctiveness heuristic was not, or could not, be successfully applied with those lineups. While we cannot rule out the possibility that the distinctiveness heuristic might be used in lineup tasks, the use of available recollection to avoid false identification that we observed is more consistent with recall-to-reject. In sum, further research is needed to identify the memorial basis and specific retrieval mechanisms involved in eyewitness decisions in simultaneous and other identification procedures and under varying conditions.

The applied importance of greater knowledge of the mechanisms underlying eyewitnesses' use of recollection and familiarity is apparent when we consider that many lineup decisions are made when recollection is poor or degraded (Brewer & Palmer, 2010). A low quality memory trace seriously constrains the kinds of procedural interventions that might improve eyewitnesses' metacognitive monitoring and control (cf., Perfect & Weber, 2012; Weber & Perfect, 2012). There is evidence that recollection degrades to a greater extent than familiarity, as well as being less likely to be effectively encoded and retained when attentional resources are overloaded (Yonelinas, 2002). Therefore, if multiple retrieval or post-retrieval mechanisms contribute to the use of available recollection in lineup decisions, it will be important to identify those that are most useful when memory is weak and recollection most difficult to retrieve. More important, perhaps, is targeted investigation of methods for aligning eyewitness strategies of metacognitive control with the known limits of familiarity-based recognition, when no recollection is available.

While we found strong evidence of the use of recollection in the lineup task in these experiments, it is possible that recall-to-reject was elicited by features peculiar to the mini-lineup paradigm, rather than being characteristic of lineup tasks in all paradigms and real contexts. For example, task difficulty has been demonstrated to prompt use of recollection in other recognition paradigms (e.g., Cook et al., 2005; Malmberg, 2008). Studying multiple targets (rather than only one or two offenders) and being required to respond to multiple test lineup trials could have led participants to perceive this task as particularly difficult, and encouraged focused and effortful recollection. Perceptions of substantial target-foil similarity within mini-lineups, or the specific demand of matching each test lineup to a given studied target, might have had a similar effect. In a more realistic identification paradigm or for a real identification test, eyewitnesses are asked to respond to a single lineup without these features, and hence would not be prompted by these task features to make full use of available recollection. Therefore, an immediate question for future research is to investigate whether our findings generalise to a more veridical, single-trial lineup paradigm. It should be noted, though, that even if eyewitness recall-to-reject was a by-product of the mini-lineup paradigm, use of the presence-recollection procedure in the mini-lineup paradigm improved discriminability still further (N. Guerin et al., 2016a). Therefore, the procedure would hold even greater potential for improving identification performance in contexts where participants tend to greater neglect of recollection.

In summary, we found substantial use of recall-to-reject by eyewitnesses in a simultaneous mini-lineup task, when recollection was available. However, when considered in the broader research context, our results suggest that mock eyewitnesses' retrieval or post-retrieval evaluation of available recollection was not fully optimal, leaving space for further potential improvement with procedural interventions. These findings demonstrate the usefulness of extending the dual-process account of recognition to lineup tasks for explaining patterns of identification responses and potentially improving identification test procedures. Future research should address the relative efficacy of specific retrieval and memory monitoring processes in both simultaneous and other lineup formats under diverse conditions. Of particular importance is further investigation of decision architecture and metacognitive instructions to establish the limits of conscious control over recognition decision-making, and minimise the unwittingly harmful influence of witnesses' idiosyncratic metacognitive approaches on measures of suspect guilt.

References

Alban, M. W., & Kelley, C. M. (2012). Variations in constrained retrieval. *Memory & Cognition*, *40*, 681-692. doi: 10.3758/s13421-012-0185-5

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390-412. doi: 10.1016/j.jml.2007.12.005

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structures for confirmatory hypothesis testing: Keeping it maximal. *Journal of Memory and Language*, 68, 255-278. doi: 10.1016/j.jml.2012.11.001

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2014). _lme4: Linear mixed-effects models using Eigen and S4_. Retrieved from http://CRAN.R-project.org/package=lme4

Benjamin, A. S. (2001). On the dual effects of repetition on false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 941-947. doi: 10.1037/0278-7393.27.4.941

Blank, H., & Launay, C. (2014). How to protect eyewitness memory against the misinformation effect: A meta-analysis of post-warning studies. *Journal of Applied Research in Memory and Cognition*, *3*, 77-88. doi: 10.1016/j.jarmac.2014.03.005

Brainerd, C. J., Reyna, V. F., Wright, R., & Mojardin, A. H. (2003). Recollection rejection: False-memory editing in children and adults. *Psychological Review*, *110*, 762-784. doi: 10.1037/0033-295x.110.4.762

Brewer, N., & Palmer, M. A. (2010). Eyewitness identification tests. *Legal and Criminological Psychology*, *15*, 77-96. doi: 10.1348/135532509x414765

Brewer, N., & Weber, N. (2008). Eyewitness confidence and latency: Indices of memory processes not just markers of accuracy. *Applied Cognitive Psychology*, *22*, 827-840. doi: 10.1002/acp.1486

Brewer, N., & Wells, G. L. (2006). The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology: Applied, 12*, 11-30. doi: 10.1037/1076-898X.12.1.11

Brewer, N., & Wells, G. L. (2011). Eyewitness identification. *Current Directions in Psychological Science*, *20,* 24-27. doi: 10.1177/0963721410389169

Brigham, J. C., Meissner, C. A., & Wasserman, A. W. (1999). Applied issues in the construction and expert assessment of photo lineups. *Applied Cognitive Psychology*, *13*, S73-S92. doi: 10.1002/(sici)1099-0720(199911)13:1+3.3.co;2-w

Charman, S. D., Gregory, A. H., & Carlucci, M. (2009). Exploring the diagnostic utility of facial composites: Beliefs of guilt can bias perceived similarity between composite and suspect. *Journal of Experimental Psychology: Applied*, *15*, 76-90. doi: 10.1037/a0014682

Clark, S. E. (2003). A memory and decision model for eyewitness identification. *Applied Cognitive Psychology*, *17,* 629-654. doi: 10.1002/acp.891

Clark, S. E. (2005). A re-examination of the effects of biased lineup instructions in eyewitness identification. *Law and Human Behavior*, *29*, 575-604. doi: 10.1007/s10979-005-7121-1

Clark, S. E., & Godfrey, R. D. (2009). Eyewitness identification evidence and innocence risk. *Psychonomic Bulletin & Review*, *16,* 22-42. doi: 10.3758/pbr.16.1.22

Clark, S. E., Howell, R. T., & Davey, S. L. (2008). Regularities in eyewitness identification. *Law and Human Behavior*, *32*, 187-218. doi: 10.1007/s10979-006-9082-4

Clark, S. E., & Tunnicliff, J. L. (2001). Selecting lineup foils in eyewitness identification experiments: Experimental control and real-world simulation. *Law and Human Behavior, 25,* 199-216. doi: 10.1023/a:1010753809988

Clark, S. E., & Wells, G. L. (2008). On the diagnosticity of multiple-witness identifications. *Law and Human Behavior*, *32*, 406-XX. doi: 10.1007/s10979-007-9115-7

Cook, G. I., Marsh, R. L., & Hicks, J. L. (2005). Revisiting the role of recollection in item versus forced-choice recognition memory. *Psychonomic Bulletin & Review*, *12*, 720-725. doi: 10.3758/bf03196763

Cutler, B. L., & Penrod, S. D. (1995). *Mistaken identification: The eyewitness, psychology and the law*. Cambridge University Press. doi: 10.1007/978-1-4615-4891-1

De Carlo, L. T. (1998). Signal detection theory and generalized linear models. *Psychological Methods, 3*, 186-205. doi: 10.1037//1082-989x.3.2.186

Deffenbacher, K. A., Bornstein, B. H., & Penrod, S. D. (2006). Mugshot exposure effects: Retroactive interference, mugshot commitment, source confusion, and unconscious transference. *Law and Human Behavior*, *30*, 287-307. doi: 10.1007/s10979-006-9008-1

Dyne, A. M., Humphreys, M. S., Bain, J. D., & Pike, R. (1990). Associative effects in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 813-824. doi: 10.1037/0278-7393.16.5.813

Fitzgerald, R. J., Price, H. L., Oriet, C., & Charman, S. D. (2013). The effect of suspect-filler similarity on eyewitness identification decisions: A meta-analysis. *Psychology, Public Policy, and Law*, *19*, 151-164. doi: 10.1037/a0030618

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition*, *38*, 833-848. doi: 10.3758/mc.38.7.833

Gallo, D. A., Bell, D. M., Beier, J. S., & Schacter, D. L. (2006). Two types of recollection-based monitoring in younger and older adults: Recall-to-reject and the distinctiveness heuristic. *Memory, 14*, 730-741. doi: 10.1080/09658210600648506

Gallo, D. A., Sullivan, A. L., Daffner, K. R., Schacter, D. L., & Budson, A. E. (2004). Associative recognition in Alzheimer's disease: evidence for impaired recall-to-reject. *Neuropsychology, 18*(3), 556-563. doi: 10.1037/0894-4105.18.3.556

Gardiner, J. M., & Parkin, A. J. (1990). Attention and recollective experience in recognition memory. *Memory & Cognition, 18*, 579-583. doi: 10.3758/bf03197100

Garrett, B. (2011). *Convicting the Innocent*: *Where Criminal Prosecutions Go Wrong*. Harvard University Press. Cambridge: Harvard.

Gelman, A., & Hill, J. (2007). *Data Analysis using Regression and Multilevel/Hierarchical Models.* New York: Cambridge University Press.

Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness, 5*(2), 189-211. doi: 10.1080/19345747.2011.618213

Goodsell, C. A., Neuschatz, J. S., & Gronlund, S. D. (2009). Effects of mugshot commitment on lineup performance in young and older adults. *Applied Cognitive Psychology, 23*, 788-803. doi: 10.1002/acp.1512

Gray, S. J., & Gallo, D. A. (2015). Disregarding familiarity during recollection attempts: Content-specific recapitulation as a retrieval orientation strategy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 134-147. doi: 10.1037/a0038363

Gronlund, S. D. (2005). Sequential lineup advantage: Contributions of distinctiveness and recollection. *Applied Cognitive Psychology,* 19, 23-37. doi: 10.1002/acp.1047

Gronlund, S. D., Mickes, L., Wixted, J. T., & Clark, S. E. (2015). Conducting an eyewitness lineup: How the research got it wrong. In B. H. Ross (Ed.), The psychology of learning and motivation (Volume 63, pp. 1– 43). Waltham, MA: Academic Press. doi: 10.1016/bs.plm.2015.03003

Gronlund, S. D., & Ratcliff, R. (1989). Time course of item and associative information: implications for global memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 846-858. doi: 10.1037/0278-7393.15.5.846

Guerin, N., Weber, N., & Horry, R. (2016a). Split decision: An account of eyewitness decision-making in simultaneous lineups and a novel presence-first identification procedure. *Under review*

Guerin, N., Weber, N., & Horry, R. (2016b). Only recollect: Improving identification accuracy by reframing eyewitness metacognition and decision structure. *Under review*

Guerin, S. A., Robbins, C. A., Gilmore, A. W., & Schacter, D. L. (2012). Retrieval failure contributes to gist-based false recognition. *Journal of Memory and Language*, *66*, 68-78. doi: 10.1016/j.jml.2011.07.002

Hancock, P. J., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences*, *4*, 330-337. doi: 10.1016/s1364-6613(00)01519-9

Higham, P. A., Perfect, T. J., & Bruno, D. (2009). Investigating strength and frequency effects in recognition memory using type-2 signal detection theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 57-80. doi: 10.1037/a0013865

Hintzmann, D. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, 95, 528–551. doi: 10.1037//0033-295x.95.4.528

Hintzmann, D.L. & Curran, T. (1994) Retrieval dynamics of recognition and frequency judgements: Evidence for separate processes of familiarity and recall. *Journal of Memory and Language,* 33, 1-18. doi: 10.1006/jmla.1994.1001

Horry, R., Memon, A., Wright, D. B., & Milne, R. (2011). Predictors of eyewitness identification decisions from video lineups in England: A field study. *Law and Human Behavior, 36,* 257-265. doi: 1007/s10979-011-9279-z

Horry, R., Wright, D. B., & Tredoux, C. G. (2010). Recognition and context memory for faces from own and other ethnic groups: A remember-know investigation. *Memory & Cognition, 38*(2), 134-141. doi: 10.3758/mc.38.2.134

Ingram, K. M., Mickes, L., & Wixted, J. T. (2012). Recollection can be weak and familiarity can be strong. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 325-339. doi: 10.1037/e520592012-514

Innocence Project (2016). Innocence project. Retrieved February 20, 2016. http://www.innocenceproject

Jacoby, L. L., Kelley, C. M., & McElree, B. D. (1999). The role of cognitive control: Early selection versus late correction. In. S. Chaiken & Y. Trope (Eds.), Dual-process theories in social psychology (pp. 383– 400). New York, NY: Guilford Press.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114,* 3-28. doi: 10.1037//0033-2909.114.1.3

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology, 103*(1), 54-69. doi: 10.1037/a0028347

Kelley, R., & Wixted, J. T. (2001). On the nature of associative information in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27,* 701-722. doi: 10.1037/0278-7393.27.3.701

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*, 490-517. doi: 10.1037/0033-295x.103.3.490

Lampinen, J. M., Neuschatz, J. S., & Payne, D. G. (1999). Source attributions and false memories: A test of the demand characteristics account. *Psychonomic Bulletin & Review*, *6*, 130-135. doi: 10.3758/bf03210820

Lampinen, J. M., Odegard, T. N., & Neuschatz, J. S. (2004). Robust recollection rejection in the memory conjunction paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30,* 332-342. doi: 10.1037/0278-7393.30.2.332

Lampinen, J. M., Watkins, K. N., & Odegard, T. N. (2006). Phantom ROC: Recollection rejection in a hybrid conjoint recognition signal detection model. *Memory*, *14,* 655-671. doi: 10.1080/09658210600648431

Lane, S. M., & Meissner, C. A. (2008). A ‘middle road’ approach to bridging the basic–applied divide in eyewitness identification research. *Applied Cognitive Psychology*, *22,* 779-787. doi: 10.1002/acp.1482

Lane, S. M., Roussel, C. C., Villa, D., & Morita, S. K. (2007). Features and feedback: Enhancing metamnemonic knowledge at retrieval reduces source-monitoring errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33,* 1131-1142. doi: 10.1037/0278-7393.33.6.1131

Lane, S. M., Roussel, C. C., Starns, J. J., Villa, D., & Alonzo, J. D. (2008). Providing information about diagnostic features at retrieval reduces false recognition. *Memory, 16*, 836–851. http://doi: 10.1080/0965821080233734

Luus, C. A., & Wells, G. L. (1991). Eyewitness identification and the selection of distracters for lineups. *Law and Human Behavior, 15,* 43. doi: 10.1007/bf01044829

Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology, 57*, 335-384. doi: 10.1016/j.cogpsych.2008.02.004

Malmberg, K. J., & Xu, J. (2007). On the flexibility and the fallibility of associative memory. *Memory & Cognition*, *35,* 545-556. doi: 10.3758/bf03193293

Malpass, R. S., & Devine, P. G. (1981). Eyewitness identification: Lineup instructions and the absence of the offender. *Journal of Applied Psychology*, *66*, 482. doi: 10.1037/0021-9010.66.4.482

Malpass, R. S., & Lindsay, R. C. L. (1999). Measuring lineup fairness. *Applied Cognitive Psychology*, *13,* S1-S7. doi: 10.1002/(sici)1099-0720(199911)13:1+3.0.co;2-9

Mäntylä, T., & Cornoldi, C. (2002). Remembering changes: Repetition effects in face recollection. *Acta Psychologica, 109,* 95-105. doi: 10.1016/s0001-6918(01)00054-3

Marcon, J. L., Susa, K. J., & Meissner, C. A. (2009). Assessing the influence of recollection and familiarity in memory for own-versus other-race faces. *Psychonomic Bulletin & Review*, *16*, 99-103. doi: 10.3758/pbr.16.1.99

Mather, M., Henkel, L. A., & Johnson, M. K. (1997). Evaluating characteristics of false memories: Remember/know judgments and memory characteristics questionnaire compared. *Memory & Cognition*, *25*(6), 826-837. doi: 10.3758/bf03211327

Meissner, C. A., Tredoux, C. G., Parker, J. F., MacLin, O. H. (2005). Eyewitness decisions in simultaneous and sequential lineups: A dual-process signal detection theory analysis. *Memory & Cognition,* 33, 783-792. doi: 10.3758/bf03193074

Mickes, L. (2015). Receiver operating characteristic analysis and confidence-accuracy characteristic analysis in investigations of system variables and estimator variables that affect eyewitness memory. *Journal of Applied Research in Memory and Cognition, 4,* 83-102. doi: 10.1016/j.jarmac.2015.01.003

Mickes, L., Johnson, E. M., & Wixted, J. T. (2010). Continuous recollection versus unitized familiarity in associative recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 843-863. doi: 10.1037/a0019755

Mickes, L., Wais, P. E., & Wixted, J. T. (2009). Recollection is a continuous process: Implications for dual-process theories of recognition memory. *Psychological Science, 20*, 509-515. doi: 10.1111/j.1467-9280.2009.02324.x

Migo, E., Montaldi, D., Norman, K. A., Quamme, J., & Mayes, A. (2009). The contribution of familiarity to recognition memory is a function of test format when using similar foils. *The Quarterly Journal of Experimental Psychology*, *62*, 1198-1215. doi: 10.1080/17470210802391599

Murayama, K., Sakaki, M., Yan, V. X., & Smith, G. M. (2014). Type I error inflation in the traditional by-participant analysis to metamemory accuracy: A generalized mixed-effects model perspective. *Journal of Experimental Psychology: Learning Memory and Cognition*, *40*, 1287-1306. doi: 10.1037/a0036914

Nobel, P. A., & Shiffrin, R. M. (2001). Retrieval processes in recognition and cued recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 384-413. doi: 10.1037/0278-7393.27.2.384

Odegard, T. N., & Lampinen, T. N. (2006). Memory editing: Knowledge, criteria, and alignment. *Memory*, *14,* 777-787. doi: 10.1080/09658210600648589

Palmer, M. A., Brewer, N., McKinnon, A. C., & Weber, N. (2010). Phenomenological reports diagnose accuracy of eyewitness identification decisions. *Acta Psychologica*, *133*, 137-145. doi: 10.1016/j.actpsy.2009.11.002

Palmer, M. A., Brewer, N., Weber, N., & Nagesh, A. (2013). The confidence-accuracy relationship for eyewitness identification decisions: Effects of exposure duration, retention interval, and divided attention. *Journal of experimental psychology: applied*, *19*, 55-71. doi: 10.1037/a0031602

Parkin, A. J., Gardiner, J. M., & Rosser, R. (1995). Functional aspects of recollective experience in face recognition. *Consciousness and Cognition*, *4*, 387-398. doi: 10.1006/ccog.1995.1046

Perfect, T. J., & Weber, N. (2012). How should witnesses regulate the accuracy of their identification decisions: One step forward, two steps back? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1810-1818. doi: 10.1037/a0028461

R Core Team. (2014). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org.

Reinitz, M. T., Morrissey, J., & Demb, J. (1994). Role of attention in face encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 161. doi: 10.1037/0278-7393.20.1.161

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and individual Differences*, *7*(1), 1-75. doi: 10.1016/1041-6080(95)90031-4

Rotello, C. M., & Heit, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory & Cognition*, *28*, 907-922. doi: 10.3758/bf03209339

Rotello, C. M., Macmillan, N. A., & Van Tassel, G. (2000). Recall-to-reject in recognition: Evidence from ROC curves. *Journal of Memory and Language*, *43*, 67-88. doi: 10.1006/jmla.1999.2701

Sauer, J. D., Brewer, N., & Weber, N. (2008). Multiple confidence estimates as indices of eyewitness memory. *Journal of Experimental Psychology: General*, 137, 528-547. doi: 10.1037/a0012712

Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language, 40,* 1-24. doi: 10.1006/jmla.1998.2611

Schreiber, T. A., & Sergent, S. D. (1998). The role of commitment in producing misinformation effects in eyewitness memory. *Psychonomic Bulletin & Review*, *5*(3), 443-448. doi: 10.3758/bf03208819

Semmler, C., Brewer, N., & Douglass, A. B. (2011). Jurors believe eyewitnesses. In B. L. Cutler (Ed.), *Conviction of the innocent: Lessons from psychological research* (pp. 185–209). Washington, DC: American Psychological Association. doi: 10.1037/13085-009

Starns, J. J., Lane, S. M., Alonzo, J. D., & Roussel, C. C. (2007). Metamnemonic control over the discriminability of memory evidence: A signal detection analysis of warning effects in the associative list paradigm. *Journal of Memory and Language*, *56*, 592-607. doi: 10.1016/j.jml.2006.08.013

Steblay, N. K., Dysart, J. E., & Wells, G. L. (2011). Seventy-two tests of the sequential lineup superiority effect: A meta-analysis and policy discussion. *Psychology, Public Policy, and Law*, *17*, 99-139. doi: 10.1037/a0021650

Tredoux, C. (2002). A direct measure of facial similarity and its relation to human similarity perceptions. *Journal of Experimental Psychology: Applied*, *8*, 180-193. doi: 10.1037/1076-898x.8.3.180

Tunnicliff, J. L., & Clark, S. E. (2000). Selecting foils for identification lineups: Matching suspects or descriptions. *Law and Human Behavior, 24,* 231-258. doi: 10.1023/a:1005463020252

Van Zandt, T. (2000). ROC curves and confidence judgments in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 582-600. doi: 10.1037/0278-7393.26.3.582

Weber, N., & Brewer, N. (2004). Confidence–accuracy calibration in absolute and relative face recognition judgments. *Journal of Experimental Psychology: Applied, 10,* 156–172. doi: 10.1037/1076-898x.10.3.156

Weber, N., & Brewer, N. (2006). Positive versus negative face recognition decisions: Confidence, accuracy and response latency. *Applied Cognitive Psychology, 20,* 17–31. doi: 10.1002/acp.1166

Weber, N., & Perfect, T. J. (2012). Improving eyewitness identification accuracy by screening out those who say they don't know. *Law and Human Behavior, 36*, 28-36. doi: 10.1037/h0093976

Wells, G. L. (1984). The Psychology of Lineup Identifications1. *Journal of Applied Social Psychology, 14,* 89-103. doi: 10.1111/j.1559-1816.1984.tb02223.x

Wells, G. L. (1993). What do we know about eyewitness identification? *American Psychologist*, *48*, 553-571. doi: 10.1037//0003-066x.48.5.553

Wells, G. L. (2008). Field experiments on eyewitness identification: Towards a better understanding of pitfalls and prospects. *Law and Human Behavior, 32*, 6-10. doi: 10.1007/s10979-007-9098-4

Wells, G. L., & Olson, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology*, *54*, 277-295. doi: 10.1146/annurev.psych.54.101601.145028

Wells, G. L., & Seelau, E. P. (1995). Eyewitness identification: Psychological research and legal policy on lineups. *Psychology, Public Policy, and Law,1,* 765-791. doi: 10.1037/1076-8971.1.4.765

Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S.M., & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behaviour, 22,* 603-647. doi: 10.1023/a:1025750605807

Wells, G. L., Steblay, N. K., & Dysart, J. E. (2015). Double-blind photo lineups using actual eyewitnesses: An experimental test of a sequential versus simultaneous lineup procedure. *Law and Human Behavior*, *39*, 1-14. doi: 10.1037/lhb0000096

Westfall, J., Judd, C. M., & Kenny, D. A. (2015). Replicating studies in which samples of participants respond to samples of stimuli. *Perspectives on Psychological Science, 10,* 390-399. doi: 10.1177/1745691614564879

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review, 114,* 152-176. doi: 10.1037/0033-295x.114.1.152

Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review*, *117,* 1025-1054. doi: 10.1037/a0020874

Wixted, J. T., Mickes, L., Clark, S. E., Gronlund, S. D., & Roediger III, H. L. (2015). Initial eyewitness confidence reliably predicts eyewitness identification accuracy. *American Psychologist, 70,* 515-526. doi: 10.1037/a0039510

Wogalter, M. S., Malpass, R. S., & McQuiston, D. E. (2004). A national survey of US police on preparation and conduct of identification lineups. *Psychology, Crime and Law*, *10,* 69-82. doi: 10.1080/10683160410001641873

Wright, D.B., & London, K. (2009). Multilevel modelling: Beyond the basic applications. *British Journal of Mathematical and Statistical Psychology,* 62, 439-456. doi: 10.1348/000711008X327632

Xu, J., & Malmberg, K. J. (2007). Modeling the effects of verbal and nonverbal pair strength on associative recognition. *Memory & Cognition*, *35*, 526-544. doi: 10.3758/bf03193292

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*, 441-517. doi: 10.1006/jmla.2002.2864

Zajac, R., & Karageorge, A. (2009). The wildcard: A simple technique for improving children's target-absent lineup performance. *Applied Cognitive Psychology*, *23*, 358-368. doi: 10.1002/acp.1511

CHAPTER 6

Deceived by the obvious: Metacognitive instructions target neglect of recollection in eyewitness identification decisions[1]

Abstract

We tested a dual-process account of the effectiveness of the novel presence-recollection lineup in enhancing discriminability relative to a standard simultaneous lineup, without significantly reducing the likelihood of accurate identification of a guilty culprit. Specifically, we proposed that eyewitnesses do not make optimal use of recollection in standard simultaneous lineup tasks. To counter this, the presence-recollection procedure prompts eyewitnesses to more effortfully retrieve and weight available recollection to increase the accuracy of their decisions. In a single face-recognition mini-lineup experiment, participants assigned to the presence-recollection procedure were given instructions designed to enhance their metamemorial knowledge of the diagnostic features of accurate recollection. Participants then made lineup decisions in two steps: first indicating only whether or not the culprit was in the lineup, before pointing out the culprit, if judged to be present. In addition, we manipulated target familiarity within-subjects by varying target study repetitions, and varied recollection between-subjects with a divided attention manipulation. Discriminability was greater for the presence-recollection procedure than the simultaneous procedure at all levels of repetitions and attention. We conclude that the presence-recollection procedure enhanced witnesses' use of recollection, even when relatively little recollective detail was

[1] Under review at *Memory & Cognition* as Guerin, N., Weber, N., & Horry, R. (2016). Deceived by the obvious: Metacognitive instructions target neglect of recollection in eyewitness identification decisions. As first author, I designed and conducted the experiments and analyses, interpreted results, and wrote the manuscript. Nathan Weber, as my PhD supervisor, advised on the experimental and analytical design, and contributed to the interpretation of results and manuscript editing. Ruth Horry, as Adjunct Supervisor, contributed to the interpretation of results and the manuscript.

encoded. When greater recollective detail had been encoded, witnesses using a standard procedure made some use of recall-to-reject to avoid false identifications, but to a lesser extent than those using the presence-recollection lineup. Results demonstrate the potential usefulness of the modified procedure to improve the reliability of identification evidence, even when limited recollective detail has been encoded or retained.

Introduction

An innocent suspect in a traditional simultaneous police identification lineup faces a well-documented risk of mistaken identification (e.g., Horry, Memon, Wright, & Milne, 2011; Wells, Steblay, & Dysart, 2015). Despite abundant evidence for the shortcomings of eyewitness memory, positive identification evidence is still frequently the sole evidence for conviction (Gross & Schaffer, 2012). Presenting lineup members one at a time in a sequential lineup format has been widely recommended to lower risk to the innocent (e.g., Wells et al., 1998). However, while sequential lineups often discourage overly liberal response bias in eyewitnesses, there is no consistent evidence that it otherwise promotes better use of memorial information than simultaneous presentation (Clark, 2012; Gronlund, Carlson, Daily, & Goodsell, 2009; Mickes, Flowe, & Wixted, 2012). We designed the presence-recollection lineup to improve eyewitnesses' use of recollection, and so improve identification accuracy. The novel procedure yielded greater discriminability than standard simultaneous lineups, without significantly or meaningfully reducing hits (N. Guerin, Weber, & Horry, 2016b). However, there was no direct evidence that recollection was instrumental in the observed improvement. Here, we aimed to directly test whether the presence-recollection advantage arose from greater use of recollection.

We extended a continuous dual-process account of recognition memory (e.g., Wixted & Mickes, 2010) to simultaneous lineup decision-making. Dual-process theories propose that two distinct types of memorial information can contribute to recognition: global familiarity, and specific recollection (e.g., Yonelinas, 2002). Continuous models further propose that both familiarity and recollection vary continuously, and in their relative contribution to recognition (e.g., Ingram, Mickes, & Wixted, 2012). We (N. Guerin, Weber, & Horry, 2016a) proposed that standard simultaneous lineups encourage an over-reliance on familiarity-based evidence of positive match to the offender, and inadequate use of recollection that would increase accuracy when the offender is not in the lineup.

The contributions of familiarity and recollection to recognition vary with task type and context (e.g., Malmberg, 2008). Target recollection is vital to avoid familiarity-based false recognition in tasks with high target-foil similarity (i.e., recall-to-reject, Rotello & Heit, 2000). However, available recollection is not always used (e.g., Dyne, Humphries, Bain & Pike, 1990; Malmberg & Xu, 2007). Participants can neglect recollection when task expectations are misaligned with task demands, and the misleading familiarity of highly-similar foils is not anticipated (e.g. Gallo, 2010; Odegard, Lampinen, & Toglia, 2005). In lineup tasks, known-innocent fillers and innocent suspects closely resemble the offender (e.g., Clark & Tunnicliff, 2001; Wogalter, Malpass, & McQuiston, 2004). Thus, lineups feature substantial target-foil similarity and call for recall-to-reject. We argue that recollection is neglected in simultaneous lineups, contributing to the risk of false identification (N. Guerin et al., 2016a).

Our best-match-first dual-process framework (N. Guerin et al., 2016a) combines this dual-process account with a best-match-first account of simultaneous lineup decision-making strategies. We extended Wells' (1984, 1993, see also Clark, 2003) relative judgment idea to argue that eyewitnesses are predisposed to find the offender in a lineup. When lineup members are presented simultaneously, this leads eyewitnesses to concentrate on finding the best-matching lineup member to memory for the offender, before deciding (with a lenient criterion) whether the best-match is the offender. We proposed that eyewitnesses rely on familiarity alone to find the best-match, and then continue to rely on this already-retrieved evidence to make their second offender-presence judgment, without effortfully retrieving potentially available recollection (e.g., Alban & Kelley, 2012). Thus, potentially disconfirming recollection is neglected, heightening eyewitnesses' susceptibility to false identification (N. Guerin et al., 2016a). In the presence-recollection lineup (N. Guerin et al., 2016b), we applied two procedural modifications to disrupt best-match-first decision-making and encourage greater recollection: a presence-first decision structure; and metacognitive instructions.

In the presence-recollection lineup, participants make their identification decisions in two steps that reverse the error-prone best-match-first order of component recognition decisions (N. Guerin et al., 2016b). First, they indicate only whether or not the offender is present in the lineup; and, second, if indicated present, they identify the offender. Eyewitnesses' first judgment is thus akin to a yes/no recognition task, for which both positive and negative evidence of match is likely to be considered in challenging tasks (e.g., Cook, Marsh, & Hicks, 2005). To counter eyewitnesses' error-prone misconceptions about lineup tasks (e.g., Clark, 2005; Malpass & Devine, 1981), we provided metacognitive information about how the phenomenology of memories can be used to diagnose their accuracy (i.e., relative vividness, clarity, and detail). Phenomenological instructions of this kind shave been shown to improve recollection in tasks that feature high target-foil similarity (e.g., Lane, Roussel, Starns, Villa, & Alonzo, 2008; Starns, Lane, Alonzo, & Roussel, 2007). We tested the presence-recollection lineup in a mini-lineup paradigm in which participants studied multiple target face photos before making identification decisions for a four-person lineup for each target (N. Guerin et al., 2016b). Discriminability was significantly greater with the presence-recollection procedure than a standard simultaneous procedure. Importantly, the presence-first decision structure alone did not improve discriminability with non-naïve participants, and metacognitive instructions were ineffective when provided with simultaneous lineups in which best-match-first decision-making was not disrupted. Both elements were necessary to improve discriminability (N. Guerin et al., 2016b).

We also used the mini-lineup paradigm to investigate use of recollection in standard simultaneous lineups (N. Guerin et al., 2016c). There is little other evidence for the use of familiarity and recollection in lineup tasks, and the few available studies rely mainly on self-report measures (cf., Gronlund, 2005; Meissner, Tredoux, Parker, & Maclin, 2005; Mickes, 2015; Palmer, Brewer, McKinnon, & Weber, 2010). Instead, we varied objectively-available familiarity and recollection orthogonally, and compared patterns of identification responses between experimental conditions. Massed repetitions have been shown to manipulate the strength of familiarity in face-recognition paradigms, without significantly affecting recollection (Mantyla & Cornoldi, 2002; Parkin, Gardiner, & Rosser, 1995). Thus, we created two levels of target familiarity by varying the number of presentations of target face photos during study (repetitions: two; six). Conversely, dividing attention has been shown to decrease recollection without influencing familiarity to the same degree (e.g., Gardiner & Parkin, 1990; Reinitz, Morrissey, & Demb, 1994). We manipulated study attention between-subjects (divided; full) to create two levels of available recollection. Contrary to predictions, participants used recollection, when available, to avoid false identification. How can this finding be reconciled with the greater discriminability previously observed with the presence-recollection lineup? We proposed that while participants used recollection to some degree in standard simultaneous lineups, it was likely that they did not exploit its full potential to maximise accuracy. Thus, relative to standard simultaneous lineups, the presence-recollection lineup further enhanced the use of recollection.

The Experiment

We used the same manipulations of study repetitions and attention to vary familiarity and recollection, respectively. Lineup type was varied between-subjects (standard simultaneous; presence-recollection). We could thus dissociate and compare participants' use of familiarity and recollection in the two procedures. We aimed to test whether the previously reported presence-recollection advantage (N. Guerin et al., 2016b) was due to greater use of recollection than for a standard simultaneous procedure. This was the first investigation of the mechanisms underlying the presence-recollection advantage. Therefore, we used the mini-lineup paradigm to collect multiple data points per participant for a large sample of stimuli to gain greater generalisability than with an identification paradigm (Wells & Windschitl, 1999; Westfall, Judd, & Kenny, 2015).

We expected participants in both procedures to use available recollection, producing greater discriminability with full than divided attention. We also predicted better use of recollection for presence-recollection lineups than standard lineups, demonstrated by a larger effect of attention on discriminability.

Method

Participants and Design

We used a 2 (lineup: standard; presence-recollection) × 2 (attention: divided; full) × 2 (repetitions: two; six) × 2 (target presence: target-present; target-absent) mixed experimental design, with lineup and attention varied between-subjects, and repetitions manipulated within-subjects. For each participant, test mini-lineups were randomly assigned to target-presence status within repetitions.

We aimed to recruit 160 participants[2], and 162 undergraduate students from Flinders University participated for course credit or payment (122 female, age: M = 21.85, SD = 6.52). All collected data were analysed.

Materials

We used 80 mini-lineup stimulus-sets comprising description-matched colour photos of five faces. In 40 mini-lineups (Weber & Brewer, 2004, 2006), each photo showed a front view of the head and neck with neutral facial expression and background (200 × 200 pixels). Each set included a randomly designated target, target-replacement (for target-absent mini-lineups), and three fillers. The remaining 40 lineups (N. Guerin et al., 2016b) were similar in all respects, except that target photo poses differed for study and test phases. Study-test variation in pose makes recognition more challenging, analogous to encoding-test differences experienced by eyewitnesses (e.g., Hancock, Bruce, & Burton, 2000). Test target photos were as described for the original set, while study target photos showed either a front view of the head and neck with a different facial expression (smiling), or a three-quarter view of the head and neck with the same facial expression. Mini-lineup sets were divided into four blocks of 20. In each block, half of the tests were target-present and half target-absent. Stimuli were presented on an 18” monitor (resolution: 1920 × 1080 pixels). All mini-lineups were presented in a row of four, with response buttons below. A text cue (name or occupation, by block) was presented with each corresponding target (study) and mini-lineup (test), centred above photos. Photo order in each mini-lineup, trial order in each block, cue and target-presence assignment, and collection order, were randomly determined for each participant.

[2] Guided by power estimates to test hypotheses with mixed-effects models (Gelman & Hill, 2007) and estimates for moderate effect size, as informed by previous findings (N. Guerin et al., 2016a, 2016b, 2016c).

Procedure

After giving informed consent, participants completed the experiment on PC in individual cubicles using a mouse to click on-screen buttons in response to written prompts. They completed four blocks of trials (after four practise trials), with block order, and assignment of name or occupation cues to blocks, counterbalanced. Each block included a study phase, a visual distractor task, and a test phase. In the study phase, target photos were presented sequentially, in the centre of the screen. Each cue was presented for 1000 ms before presentation of the target, and throughout all repetitions of that target (500 ms per repetition with an inter-stimulus interval of 1000 ms). Following each study phase, participants worked on a spatial memory task for 3 min. In each test phase, participants were given unbiased instructions that each mini-lineup may or may not include a target (Malpass & Devine, 1981) and completed 20 mini-lineup trials. Response latency was recorded for all decisions.

Repetitions. For all participants, half of the target photos were presented for study with two massed (consecutive) repetitions and the remainder with six. Targets were assigned randomly to repetitions condition for each participant. Participants were informed that photos would be presented a variable number of times.

Attention. We varied attention between-subjects on two levels (divided; full) during study. For all participants, the presentation of targets was accompanied by audio tones randomised for pitch (high, low) and intervening interval (1000 or 2000 ms). In the divided attention condition, participants responded to tones with a mouse click (left-button for high, right for low). In the full attention condition, participants were told to ignore the audio tones. The attention manipulation was included in the practice trials.

Lineup. We tested two identification lineup procedures, between-subjects. In the standard lineup condition, participants provided an identification response for each mini-lineup by identifying a lineup member as the studied target (by clicking on the photo), or clicking the *Not Present* button.

In the presence-recollection lineup condition, following the procedure reported previously (N. Guerin et al., 2016b, Experiment 3), participants made two blocked identification decisions for each mini-lineup. Participants first completed two blocks of trials making only their first decision for each mini-lineup (though they were informed that they would later be required to make a second, target identification decision). For this first presence decision, they indicated whether the studied face was *Present* or *Not Present* in the mini-lineup. After all trials were completed, participants were presented with the series of lineups for which they had made a positive presence decision (*Present)* and asked to click on the face identified as the target. This procedure was repeated for the final two blocks of trials.

In addition, presence-recollection lineup participants were given metacognitive instructions that included information about the diagnostic phenomenology of accurate remembering (full phenomenological instructions are included in an online supplement).[3] Participants were asked to use this information to improve the accuracy of identification decisions and were reminded of these instructions at the start and half-way points of each test block.

Results

Overview

We used an alpha level of .05 for inferential analyses, and analysed trial-by-trial performance with maximal logistic mixed-effects regression models for confirmatory hypothesis testing (Barr, Levy, Scheepers, & Tily, 2013; Gelman & Hill, 2007; Judd, Westfall, & Kenny, 2012). The main outcome variable was identification decision type (lineup rejection; positive identification) in models used to evaluate the effects of repetitions,

[3] Supplemental materials for this chapter are presented in Appendix 4

attention, and lineup type on discriminability (accuracy, regardless of response bias) and choosing (De Carlo, 1999; Wright & London, 2009). Data were analysed using the lme4 package (Bates, Maechler, Bolker, & Walker, 2014) for R (R Core Team, 2014).

We made planned pairwise comparisons between conditions of main interest (Gelman, Hill & Yajima, 2012). Logistic regression models predict a binary categorical outcome variable and give the log odds of a given outcome, with values symmetrical around zero. Values greater than 0 indicate the predicted outcome (e.g., a positive identification) is more likely, and values lower than 0 indicate the alternative (e.g., a lineup rejection) is more likely. Binary fixed factor coefficients indicate the difference in the log odds of the outcome between factor levels (log odds ratio, *lnOR*), and we took effects to be significant if the coefficient CI excluded 0. We discuss odds ratios (*ORs*) for differences between experimental cells and present relevant aggregate descriptive statistics.

In models with decision type (lineup rejection or positive identification) as the outcome variable, target presence (target-absent = 0, target-present = 1), repetitions (two = 0, six = 1), attention (divided = 0, full = 0), and their interactions were fixed factors. The coefficient for the fixed factor of repetitions indexed differences in participants' tendency to choose (response bias) depending on target repetitions. Similarly, the coefficient for attention indexed any difference in participants' response bias between attention conditions. The coefficient for target presence indexed the accuracy of choosing, independent of bias (discriminability). Variation in discriminability between levels of attention was thus indexed by the attention × target presence interaction term. Similarly, variation in discriminability between levels of repetitions was indexed by the repetitions × target presence interaction term. Moderation of the effect of attention on discriminability by repetitions was evaluated by examining repetitions interaction terms (with attention and target presence). Likewise, to evaluate differences in the effects of attention on discriminability between lineup procedures, lineup was added to a model predicting decision type as a fixed factor (standard = 0; presence-recollection = 1), with all interaction terms. Effects are most easily interpreted with reference to plotted models.

We report analyses with and without filler identifications in the dataset. With filler identifications (from target-absent and target-present lineups) excluded, discriminability indexed participants' ability to discriminate guilty from innocent suspects. With filler identifications included, discriminability indexed the accuracy of target *presence* decisions, whether the lineup member identified was the suspect or a filler.

We report manipulation checks for the effects of repetitions on target familiarity, and divided attention on available recollection in the online supplement. In sum, the repetitions manipulation produced two levels of familiarity, unconfounded by attention, in each procedure. Thus, we successfully disaggregated the effects of familiarity and recollection. Tone monitoring responses in the divided attention conditions were consistent with participants having attended to tones.

Effect of Repetitions

Specific predictions about the effects of repetitions and attention on false identification depended on whether the lineup tasks yielded an associative pattern of recognition (e.g., Malmberg, 2008; Yonelinas, 2002). We previously established (N. Guerin et al., 2016c) that recognition for the 40 mini-lineups used in the previous experiment did not follow an associative pattern with a standard simultaneous lineup: target-replacement (and filler) familiarity did not increase with increased target familiarity, though substantial target-foil similarity was evident. We evaluated this by examining the effect of repetitions on false identification in the divided-attention condition, when less recollection to counter misleading filler and innocent suspect familiarity was expected to be available (see, e.g., Malmberg, 2008; Dyne et al., 1990). This pattern of recognition replicated in the current experiment (with 40 additional mini-lineups, see Tables 1 & 2). There was no meaningful difference in the likelihood of false identification between repetitions levels with divided attention for standard ($OR_{2\ repetitions:6\ repetitions}$ =1.13) or presence-recollection lineups ($OR_{2\ repetitions:6\ repetitions}$ =1.04). Figure 1 shows that the significant effect of repetitions on choosing with divided attention applied only to target-present lineups (i.e., no significant effect of repetitions; a significant, positive repetitions × target-presence interaction term; a significant negative effect of attention, moderated by a positive target-presence × attention coefficient; and no significant target presence × repetitions × attention interaction terms). Neither procedure demonstrated an associative recognition pattern.

Effect of Attention

Our main aim was to test whether better use of recollection than for standard lineups would explain the greater discriminability previously observed for the presence-recollection procedure (N. Guerin et al., 2016b). We predicted greater discriminability with full than divided attention, with this effect being greater in the presence-recollection than the standard procedure. Separate models for standard and presence-recollection procedures predicting decision type were used to compare discriminability between full and divided attention conditions (see Tables 1 & 2, Figure 1). An omnibus model that included observations from both procedures, and the additional fixed factor of lineup type, tested moderation of this effect by lineup type (see Tables 1 &s 2, Figure 2).

The model for standard lineups showed a clear effect of attention on discriminability (see Figure 1, Panels A, B, E, & F) replicating earlier findings (N. Guerin et al., 2016c). Attention, target presence, and their interaction term significantly predicted decision type (Tables 1 & 2). The interaction term demonstrated greater discriminability with full than divided attention due to the lower likelihood of false identification, characteristic of better recollection retrieval or monitoring (e.g., Gallo, Bell, Beier & Schacter, 2006; Rotello & Heit, 2000). Likewise, for the presence-recollection procedure, participants demonstrated greater discriminability with full than divided attention (Tables 1 & 2, Figure 1, Panels C, D, G, & H). Target presence and the target presence × attention interaction term were significant predictors, and greater discriminability with full than divided attention was largely due to a higher likelihood of accurate target identification. Thus, mock witnesses used recollection in both procedures.

Table 1

Model Fit Statistics for Mixed-Effects Model Predicting Decision Type (Positive Identification; Lineup Rejection) by Attention and Repetitions

Filler Identifications	Lineup Type	χ^2	*df*	*P*
Included	Standard	149.42	7	< .001
	Presence-recollection (PR)	128.6	7	< .001
	All	242.78	7	< .001
Excluded	Standard	197.82	7	< .001
	PR	163.98	7	< .001
	All	317.61	7	< .001

Number of observations, fillers included (fillers excluded): Standard = 6640 (5309), *n* = 83 (83); PR = 6240 (5281), *n* = 78 (78).

Table 2

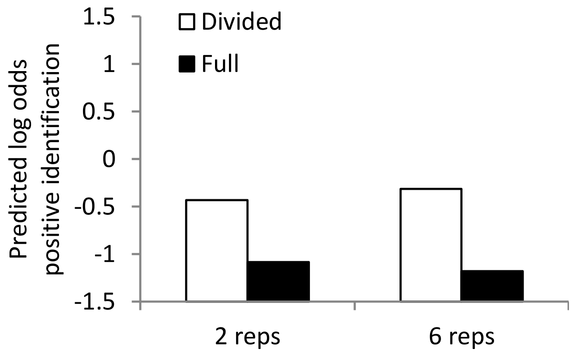
Fixed Effect Coefficients for Mixed Effects Model Predicting Decision Type (Lineup Rejection; Positive Identification) by Lineup Type, Attention (Attn), Repetitions (Reps), and Target Presence (TP)

		Included			Excluded		
Lineup Type	Fixed Effect	*b*	SE_b	95% CI[b]	*b*	SE_b	95% CI[b]
Standard	Intercept	-0.43			-2.06		
	Reps	0.12	0.11	-0.09, 0.34	0.14	0.18	-0.21, 0.49
	Attn	-0.65*	0.24	-1.11, -0.19	-0.70*	0.26	-1.22, -0.19
	TP	1.03*	0.18	0.67, 1.39	2.25*	0.25	1.76, 2.75
	Reps × TP	0.36*	0.15	0.05, 0.66	0.40	0.21	-0.02, 0.81
	Attn × TP	0.84*	0.25	0.36, 1,32	1.10*	0.32	0.47, 1.74
	Reps × Attn	-0.22	0.16	-0.54, 0.10	-0.20	0.28	-0.74. 0.34
	Reps × Attn × TP	0.27	0.23	-0.17, 0.72	0.26	0.32	-0.37, 0.90
PR	Intercept	-1.15			-2.50		
	Reps	0.04	0.14	-0.23, 0.30	-0.20	0.23	-0.65, 0.24
	Attn	-0.22	0.21	-0.63, 0.19	-0.36	0.28	-0.91, 0.18
	TP	1.39*	0.20	0.99, 1.79	2.30*	0.27	1.77, 2.84
	Reps × TP	0.32	0.18	-0.04, 0.67	0.67*	0.27	0.14, 1.20
	Attn × TP	0.86*	0.28	0.32, 1.41	1.28*	0.37	0.55, 2.00
	Reps × Attn	-0.08	0.19	-0.45, 0.29	0.37	0.32	-0.25, 1.00
	Reps × Attn × TP	-0.02	0.25	-0.52, 0.47	-0.59	0.37	-1.32, 0.14
All	Intercept	-0.36			-1.91		

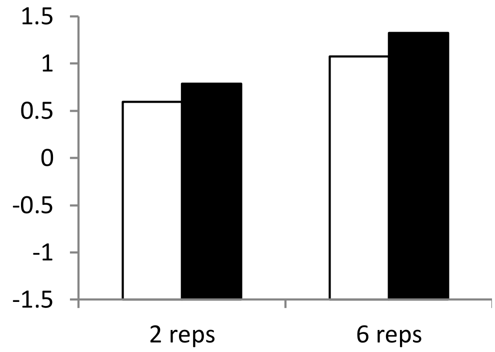
Attn	-0.74*	0.20	-1.13, -0.36	-0.73*	0.20	-1.13, -0.33
TP	1.22*	0.17	0.88, 1.55	2.39*	0.22	1.96, 2.82
Lineup	-0.77*	0.20	-1.16, -0.39	-0.71*	0.20	-1.11, -0.30
Lineup × TP	0.32	0.23	-0.13, 0.76	0.24	0.29	-0.33, 0.80
Attn × TP	0.95*	0.23	0.50, 1.40	1.15*	2.84	0.59, 1.71
Lineup × Attn	0.44	0.28	-0.11, 0.99	0.39	0.30	-0.19, 0.97
Lineup × Attn × TP	-0.04	0.33	-0.68, 0.60	-0.00	0.41	-0.81, 0.81

Note: * indicates significant predictors. Number of observations, fillers included (fillers excluded): Standard = 6640 (5309), *n* = 83 (83); PR = 6240 (5281), *n* = 78 (78).

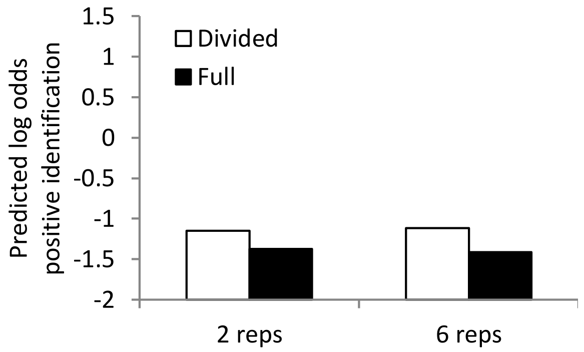
Panel A (Standard, TA, fillers included)



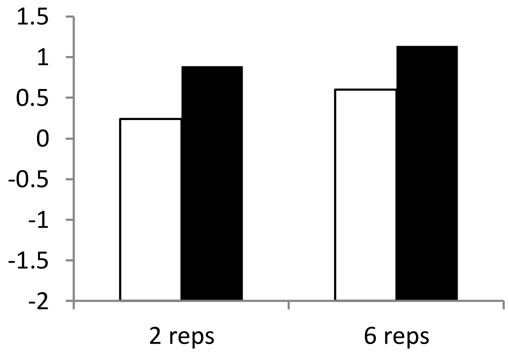
Panel B (Standard, TP, fillers included)



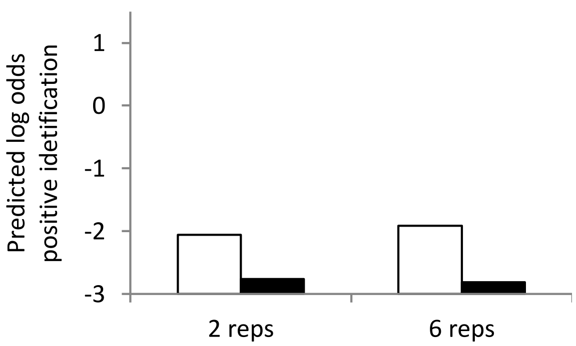
Panel C (PR, TA, fillers included)



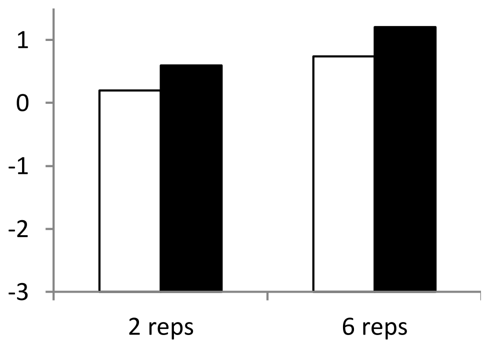
Panel D (PR, TP, fillers included)



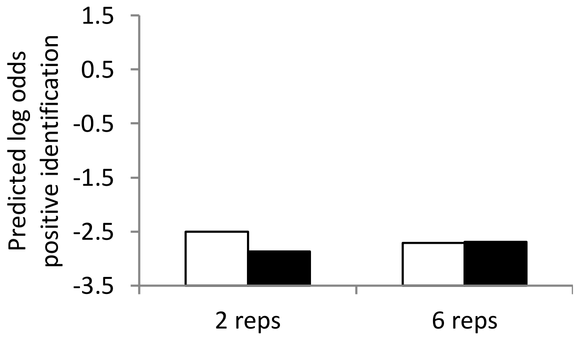
Panel E (Standard, TA, fillers excluded)



Panel F (Standard, TP, fillers excluded)



Panel G (PR, TA, fillers excluded)



Panel H (PR, TP, fillers excluded)

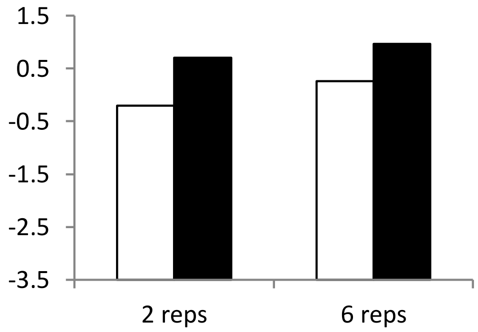


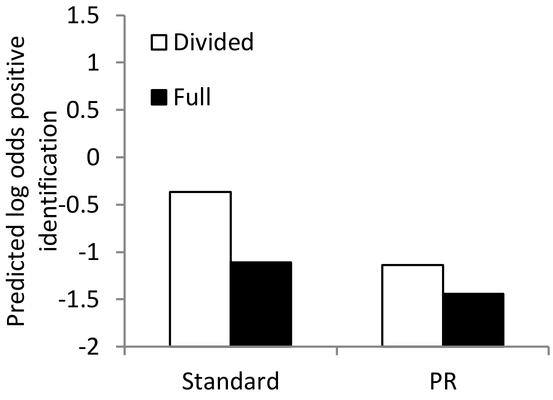
Figure 1. Plots of predicted log odds of a positive identification decision by repetitions and attention, separately for target-absent (TA) and target-present (TP) lineups, for standard and presence-recollection lineups. Values greater than 0 indicate a positive identification from the mini-lineup is more likely, and values lower than 0 indicate a lineup rejection is more likely. Coefficient standard errors, indicative of the variability of the estimates, are displayed in Table 2.

Effect of Lineup

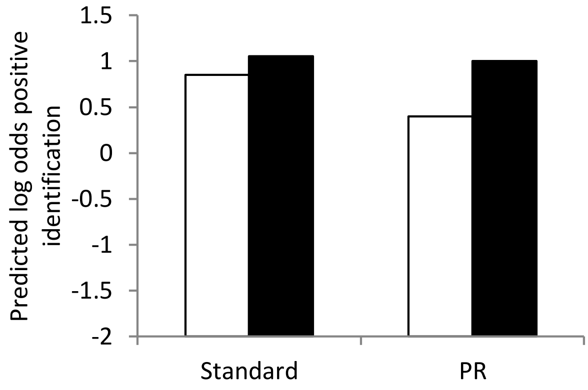
To test the moderating effect of lineup type on the effect of attention on discriminability, we evaluated the omnibus model including fixed effects of target presence, attention, lineup, and all interaction terms (attention × target presence × lineup!), with observations collapsed across repetitions.[4] Significant predictors are reported in Table 2, with the model plotted in Figure 2. In the divided-attention condition, choosing was significantly lower for the presence-recollection than the standard procedure across target presence (lineup was a significant predictor, the interaction terms of lineup × target-presence, and lineup × attention were not, see Table 2, Figure 2). The presence-recollection procedure might have encouraged more conservative response bias. Alternatively, consistent with previously reported evidence of participants weighting familiarity and recollection more appropriately in the presence-recollection lineup (N. Guerin et al., 2016b), presence-recollection participants might have been more reluctant to offer familiarity-based identifications when recollection was less available (with divided attention). If so, a lower likelihood of both false and accurate identification would be expected with divided attention, but not with full attention.

[4] Response patterns were the same with or without the effect of repetitions, with no significant difference between low and high familiarity targets.

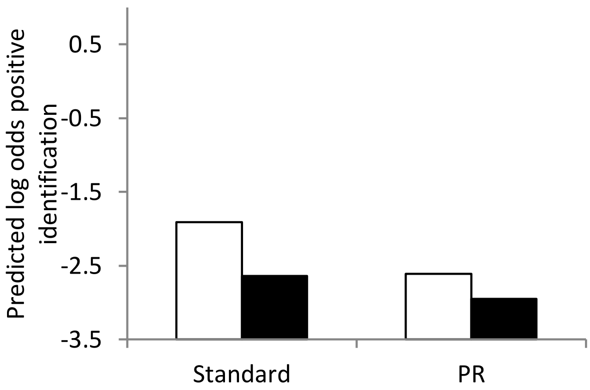
Panel A (TA, fillers included)



Panel B (TP, fillers included)



Panel C (TA, fillers excluded)



Panel D (TP, fillers excluded)

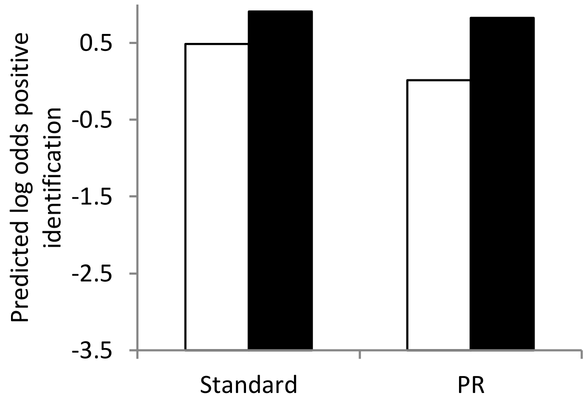


Figure 2. Plots of predicted log odds of a positive identification decision by attention and lineup, separately for target-absent (TA) and target-present (TP) lineups. Values greater than 0 indicate a positive identification from the mini-lineup is more likely, and values lower than 0 indicate a lineup rejection is more likely. Coefficient standard errors, indicative of the variability of the estimates, are displayed in Table 2.

While the greater discriminability in each procedure with full than divided attention is clearly evident in the plotted model (Figure 2), there was no significant difference in the effect on discriminability overall between the two procedures (the lineup × attention and other lineup interaction terms did not significantly predict decision type). That is, with greater availability of recollection, the extent of improvement in discriminability was not significantly greater in the presence-recollection than the standard procedure. Therefore, we have no strong evidence that participants in the presence-recollection condition benefitted more from full attention than those in the standard condition.

It is possible that the greater difficulty of recognition for mini-lineup stimuli used in this experiment limited the possible range of effects. The pattern and size of effects with full attention in the omnibus model (see also Table 3 for aggregate descriptives) suggests a non-significant difference in the direction of greater discriminability with presence-recollection than standard lineups (Figure 2). For standard lineups, as noted above, greater discriminability with full than divided attention was largely due to a lower likelihood of false identification. However, with divided attention, those in the presence-recollection condition were less likely to make familiarity-based identifications than participants in the standard condition. Thus, the likelihood of false identification for presence-recollection participants was already lower (odds = 0.32) than for the standard procedure when attention was divided (odds = 0.69). With the benefit of full attention, the likelihood of false identification was lower for both lineup types, but lowest for presence-recollection lineups ($odds_{presence\text{-}recollection}$ = 0.24; $odds_{standard}$ = 0.33). For lineup type to moderate the effect of attention on false identifications as predicted, false identifications would have had to be substantially lower still from presence-recollection lineups. However, the already low likelihood of false identification from presence-recollection lineups is likely to have limited further improvement.

With divided attention, the likelihood of positive identifications from target-present lineups was significantly lower in presence-recollection lineups than standard lineups ($odds_{standard} = 0.85$, $odds_{presence\text{-}recollection} = 0.40$). This effect was not evident with full attention ($odds_{standard} = 1.06$, $odds_{presence\text{-}recollection} = 1.00$), providing no evidence of significantly lower response bias for presence-recollection lineups. This suggests that while presence-recollection participants were less willing to provide familiarity-based identifications in the absence of recollection, they were no less likely to identify a target than those in the standard procedure when greater recollection was available. On this view, presence-recollection participants were not more conservative choosers overall, but gave less weight to familiarity-without-recollection than those in the standard condition. However, further testing is needed, in experimental conditions more clearly able to capture the full range of potential effects.

In sum, participants in both procedures demonstrated greater discriminability with full than divided attention with no significant moderating effect of lineup. But results also suggest that presence-recollection participants might have placed greater weight on recollection in recognition decisions than those in the standard condition. Presence-recollection discriminability was underpinned by a lower absolute likelihood of false identification, which carries intrinsic value in applied contexts when the base rate of innocent suspects is not negligible.

Table 3

Descriptive Statistics for Identification Responses (Hit rates, Filler ID Rates, Correct Rejection Rates and False ID Rates) by Lineup Type, Repetitions and Attention

				Response Proportion				
Lineup Type	Repetitions	Attention	Statistic	Hits	Filler IDs (TP)	Correct Rejections	Filler IDs (TA)	False IDs
Standard	Two	Divided	*M*	.44	.19	.59	.30	.11
			SD	.17	.18	.21	.17	.07
			95% CI	.38, .49	.14, .24	.53, .65	.25, .35	.08, .13
		Full	*M*	.57	.10	.71	.21	.08
			SD	.19	.11	.20	.17	.06
			95% CI	.51, .62	.07, .14	.65, .77	.16, .26	.06, .09
	Six	Divided	*M*	.53	.19	.56	.32	.12
			SD	.18	.15	.22	.18	.09
			95% CI	.48, .59	.14, .23	.50, .63	.26, .37	.09, .14
		Full	*M*	.68	.09	.73	.19	.08

			SD	.19	.10	.18	.15	.06
			95% CI	.62, .73	.06, .12	.67, .78	.15, .24	.06, .10
PR	Two	Divided	*M*	.40	.16	.73	.18	.09
			SD	.16	.10	.14	.12	.07
			95% CI	.35, .45	.13, .19	.69, .78	.14, .22	.06, .11
		Full	*M*	.59	.10	.77	.17	.06
			SD	.18	.09	.15	.11	.06
			95% CI	.53, .65	.07, .13	.72, .81	.13, .20	.04, .08
	Six	Divided	*M*	.48	.15	.72	.20	.08
			SD	.21	.12	.12	.10	.06
			95% CI	.41, .54	.12, .19	.68, .76	.17, .23	.06, .10
		Full	*M*	.64	.10	.77	.15	.07
			SD	.21	.11	.16	.12	.07
			95% CI	.57, .70	.06, .14	.72, .83	.11, .19	.05, .10

Discussion

In a mini-lineup experiment, we investigated participants' use of familiarity and recollection to make simultaneous lineup decisions using either a standard procedure, or the presence-recollection lineup. We had two main research questions. First, would participants in both procedures use recollection to improve identification accuracy? Second, would presence-recollection participants make better use of recollection than those using a standard simultaneous procedure? Consistent with previous research (N. Guerin et al., 2016c), we found evidence that participants in both procedures used recall-to-reject (Rotello & Heit, 2000) to avoid false identifications. Results suggested greater weighting of recollection in participants' responses to presence-recollection lineups, but we did not find a significantly greater benefit of recollection than for standard lineups.

As expected, and replicating previous results for standard simultaneous lineups (N. Guerin et al., 2016c), discriminability was greater with greater availability of recollection in both procedures. Recollection strengthened familiarity-based evidence for accurate identifications, and participants also used recall-to-reject to avoid familiarity-based false identifications of foils that were misleadingly familiar because of their substantial similarity to targets. With greater encoded recollection, discriminability tended to be greater for presence-recollection lineups than for simultaneous lineups, due to a lower likelihood of false identification. However, with less encoded recollection, the effect of lineup on discriminability was not significant due to a lower likelihood of making familiarity-based target identifications for presence-recollection lineups, as well as fewer false identifications. We conclude that while participants following the standard procedure used recollection, it is possible that recollection was not used to its full potential, through either a failure to maximise recall-to-reject, or by underweighting the importance of retrieved recollection during post-retrieval monitoring (e.g., Dodson & Schacter, 2002; Johnson, Hashtroudi, &

Lindsay, 1993). In contrast, patterns of responses from participants following the presence-recollection procedure suggested a greater reluctance to make positive identifications based predominantly on familiarity. Therefore, responses to presence-recollection lineups would also tend to be more reliable than those from standard lineups, even without significantly different discriminability. Results thus provide valuable evidence about the influence of judgment type, and the alignment of task understandings and demands, on metacognitive control of memory and decision processes for identically presented stimuli. Findings provide some support for further investigation of the efficacy of the presence-recollection lineup for enhancing the memorial basis of simultaneous lineup decisions.

While results are consistent with participants having used recall-to-reject in both lineup procedures, our design does not allow us to disaggregate mechanisms that potentially contributed to the use of recollection in this task. Research in other recognition paradigms suggests several mechanisms that may have been used by participants (Brainerd, Reyna, Wright & Mojardin, 2003; Jacoby, Kelley, & McElree, 1999; Odegard & Lampinen, 2006). We argue that multiple mechanisms were likely to have contributed to retrieval and metacognitive monitoring in both lineup procedures (e.g., Gallo, 2010). These mechanisms operate at different points of decision-making from initial retrieval (e.g., Gray & Gallo, 2015; Hunt & Smith, 2014), through post-retrieval monitoring (e.g., Schacter, Israel, & Racine, 1999); in further post-monitoring retrieval (Jacoby et al., 1999); and as part of controlled decision-making (e.g., Koriat & Goldsmith, 1996). Importantly, processes might have differed between lineup types to produce the observed variability in patterns of discriminability. As each process, or their combination, will have specific implications for identification accuracy in diverse witnessing conditions, and thus for the efficacy of the presence-recollection procedure in improving accuracy over and above that observed for standard lineups, future research should be directed at investigating their separate contributions to participants' use of recollection in lineup tasks.

One key consideration in identifying underlying mechanisms for the use of recollection is that the quality of eyewitness recollection will frequently be poor, due to difficult encoding conditions during a witnessed event, or with post-event degradation over time. In these cases, mechanisms that rely on recollection retrieval will have limited effectiveness. In contrast, some post-retrieval mechanisms might be less affected by the absence of recollection. For example, the distinctiveness heuristic (Schacter et al., 1999), that relies on comparison of the quality of memory retrieved with that expected in response to a target.

However, many post-retrieval mechanisms, including the distinctiveness heuristic, might rely on some baseline recollection, or the ability to compare recollective with non-recollective retrieval over a number of trials. A very weak recollective trace that was only successfully cued by target presence might not contribute to use of the distinctiveness heuristic for target-absent lineups (e.g., S. A. Guerin, Robbins, Gilmore, & Schacter, 2012). There is evidence that for familiarity-based memories, eyewitness feeling-of-knowing is not a good predictor of recognition accuracy (Perfect & Hollins, 1996, 1999), leaving innocent suspects vulnerable to false identification. Further, this weakness would not be reflected in markers of accuracy, such as decision confidence. It is possible that feeling-of-knowing might be enhanced by metacognitive instructions. But, if not, the distinctiveness heuristic or other monitoring processes might not be useful when recollection is minimal, nor would any procedure that relied on them. This example highlights the importance of identifying the specific recollection retrieval and monitoring mechanisms that operate in lineup tasks. And the need to investigate multiple methods for enhancing eyewitness metacognition and performance, to cater for both varying witnessing and test conditions, and specific witness populations.

While the mini-lineup paradigm allowed the collection of ample data to observe regularities in performance, we cannot rule out the possibility that paradigm-specific effects also influenced participants' use of recollection. It is possible that greater neglect of recollection in standard simultaneous lineups, as proposed by the best-match-first dual-process framework, would be evident in a more realistic identification paradigm. For example, recall-to-reject or more rigorous monitoring might have been encouraged by the relatively high difficulty of this task (e.g., Cook et al., 2005). Alternatively, the opportunity to evaluate variations in memory quality across multiple trials might have prompted recollection and/or relativistic memory editing strategies (e.g., the distinctiveness heuristic). Such features are not characteristic of real lineup conditions. However, findings from other research using the mini-lineup paradigm to investigate eyewitness memory and metacognition have translated well to the identification paradigm (e.g., Sauer, Brewer, & Weber, 2008; Weber & Brewer, 2004, 2006). We also found no evidence of a different pattern of responses from evaluating only first trial responses for this and previous mini-lineup experiments (N. Guerin et al., 2016b, 2016c). However, generalisation of these results to a single-trial identification task is an open question.

The complexities of real world witnessing and test conditions demand extensive testing under diverse conditions, guided by investigation of the mechanisms that contributed to the originally-observed presence-recollection advantage, before conclusions can be made about the usefulness of the modified procedure for practical application. However, these results provide important evidence for further research into the effects of targeted structural changes and metacognitive instructions on lineup decisions, as part of the endeavour to maximise eyewitness memory within its inescapable limits.

References

Alban, M. W., & Kelley, C. M. (2012). Variations in constrained retrieval. *Memory & Cognition*, *40*, 681-692. doi: 10.3758/s13421-012-0185-5

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structures for confirmatory hypothesis testing: Keeping it maximal. *Journal of Memory and Language*, 68, 255-278. doi: 10.1016/j.jml.2012.11.001

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2014). _lme4: Linear mixed-effects models using Eigen and S4_. Retrieved from http://CRAN.R-project.org/package=lme4

Brainerd, C. J., Reyna, V. F., Wright, R., & Mojardin, A. H. (2003). Recollection rejection: False-memory editing in children and adults. *Psychological Review*, *110*, 762-784. doi: 10.1037/0033-295x.110.4.762

Clark, S. E. (2003). A memory and decision model for eyewitness identification. *Applied Cognitive Psychology*, *17*, 629-654. doi: 10.1002/acp.891

Clark, S. E. (2005). A re-examination of the effects of biased lineup instructions in eyewitness identification. *Law and Human Behavior*, *29*, 575-604. doi: 10.1007/s10979-005-7121-1

Clark, S. E. (2012). Costs and benefits of eyewitness identification reform psychological science and public policy. *Perspectives on Psychological Science*, *7*, 238-259. doi: 10.1177/1745691612439584

Clark, S. E., & Tunnicliff, J. L. (2001). Selecting lineup foils in eyewitness identification experiments: Experimental control and real-world simulation. *Law and Human Behavior, 25,* 199-216. doi: 10.1023/a:1010753809988

Cook, G. I., Marsh, R. L., & Hicks, J. L. (2005). Revisiting the role of recollection in item versus forced-choice recognition memory. *Psychonomic Bulletin & Review*, *12*, 720-725. doi: 10.3758/bf03196763

De Carlo, L. T. (1998). Signal detection theory and generalized linear models. *Psychological Methods, 3*, 186-205. doi: 10.1037//1082-989x.3.2.186

Dodson, C. S., & Schacter, D. L. (2002). When false recognition meets metacognition: The distinctiveness heuristic. *Journal of Memory and Language*, *46*, 782-803. doi: 10.1037/e501882009-058

Dyne, A. M., Humphreys, M. S., Bain, J. D., & Pike, R. (1990). Associative effects in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 813-824. doi: 10.1037/0278-7393.16.5.813

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition*, *38*, 833-848. doi: 10.3758/mc.38.7.833

Gallo, D. A., Bell, D. M., Beier, J. S., & Schacter, D. L. (2006). Two types of recollection-based monitoring in younger and older adults: Recall-to-reject and the distinctiveness heuristic. *Memory, 14*, 730-741. doi: 10.1080/09658210600648506

Gardiner, J. M., & Parkin, A. J. (1990). Attention and recollective experience in recognition memory. *Memory & Cognition*, *18*, 579-583. doi: org/10.3758/bf03197100

Gelman, A., & Hill, J. (2007). *Data Analysis using Regression and Multilevel/Hierarchical Models.* New York: Cambridge University Press.

Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness*, *5*(2), 189-211. doi: 10.1080/19345747.2011.618213

Gray, S. J., & Gallo, D. A. (2015). Disregarding familiarity during recollection attempts: Content-specific recapitulation as a retrieval orientation strategy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 134-147. doi: 10.1037/a0038363

Gronlund, S. D. (2005). Sequential lineup advantage: Contributions of distinctiveness and recollection. *Applied Cognitive Psychology*, 19, 23-37. doi: 10.1002/acp.1047

Gronlund, S. D., Carlson, C. A., Daily, S. B., & Goodsell, C. A. (2009). Robustness of the sequential lineup advantage. *Journal of Experimental Psychology: Applied,* 15, 140-152. doi: 10.1037/a0015082

Gross, S. R., & Schaffer, M. (2012). Exonerations in the United States, 1989 – 2012: Report by the National Registry of Exonerations. *University of Michigan Public Law Working Paper*, 277. doi: 10.2139/ssrn.2092195

Guerin, N., Weber, N., & Horry, R. (2016a). Split decision: An account of eyewitness decision-making in simultaneous lineups and a novel presence-first identification procedure. *Under review*

Guerin, N., Weber, N., & Horry, R. (2016b). Only recollect: Improving identification accuracy by reframing eyewitness metacognition and decision structure. *Under review*

Guerin, N., Weber, N., & Horry, R. (2016c). Do eyewitnesses recall-to-reject? Familiarity and recollection in simultaneous lineups. *Under review*

Guerin, S. A., Robbins, C. A., Gilmore, A. W., & Schacter, D. L. (2012). Retrieval failure contributes to gist-based false recognition. *Journal of Memory and Language*, *66*, 68-78. doi: 10.1016/j.jml.2011.07.002

Hancock, P. J., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences*, *4*, 330-337. doi: 10.1016/s1364-6613(00)01519-9

Horry, R., Memon, A., Wright, D. B., & Milne, R. (2011). Predictors of eyewitness identification decisions from video lineups in England: A field study. *Law and Human Behavior*, *36*, 257-265. doi: 1007/s10979-011-9279-z

Hunt, R. R., & Smith, R. E. (2014). How distinctive processing enhances hits and reduces false alarms. *Journal of Memory and Language*, *75,* 45-57. doi: 10.1016/j.jml.2014.04.007

Ingram, K. M., Mickes, L., & Wixted, J. T. (2012). Recollection can be weak and familiarity can be strong. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 325-339. doi: 10.1037/e520592012-514

Jacoby, L. L., Kelley, C. M., & McElree, B. D. (1999). The role of cognitive control: Early selection versus late correction. In. S. Chaiken & Y. Trope (Eds.), Dual-process theories in social psychology (pp. 383– 400). New York, NY: Guilford Press.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3-28. doi: 10.1037//0033-2909.114.1.3

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology, 103*(1), 54-69. doi: 10.1037/a0028347

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, *103*, 490-517. doi: 10.1037/0033-295x.103.3.490

Lane, S. M., Roussel, C. C., Starns, J. J., Villa, D., & Alonzo, J. D. (2008). Providing information about diagnostic features at retrieval reduces false recognition. *Memory*, *16*, 836–851. http://doi: 10.1080/0965821080233734

Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology*, *57*, 335-384. doi: 10.1016/j.cogpsych.2008.02.004

Malmberg, K. J., & Xu, J. (2007). On the flexibility and the fallibility of associative memory. *Memory & Cognition, 35,* 545-556. doi: 10.3758/bf03193293

Malpass, R. S., & Devine, P. G. (1981). Eyewitness identification: Lineup instructions and the absence of the offender. *Journal of Applied Psychology, 66*, 482. doi: 10.1037/0021-9010.66.4.482

Mäntylä, T., & Cornoldi, C. (2002). Remembering changes: Repetition effects in face recollection. *Acta Psychologica*, *109*, 95-105. doi: 10.1016/s0001-6918(01)00054-3

Meissner, C. A., Tredoux, C. G., Parker, J. F., MacLin, O. H. (2005). Eyewitness decisions in simultaneous and sequential lineups: A dual-process signal detection theory analysis. *Memory & Cognition,* 33, 783-792. doi: 10.3758/bf03193074

Mickes, L. (2015). Receiver operating characteristic analysis and confidence-accuracy characteristic analysis in investigations of system variables and estimator variables that affect eyewitness memory. *Journal of Applied Research in Memory and Cognition, 4,* 83-102. doi: 10.1016/j.jarmac.2015.01.003

Mickes, L., Flowe, H. D., & Wixted, J. T. (2012). Receiver operating characteristic analysis of eyewitness memory: comparing the diagnostic accuracy of simultaneous versus sequential lineups. *Journal of Experimental Psychology: Applied*, *18*, 361-376. doi:

Odegard, T. N., & Lampinen, T. N. (2006). Memory editing: Knowledge, criteria, and alignment. *Memory*, *14,* 777-787. doi: 10.1080/09658210600648589

Odegard, T. N., Lampinen, J. M., & Toglia, M. P. (2005). Meaning's moderating effect on recollection rejection. *Journal of Memory and Language*, *53*, 416-429. doi: 10.1016/j.jml.2005.04.004

Palmer, M. A., Brewer, N., McKinnon, A. C., & Weber, N. (2010). Phenomenological reports diagnose accuracy of eyewitness identification decisions. *Acta Psychologica*, *133*, 137-145. doi: 10.1016/j.actpsy.2009.11.002

Parkin, A. J., Gardiner, J. M., & Rosser, R. (1995). Functional aspects of recollective experience in face recognition. *Consciousness and Cognition*, *4,* 387-398. doi: 10.1006/ccog.1995.1046

Perfect, T. J., & Hollins, T. S. (1996). Predictive feeling of knowing judgements and postdictive confidence judgements in eyewitness memory and general knowledge. *Applied Cognitive Psychology*, *10*, 371-382. doi: 10.1002/(sici)1099-0720(199610)10:5<371::aid-acp389>3.0.co;2-o

Perfect, T. J., & Hollins, T. S. (1999). Feeling-of-knowing judgments do not predict subsequent recognition performance for eyewitness memory. *Journal of Experimental Psychology: Applied*, *5*, 250-264. doi: 10.1037/1076-898x.5.3.250

R Core Team. (2014). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org.

Reinitz, M. T., Morrissey, J., & Demb, J. (1994). Role of attention in face encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 161. doi: 10.1037/0278-7393.20.1.161

Rotello, C. M., & Heit, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory & Cognition*, *28*, 907-922. doi: 10.3758/bf03209339

Sauer, J. D., Brewer, N., & Weber, N. (2008). Multiple confidence estimates as indices of eyewitness memory. *Journal of Experimental Psychology: General*, 137, 528-547. doi: 10.1037/a0012712

Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language*, *40,* 1-24. doi: 10.1006/jmla.1998.2611

Starns, J. J., Lane, S. M., Alonzo, J. D., & Roussel, C. C. (2007). Metamnemonic control over the discriminability of memory evidence: A signal detection analysis of warning effects in the associative list paradigm. *Journal of Memory and Language*, *56*, 592-607. doi: 10.1016/j.jml.2006.08.013

Weber, N., & Brewer, N. (2004). Confidence–accuracy calibration in absolute and relative face recognition judgments. *Journal of Experimental Psychology: Applied, 10,* 156–172. doi: 10.1037/1076-898x.10.3.156

Weber, N., & Brewer, N. (2006). Positive versus negative face recognition decisions: Confidence, accuracy and response latency. *Applied Cognitive Psychology, 20,* 17–31. doi: 10.1002/acp.1166

Wells, G. L. (1984). The Psychology of Lineup Identifications1. *Journal of Applied Social Psychology*, *14*, 89-103. doi: 10.1111/j.1559-1816.1984.tb02223.x

Wells, G. L. (1993). What do we know about eyewitness identification? *American Psychologist*, *48*, 553-571. doi: 10.1037//0003-066x.48.5.553

Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S.M., & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behaviour*, *22*, 603-647. doi: 10.1023/a:1025750605807

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin*, *25,* 1115-1125. doi: 10.1177/01461672992512005

Wells, G. L., Steblay, N. K., & Dysart, J. E. (2015). Double-blind photo lineups using actual eyewitnesses: An experimental test of a sequential versus simultaneous lineup procedure. *Law and Human Behavior, 39,* 1-14. doi: 10.1037/lhb0000096

Westfall, J., Judd, C. M., & Kenny, D. A. (2015). Replicating studies in which samples of participants respond to samples of stimuli. *Perspectives on Psychological Science, 10,* 390-399. doi: 10.1177/1745691614564879

Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review, 117*, 1025-1054. doi: 10.1037/a0020874

Wogalter, M. S., Malpass, R. S., & McQuiston, D. E. (2004). A national survey of US police on preparation and conduct of identification lineups. *Psychology, Crime and Law, 10*, 69-82. doi: 10.1080/10683160410001641873

Wright, D.B., & London, K. (2009). Multilevel modelling: Beyond the basic applications. *British Journal of Mathematical and Statistical Psychology*, 62, 439-456. doi: 10.1348/000711008X327632

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*, 441-517. doi: 10.1006/jmla.2002.2864

CHAPTER 7

The Presence-Recollection Lineup:

Can Eyewitnesses Make Better Use of Recollection in Identification Contexts?[1]

Abstract

We proposed that eyewitnesses typically select a best-match from a simultaneous lineup before deciding if the best-match is the offender, promoting over-reliance on familiarity and neglect of recollection. The presence-recollection lineup was designed to encourage recollection. Eyewitnesses received metacognitive instructions and made two-step identification responses: first, the witness indicated whether they thought the offender was present (*Present/Not present*); second, following a 'present' decision, the witness indicated the identity of the offender. In an ecologically valid identification paradigm, we found no evidence of greater discriminability than with standard simultaneous or sequential lineups. Eyewitnesses might have been unable to recollect enough detail of offenders' appearance to improve decision accuracy with recall-to-reject or recollection-dependent post-retrieval memory editing. However, if bias toward making a positive identification resisted metacognitive intervention, even when memory was poor, radical changes to lineup structure will be needed to maximise the memorial basis of identifications in real criminal investigations.

[1] Submitted to *Journal of Applied Research in Memory and Cognition,* as Guerin, N., Weber, N., & Horry, R. (2016). The presence-recollection lineup: Can eyewitnesses make better use of recollection in identification contexts? As first author, I designed and conducted the experiments and analyses, interpreted results, and wrote the manuscript. Nathan Weber, as my PhD supervisor, advised on the experimental and analytical design, and contributed to the interpretation of results and manuscript editing. Ruth Horry, as Adjunct Supervisor, contributed to the interpretation of results and the manuscript.

Introduction

Police frequently gather eyewitness identification evidence to guide the course of investigation into both petty and serious crime (Wogalter, Malpass, & McQuiston, 2004). But memory for events and previously unknown people can be fragile and inaccurate, and the identification evidence gained from eyewitnesses has consistently been shown to be unreliable (e.g., Behrman & Richards, 2005; Clark, Howell, & Davey, 2008; Horry, Memon, Wright, & Milne, 2011). Yet, an identification is often the only evidence for prosecution, and many defendants are tried and convicted on this basis (Garrett, 2011; Goldstein, Chance, & Schneller, 1989; Gross, Jacoby, Matheson, Montgomery, & Patil, 2005). False convictions of the innocent can be a direct result of this over-reliance on flawed evidence, with false identification evidence implicated in over 70% of DNA exonerations in the U.S. (Innocence Project, 2016). Equally, missed identifications can frustrate the just resolution of criminal investigations. These misses contribute to systemic reluctance to adopt recommended reforms of identification procedure that have been shown to reduce false identification, but are also likely to reduce rates of accurate identification (e.g., National Research Council, 2014). Therefore, an important goal of eyewitness identification research is the development of test formats that will facilitate better use of available memory by eyewitnesses in both directions, to avoid falsely identifying innocent suspects, but also to avoid missing guilty suspects. We designed the novel presence-recollection procedure to change the memorial basis of eyewitness identification decisions. Specifically, we aimed to maximise eyewitnesses' recollection of encoded details of the offender's appearance to both strengthen accurate positive identifications, and more strongly resist false identifications of innocent suspects who resemble the offender.

Alternatives to the traditional simultaneous lineup procedure have been few. Following the innovation of the sequential lineup (Lindsay & Wells, 1985), multiple sequential variants have been tested (Horry, Palmer, & Brewer, 2012; Wells, Steblay, & Dysart, 2015), and adopted in a growing number of jurisdictions (e.g. Horry et al., 2011; National Research Council, 2014; National VIPER Bureau, 2009). However, extensive debate has highlighted the lack of consistent results from comparisons of identification performance between simultaneous and sequential formats (e.g., Clark & Davey, 2005; Dobolyi & Dodson, 2013; Gronlund, Carlson, Daily, & Goodsell, 2009; Palmer & Brewer, 2012; Steblay, Dysart, & Wells, 2011; Wells, 2014). Even if one of these formats was most appropriate in given circumstances, and boundary conditions could be established, identification performance remains unacceptably poor for both, overall. Yet, the main focus of research into existing identification procedures has been on how eyewitnesses make decisions about, and report, memorial information, rather than attempting to improve memory retrieval itself. While some novel identification test formats have been tested in the laboratory (e.g., Brewer, Weber, Wootton, & Lindsay, 2012; Pozzulo & Lindsay, 1999; Pryke, Lindsay, Dysart, & Dupuis, 2004; Sauer, Brewer, & Weber, 2008; Sauerland & Sporer, 2009; Weber & Perfect, 2012; Zajac & Karageorge, 2009) potentially fruitful avenues of research into memory retrieval and editing processes remain relatively unexplored. Our best-match-first dual-process framework proposes that eyewitnesses routinely neglect recollective aspects of memory in evaluating lineup members, leaving them prone to false identifications based on sometimes misleading global familiarity. The presence-recollection simultaneous lineup format was designed to tap that unused evidence to reduce risk to the innocent, while maintaining or increasing identifications of the guilty. Greater use of recollection would provide clearer evidence of mismatch between the offender and an innocent suspect, but bolster positive match evidence for a guilty suspect.

Eyewitnesses using the presence-recollection lineup are given metacognitive instructions about memory qualities known to be diagnostic of accuracy, before providing an identification response in two steps. First, they indicate whether or not the offender is in the lineup (*Present, Not present*). Second, those witnesses who indicated that the lineup includes the offender, are asked to indicate the identity of the offender. The procedure was developed from our best-match-first dual-process framework for the analysis of simultaneous lineups (N. Guerin, Weber, & Horry, 2016a). The first leg of the framework is a best-match-first account of simultaneous lineup decision-making that formalises Wells' (1984) relative judgment idea. Broadly, the best-match-first account proposes that witnesses tend to first select a best-match candidate, before deciding whether or not the best-match is the offender (Clark, 2003; Wells, 1984, 1993). If eyewitnesses adopt a too-lenient criterion for the second offender-presence decision, or fail to consider all available memorial information, they risk falsely identifying an innocent suspect with a close-enough resemblance to the offender to be selected as the best-match (Clark, 2005; Wells, 1984, 1993, 2014).

A continuous dual-process model of memory forms the second leg of the framework, and proposes that recognition can be informed by two memory retrieval processes: global-matching familiarity, and more effortful recollection (e.g., Wixted & Mickes, 2010; Yonelinas, 2002). In the framework, we proposed that the comparison processes used to select a best-match from among multiple options mislead witnesses to focus primarily on evaluating familiarity-based evidence for positive match. Witnesses then proceed to neglect the importance of recollected detail when deciding whether or not the best-match is the offender. This has two serious consequences for identification errors. First, without retrieval or appropriate weighting of recollection of the offender, eyewitnesses will underestimate the evidence for positive identification for a guilty suspect, and risk missing the culprit. Second, an innocent suspect who resembles the offender will seem familiar, and recollective detail is needed to offset this misleading familiarity (i.e., recall-to-reject, Rotello & Heit, 2000) to avoid a false identification. Therefore, identification-test design should 1) disrupt the best-match-first order of decision-making; and/or, 2) improve eyewitnesses' use of recollection. The presence-recollection procedure was designed to achieve both aims via two components: the modified simultaneous task structure and metacognitive instructions.

The presence-recollection lineup reverses the best-match-first order of decision-making when eyewitnesses provide responses in two steps: 1) indicating whether or not the offender is in the lineup; and 2) if a positive decision was made at step 1, indicating which member of the lineup is the offender. Requiring eyewitnesses to decide first on offender-presence, before focussing on pointing out the offender, if present, should avert a premature narrowing of attention to a targeted search for the single best-match. In this way, a more balanced evaluation of evidence for and against memory-match for each lineup member would be facilitated. In addition, pre-test metacognitive instructions about memory qualities that are diagnostic of accuracy were tailored to improve retrieval and monitoring of recollection. Substantial research has shown greater accuracy in a variety of memory tasks with high target-foil similarity when people are informed before test about the phenomenology of accurate remembering (e.g., Gallo, Roberts, & Seamon, 1997; Lane, Roussel, Starns, Villa, & Alonzo, 2008; Schooler, Gerhard, & Loftus, 1978). Presence-recollection instructions focus on the greater vividness, clarity and detail likely to be associated with recollection versus global familiarity. We anticipated that these phenomenological instructions would increase eyewitnesses' use of recollection through recall-to-reject (Rotello & Heit, 2000) or post-retrieval memory editing mechanisms (Dodson & Schacter, 2002; Gallo, Bell, Beier, & Schacter, 2006). In contrast, eyewitnesses following a simultaneous procedure were expected to follow structural cues to best-match-first decision-making, and familiarity-based selection and evaluation of a single best-match.

People do not always use detailed recollection to make recognition decisions, even if the memory trace is available (e.g., Lane et al., 2008; Malmberg & Xu, 2007). However, this kind of recollection is vital to prevent identification errors caused by over-reliance on the familiarity of lineup-members and their degree of positive global match with the offender. In basic memory research, higher accuracy and a stronger confidence-accuracy relationship have been observed for recognition decisions informed by recollection, than those based on familiarity (e.g., Ingram, Mickes, & Wixted, 2012; Malmberg, 2008; Wixted & Mickes, 2010; Yonelinas, 2002). In identification tests, this pattern of performance would increase the diagnosticity of individual identification responses. We investigated these effects in lineup identification tasks in two previous experiments using a mini-lineup paradigm, by manipulating the objective amount of familiarity and recollective evidence participants were able to encode (N. Guerin, Weber, & Horry, 2016c). Results showed that when identification decisions for a standard simultaneous lineup were informed by relatively higher recollection, discriminability of guilty from innocent suspects was greater than when less recollection was available. The positive confidence-accuracy relationship was also stronger for these decisions. In a small number of identification studies, researchers have used Tulving's (1985, 1989; see also Gardiner & Richardson-Klavehn, 2000) Remember-Know procedure to measure eyewitnesses' subjective recollection and familiarity ratings for identification decisions, and reported similar patterns of results (e.g., Meissner, Tredoux, Parker, & Maclin, 2005; Palmer, Brewer, McKinnon, & Weber, 2010; Mickes, 2015).

In previous research, participants used the presence-recollection procedure in a mini-lineup paradigm to make identification decisions for multiple lineups, each for a different offender seen previously in studied photos. Participants were better at differentiating guilty from innocent suspects when following the presence-recollection procedure than a standard simultaneous procedure (N. Guerin, Weber, & Horry, 2016b). This effect was driven by a lower likelihood of false identification. When little recollective detail had been encoded, presence-recollection witnesses offered fewer familiarity-based responses than standard witnesses (N. Guerin, Weber, & Horry, 2016d). However, presence-recollection participants used recollection when available to increase hits, and to counter the misleading familiarity of lineup members who matched the offender's description to successfully avoid false identifications (N. Guerin et al., 2016d). While we did not measure decision confidence, this greater contribution of recollection to accurate identifications of guilty suspects would also be expected to improve confidence-accuracy calibration for both rejections and identifications, even with no change in the number of accurate positive identifications (e.g., Starns, Lane, Alonzo, & Roussel, 2008). Hence, with the novel procedure, we would expect confidence to more reliably predict identification accuracy, and provide a better estimation of the probative value of a single identification decision.

It is important to note that previously reported results demonstrated that both elements of the presence-recollection procedure (two-step presence-first decision structure; recollection instructions) were needed to improve discriminability. There was no evidence of a significant effect of recollection instructions when provided with a standard simultaneous procedure. When eyewitnesses knew they might need to point out a guilty offender from a lineup, neither a presence-first task without recollection-instructions (N. Guerin et al., 2016a, Experiment 1), or recollection-instructions alone (i.e., with a standard simultaneous lineup, N. Guerin et al., 2016b, Experiments 2 & 3), led to greater discriminability than for a standard simultaneous procedure. The focus of the experiment we report here was comparison of presence-recollection and standard simultaneous procedures in an ecologically valid paradigm, not the effectiveness of the individual elements of the presence-recollection procedure. Thus, the confounding of manipulations of instructions and lineup structure in the presence-recollection lineup was evidence-based. Whether recollection instructions would enhance eyewitnesses' metacognitive strategies with a sequential procedure remains an open question, beyond the scope of this study.

The Experiment

The potential increase in the probative value of identification evidence to be gained from better use of recollection by eyewitnesses is substantial. However, previous evidence of the efficacy of the presence-recollection lineup, was gained in the mini-lineup paradigm that controls many extrinsic influences on eyewitness decision-making. Here, we report the first test of the novel presence-recollection procedure in a more realistic identification paradigm. Participants viewed mock crime stimulus videos for three targets committing non-violent mock crimes, and responded to a six-person lineup for each offender at test. We predicted greater discriminability with the presence-recollection procedure than a standard simultaneous procedure.

We also included a sequential lineup control. The best-match-first dual-process framework makes no predictions about the use of familiarity and recollection in a sequential procedure. Clearly, sequential presentation would disrupt best-match-first decision-making. However, the use of familiarity and recollection in yes/no single-item recognition varies with task specifics (Cook, Marsh, & Hicks, 2005; Malmberg & Xu, 2007). Therefore, eyewitnesses' use of familiarity and recollection in sequential procedures is an open empirical question (but see Meissner et al., 2005). Comparison of the presence-recollection procedure with a sequential procedure was exploratory. This design allowed the first experimental comparison of sequential and presence-recollection procedures, as well as a useful replication of the simultaneous-sequential comparison with previously untested stimuli.

Method

Participants and Design

A priori, we aimed to collect data from 100 participants per lineup condition (300 in all). 302 undergraduate students from Flinders University (222 female, age: *M* = 23.74 years, *SD* = 7.77 years) participated for payment or course credit. All collected data were analysed. Participants were randomly assigned to lineup type (simultaneous; sequential; presence-recollection) and target presence was varied within-subjects.

Materials

The stimulus events were three short videos previously used in our laboratory. Archival laboratory data suggested that these stimulus videos produced different patterns of identification test responses (i.e., choosing and accuracy). Each participant viewed all three videos and made lineup decisions for each, to increase generalizability by testing the modified procedure across multiple stimuli. Each video showed a different simulated non-violent crime or misdemeanour involving a different offender. Videos were presented on an 18” monitor (resolution: 1920 × 1080 pixels) and presentation order was counterbalanced. The *park* video (4 min 26 s) showed a young man strolling around a park, dropping litter and bumping into another park-goer, before helping a woman to fold her picnic blanket. The offender was in view with a full or partial view of his face available for 3 min 23 s. In the *liquor-store* video (26 s), a young man was shown entering a liquor store shortly after a similar-looking customer. He selected a bottle of wine from a display and hid it in his coat before leaving the shop. The offender was in view with a full or partial view of his face available for 12 s. Finally, in the *wallet* video (29 s), a young woman walked around a ground-level car park, got into a car and rummaged through the glove compartment, and took money from a wallet found inside a sports bag on the seat before walking away from the car. The offender was in view for 29 s with a full or partial view of her face available for 19 s.

Following the retention interval, participants were told, “*Now we would like to show you some photo lineups related to the three videos you saw earlier. There will be one lineup for each of the three videos you watched at the beginning of this experiment. One will relate to the male who littered in the park, another to the female thief who took money from a wallet* *in a car, and another to the male thief who stole a bottle from the liquor store. In each lineup the culprit may or may not be present*." Each lineup comprised six lineup photos showing a front view from the chest up with a neutral facial expression and background (200 × 200 pixels). Target-present lineups included the offender and five fillers. Target-absent lineups included a designated innocent suspect and the same five fillers. The position of lineup members in each lineup was random for each participant. Lineup fillers were chosen by description-match (Wells et al., 1998) and all lineup members wore the same clothes.

Lineup fairness. We tested lineup fairness and effective size (Malpass, 1981; Tredoux, 1999), following the mock eyewitness paradigm originated by Doob and Kirshenbaum (1973; see also Malpass & Lindsay, 1999) to provide performance indices that can be compared across reported identification experiments (Malpass, Tredoux, & McQuiston-Surrett, 2007). Six independent observers of the stimulus videos provided a free-recall description of each offender. These were used to create modal descriptions. Eighty-one participants (in addition to the 302 participants who completed the main experiment) were shown the target-present and target-absent lineups for all three offenders. Lineup order was randomised for each participant. Participants indicated for each lineup which member best fit the modal description.

We used these responses to calculate two measures. First, how frequently the suspect (guilty or innocent) was chosen as the best-match for the modal description, as compared with the rate at which a lineup member would be expected to be chosen by chance (for a six person lineup: 1 in 6 times, or by 17% of witnesses). The values for each lineup are shown in Table 1. The average frequency of suspect selections across lineups was 19.46% (95% CI [9.05,23.87], *SD* = 12.53). These values are similar to averages reported by other researchers for six-person description-matched lineups (e.g., Dobolyi & Dodson, 2013, 18% [10,26]; Meissner et al., 2005, 18% [15,20]).

Second, we used best-match choices to calculate Tredoux's *E* for each lineup (Tredoux, 1998; see also Malpass, 1981). Tredoux's *E* considers both lineup bias and size, and has a range of 1 to 6 (the nominal size) for six person lineups. The value of *E* decreases depending on how many lineup members are chosen as the best-match at below-chance rates. The values for each lineup are shown in Table 1. Average *E* across our lineups was 3.38 (95% CI [3.21,4.90]), comparatively lower than those reported for six-person lineups by Meissner et al. (2005), 4.99 [4.80, 5.19]; and Dobolyi and Dodson (2013), 4.66 [4.26, 5.07], but close to the range described by Gronlund et al. (2009) as fair (3.75-4.51). Average *E* was lowered by the park offender lineups. Interestingly, the filler lineup member who drew by far the most best-match choices in the target-absent park lineup attracted no picks in the target-present lineup, highlighting the difficulty of interpreting lineup fairness measures across context (e.g., Malpass & Lindsay, 1999; Wells & Bradfield, 1999), including how fairness measures relate to actual identification performance for the same lineup by real and mock eyewitnesses who have seen the guilty suspect. As an additional measure, following their best-match selection, we asked mock witnesses to indicate all lineup members that matched the modal description for each lineup (see Table 1). The mean number of photos selected was 3.34 (*SD* = 0.63).

Table 1

Measures of Lineup Fairness and Effective Size as a Function of Stimulus Event (Video) Offender and Target Presence

		Fairness Measures					
		Suspect Best-Match		Effective Size		No. Lineup Members Matching Modal Description	
Offender	Target Presence (TP/TA)	Proportion (%)	95% CIs	Tredoux's *E* (Range: 1-6)	95% CIs	*M (SD)*	95% CIs
Park	TP	35	[.27, .49]	2.88	[2.57, 3.29]	2.79 (1.01)	[2.57, 3.01]
	TA	5*	[.00, .10]	1.84	[1.51, 2.37]	2.37 (1.26)	[2.10, 2.64]
Liquor store	TP	15*	[.07, .23]	3.99	[3.26, 5.13]	3.37 (1.54)	[3.03, 3.71]
	TA	23*	[.14, .33]	4.71	[3.96, 5.80]	3.31 (1.43)	[3.00, 3.62]
Wallet	TP	7*	[.02, .13]	3.27	[2.66, 4.24]	4.11 (1.17)	[3.86, 4.36]
	TA	10*	[.03, .16]	3.56	[2.98, 4.41]	4.06 (1.34)	[3.77, 4.35]

N = 81. *Note:* * indicates 95% confidence intervals include or are lower than the proportion expected by chance (17%).

Procedure

After giving informed consent, participants completed the entire experiment on PC in individual cubicles using a mouse to click on-screen buttons in response to written prompts. The experiment comprised a stimulus event phase, a 20-25 min retention interval, and the lineup phase. In the stimulus event phase, participants were informed: "*You are going to be shown a series of short videos. Pay close attention to them because you will be asked some questions later*." Participants were then told to click the *Next* button when they were ready to watch the video. A count-down was presented in the centre of the screen and the video started when the count-down expired. When the first video ended, participants were told "*Here is the next video. Pay close attention to it because you will be asked some questions later."* After watching all three videos, participants completed an unrelated word recognition experiment that lasted 20-25 min.

In the lineup phase, participants completed an identification lineup for the offender in each video, in the same order as the videos were presented. For each lineup, participants were given unbiased instructions (Malpass & Devine, 1981) that the offender may or may not be present. The decision itself and response latency were recorded for all decisions. We varied lineup type on three levels (simultaneous; sequential; presence-recollection) between-subjects.

In the *simultaneous* lineup condition, the six lineup photos were centred on the screen in two rows of three faces, and a central *Not Present* button appeared at the bottom of the screen. Participants provided a single response for each lineup. They identified a lineup member as the offender from the video (by clicking on the photo), or clicked the *Not present* button.

In the *sequential* lineup condition, the six lineup photos were presented sequentially, centred on the screen. *Yes* and *No* response buttons appeared centred at the bottom of the screen. Participants were not told how many photos were in the lineup and had as much time to respond to each photo as they desired. A *Yes/No* decision was required for each face. If the witness chose the presented lineup member as the offender (by clicking *Yes*), the lineup ended. If the witness did not choose the lineup member presented as the offender (by clicking *No*), the next face was shown, until a lineup member was chosen or the lineup ended. We used this stopping rule (i.e., participants did not make yes/no decisions for the remaining faces in a lineup after giving a *Yes* response to a lineup member) as our focus was on testing memory and decision processes used when participants could not hedge their bets on a *Yes* decision, knowing that they could also say *Yes* for later lineup members. We tested participants for three offenders, therefore allowing them to make decisions for every lineup member in the first lineup would have alerted them to this possibility in following lineups).

In the *presence-recollection* lineup condition, the six lineup photos were centred on the screen in two rows of three faces, and two buttons appeared centred at the bottom of the screen: *Present* and *Not Present*. Participants were informed that they would make two identification decisions for each lineup, and given *recollection instructions* that outlined aspects of the phenomenology of remembering that are diagnostic of accurate recollection, as we described previously (N. Guerin et al., 2016b, Experiment 3). Specifically, participants were instructed:

> "*Research has shown that accurate decisions about the culprit's face being present are likely to be accompanied by more vivid and clear memories of seeing the face in the video than when a face that has not been seen before is inaccurately and mistakenly recognised. These can include memories of how events or items from the video were related to your thoughts about the culprit's face and other thoughts and feelings that were experienced when the culprit's face was seen.*"

Participants were also told:

> "*We would like you to use this information to accurately decide when a culprit's face is present in the lineup and when it is not. When you think that the culprit's face is present, think about how clear, vivid, and detailed your memory is. Can you remember the thoughts and feelings you experienced when you saw the culprit's face in the video? Remember, a more clear, vivid, and detailed memory of the culprit's face is more likely to be accurate.*"

Following these instructions, participants made their first decision by indicating whether the offender from the video was *Present* or *Not Present* in the lineup. Then, if they had indicated that the offender was *Present*, they were shown the lineup again (without any accompanying response buttons), and asked to click on the face identified as the offender to provide their second decision response. If they had indicated that the offender was not present, they were presented with the next lineup, until the experiment ended.

Results

Overview

We used an alpha level of .05 for all inferential analyses. Our focus was on trial-by-trial performance and we used logistic mixed-effects regression models to test predictions for differences in identification performance by lineup procedure (Barr, Levy, Scheepers, & Tily, 2013; Judd, Westfall, & Kenny, 2012). Identification decision (lineup rejection, positive identification) was the main outcome variable. Models were constructed to evaluate the effects of lineup type on choosing (response bias) and discriminability (accuracy, regardless of response bias) in planned pairwise comparisons (Gelman, Hill & Yajima, 2012). Crossed random effects of intercept by participant and stimulus, and slope by participant for target-presence (as varied within-subjects), and by stimulus for target-presence and lineup type (as varied within-stimulus) were included in models. Accounting for these sources of random variability avoids the risk of missing or artificially inflating real effects that is incurred when analyses instead focus on aggregate measures based on a binary outcome variable (e.g., Baayen, Davidson, & Bates, 2008; Gelman & Hill, 2007; Murayama Sakaki, Yan & Smith, 2014; Quene & van den Bergh, 2008). In eyewitness identification paradigms, bias and discriminability can vary considerably between lineups, offenders, and witnesses, even when using the same lineup procedure (Brewer & Palmer, 2010). This reflects the variability of real witnesses in responses to police eyewitness identification procedures (e.g., Cutler & Penrod, 1995; Horry et al., 2011). Therefore, to increase generalizability, we aimed to test key effects across this variability with multiple offenders (Brewer & Wells, 2011; Gronlund, Mickes, Wixted, & Clark, 2015). Data were analysed using the lme4 package (Bates, Maechler, Bolker, & Walker, 2014) for R (R Core Team, 2014).

Logistic regression models that predict a binary categorical outcome variable give the log odds of observing a given outcome for specific values of the predictors. Log odds values are symmetrical around 0. Values greater than 0 indicate the predicted outcome (i.e., a positive identification decision) is more likely, and values lower than 0 indicate the alternative outcome (i.e., a lineup rejection) is more likely. Coefficients of individual binary fixed factors in a model represent the log odds ratio of the two factor levels (lnOR). We examined these coefficients in maximal models to evaluate the significance of the effects of specific factors on the outcome variable. Confidence intervals (CIs) for coefficients for fixed factors (lnOR ± [1.96 * *SE*]) were interpreted to show a statistically significant effect if they excluded 0. When discussing comparisons of the levels of binary fixed effects, we also report odds ratios (ORs). ORs are the exponents of the log odds ratio (lnOR) for the two levels of the fixed factor given by the regression model. Aggregate descriptive statistics are presented when relevant and as a familiar index of identification performance (see Table 6).

Positive identifications of known-innocent lineup fillers are known errors and provide evidence of suspect innocence (e.g., Clark & Wells, 2008; Wells & Olson, 2002). Therefore, filler identifications are not always important for evaluating the practical reliability of a lineup procedure as a test of suspect guilt (except insofar as they can be an indication of lineup fairness, Wells, 2008). However, patterns of filler identifications also index memory and decision processes, and can therefore have important theoretical implications (e.g., in terms of the use of familiarity and recollection). Therefore, we report analyses both with and without filler identifications. When filler identifications (from both target-absent and target-present lineups) were excluded (i.e., removed from the dataset, not coded as lineup rejections), our measure of discriminability indexed the accuracy of positive identification decisions (i.e., witnesses' ability to discriminate guilty from innocent suspects). When filler identifications were included, discriminability indexed the accuracy of target *presence* decisions, regardless of whether the lineup member identified was the suspect (innocent or guilty) or a filler.

Accuracy and Decision Type

Our main research question was whether witnesses using the presence-recollection procedure would demonstrate better performance than those using either standard simultaneous or standard sequential lineup procedures. In models used to test predictions for the effects of lineup on decision type (lineup rejection or positive identification), lineup, target presence (target-absent = 0, target-present = 1), and their interactions were fixed factors. The coefficient for lineup indexed differences between procedures in witnesses' tendency to choose (response bias), while the coefficient for target presence indexed the accuracy of choosing decisions, independent of bias (discriminability). Variation in discriminability between lineups was thus indexed by the lineup × target presence interaction term (De Carlo, 1998; Wright & London, 2009). Planned pairwise comparisons were conducted between simultaneous and presence-recollection procedures to test our predictions. We also compared sequential and presence-recollection procedures, and simultaneous and sequential procedures, to further explore differences in identification performance across lineup types. Results are most easily interpreted with reference to plots of the models (Gelman & Hill, 2007).

Simultaneous vs presence-recollection. We predicted that if witnesses made better use of available recollection when using the presence-recollection procedure rather than the simultaneous procedure in making their identification decisions, greater discriminability would be observed for the modified procedure. On the contrary, we found no significant difference in choosing or discriminability between presence-recollection and simultaneous lineups, whether or not filler identifications were included in analyses (see Tables 2 & 3, Figure 1). Further, examination of Figure 1 reveals a marginally significant difference in discriminability between lineups in the opposite direction than predicted. Witnesses in the simultaneous procedure tended to be better at differentiating guilty from innocent suspects than those using the presence-recollection procedure. Finally, a model predicting the likelihood of a filler identification (other response; filler identification) showed a significantly greater likelihood of a filler identification for target-absent than target-present lineups (target presence was a significant predictor), as would be expected for fair target-absent lineups. But this did not differ with lineup type, as demonstrated by the absence of a significant effect of lineup or the target presence × lineup interaction term (Tables 4 & 5). However, as the aggregate proportions in Table 6 also demonstrate, the likelihood of false identification was low for both procedures, suggesting that performance was near ceiling for target-absent lineups. Therefore, this suggestion of a difference in discriminability might be artefact.

Table 2

Model Fit Statistics for Mixed-Effects Model Predicting Decision Type (Positive Identification; Lineup Rejection) by Lineup Type

Filler Identifications	Comparison	χ^2	*df*	*p*
Included	Simultaneous (Sim) vs Sequential (Seq)	9.44	3	.023
	Sim vs Presence-Recollection (PR)	5.53	3	.14
	Seq vs PR	8.38	3	.039
Excluded	Sim vs Seq	63.72	3	< .001
	Sim vs PR	16.10	3	.001
	Seq vs PR	37.20	3	< .001

Number of observations, fillers included (fillers excluded): Sim vs Seq = 600 (413), *n* = 200 (194); Sim vs PR = 606 (400), *n* = 202 (195); Seq vs PR: 606 (429), *n* = 202 (197).

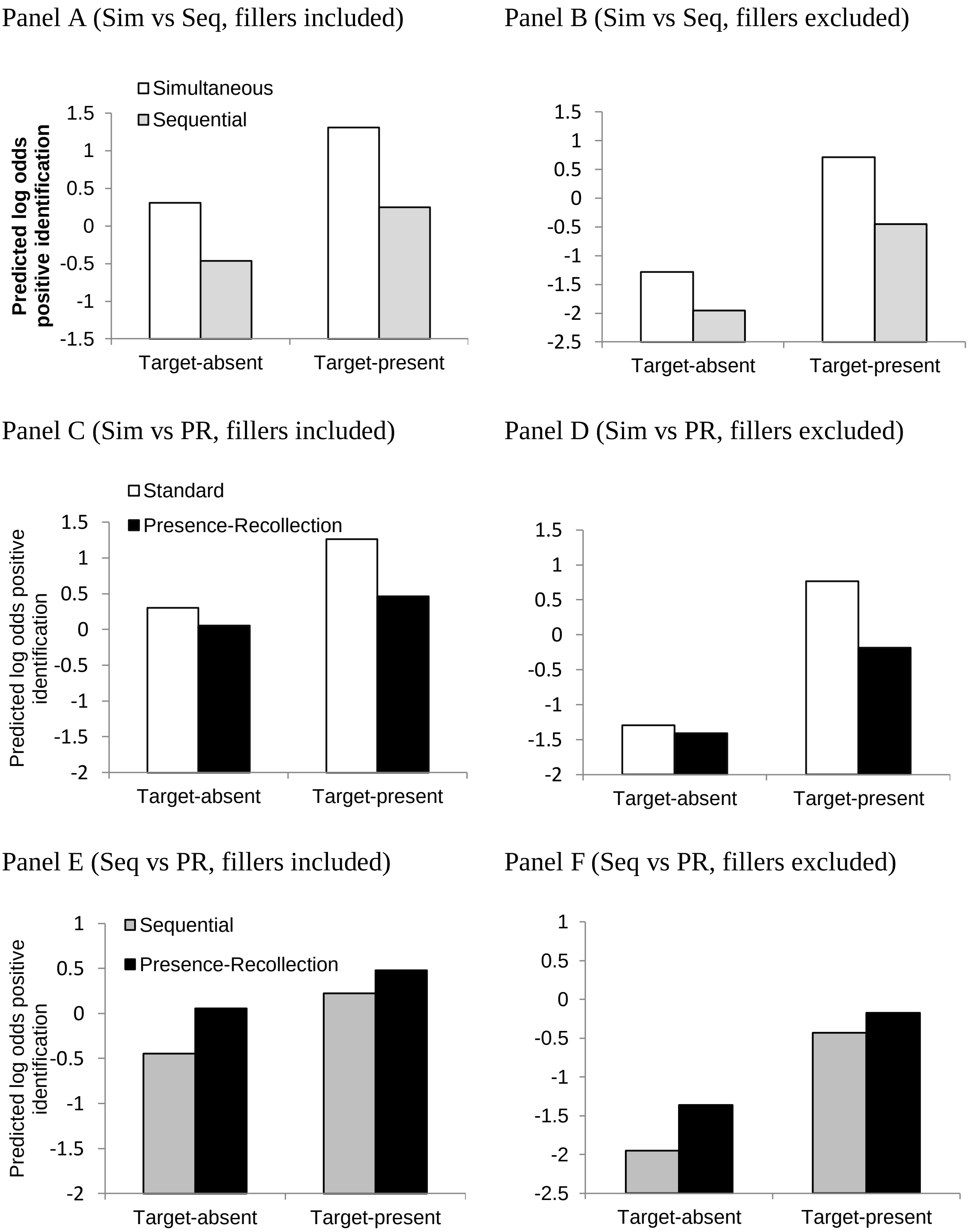


Figure 1. Plots of predicted log odds of a positive identification decision by lineup type and target presence. Values greater than 0 indicate a positive identification from the lineup is more likely, and values lower than 0 indicate a lineup rejection is more likely. Coefficient standard errors, indicative of the variability of the estimates, are displayed in Table 3.

Table 3

Fixed Effect Coefficients for Mixed-Effects Model Predicting Decision Type (Lineup Rejection; Positive Identification) by Lineup Type

			Lineup Comparisons								
			#Sim vs Seq			#Sim vs PR			#Seq vs PR		
Filler Identifications	Effect		*b*	SE_b	95% CI_b (*SD*)+	*B*	SE_b	95% CI_b (*SD*)	*b*	SE_b	95% CI_b (*SD*)
Included	Random	Participant (P)			(0.34)						(0.00)
		Stim			(0.16)						(0.24)
		TP|Stim			(0.48)			(0.15)			(0.35)
		Lineup|Stim			(0.11)			(0.45)			NA
		Lineup*TP|Stim			(0.02)			(0.22)			NA
	Fixed	Intercept	0.31			0.30			0.06		
		Lineup	-0.77*	0.25	-1.26, -0.27	-0.25	0.24	-0.71, 0.21	-0.50*	0.25	-0.96, -0.04
		TP	1.00*	0.38	0.25, 1.74	0.96*	0.37	0.23, 1.69	0.42	0.31	-0.19, 1.03
		Lineup × TP	-0.29	0.36	-0.99, 0.41	-0.55	0.38	-1.29, 0.19	0.25	0.34	-0.41, 0.91

Excluded	Random	Stim			0.66			0.60			0.76
		Lineup|Stim			0.74			0.71			0.58
	Fixed	Intercept	-1.29			-1.30			-1.95		
		Lineup	-0.67	0.41	-1.47, 0.14	-0.11	0.37	-0.84, 0.61	0.59	0.41	-0.21, 1.39
		TP	2.00*	0.35	1.31, 2.69	2.06*	0.51	1.06, 3,06	1.52*	0.36	0.82, 2.22
		Lineup × TP	-0.50	0.50	-1.48, 0.48	-0.83	0.47	-1.76, 0.10	-0.33	0.48	-1.27, 0.61

Note: * indicates significant predictors, # indicates the reference lineup condition. + Standard Deviation (*SD)* reflects variability due to the random effect. Number of observations, fillers included (excluded): Sim vs Seq = 600 (413), *n* = 200 (194). Sim vs PR = 606 (400), *n* = 202 (195). Seq vs PR = 606 (429), *n* = 202 (197).

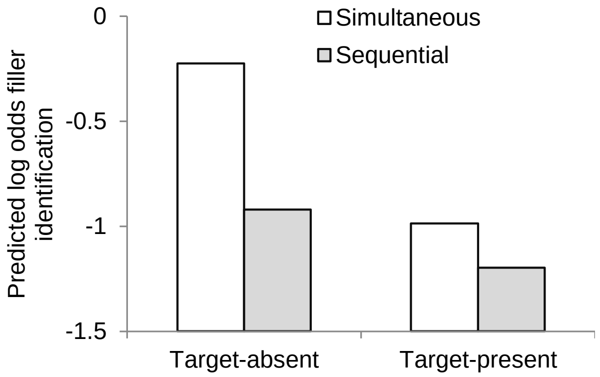
Table 4

Model Fit Statistics for Mixed-Effects Model Predicting Filler Identification (Other Response; Filler Identification) by Lineup Type

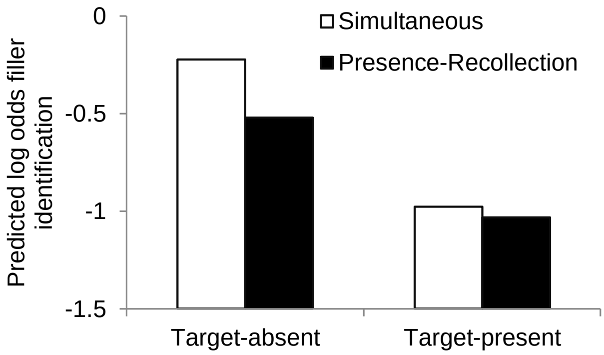
Comparison	χ^2	*df*	*p*
Sim vs Seq	17.35	3	< .001
Sim vs PR	14.76	3	.002
Seq vs PR	7.72	3	.052

Number of observations: Sim vs Seq = 600, n = 200; Sim vs PR = 606, n = 202; Seq vs PR: 606, n = 202

Panel A (Sim vs Seq)



Panel B (Sim vs PR)



Panel C (Seq vs PR)

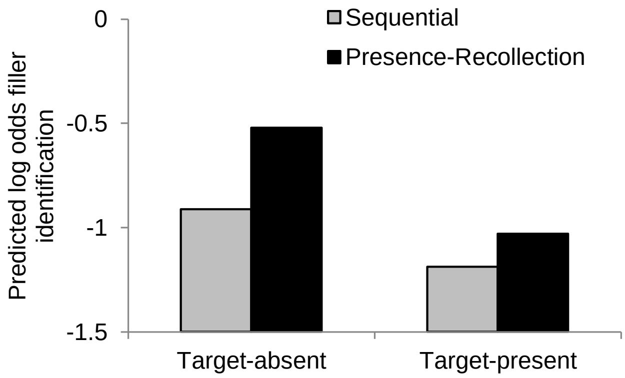


Figure 2. Plots of predicted log odds of a filler identification by lineup type and target presence. Values greater than 0 indicate a filler identification from the lineup is more likely, and values lower than 0 indicate a lineup rejection is more likely. Coefficient standard errors, indicative of the variability of the estimates, are displayed in Table 5.

Table 5

Fixed Effect Coefficients for Mixed Effects Model Predicting Filler Identification (Other Response; Filler Identification) by Lineup Type

		Lineup Comparisons								
		#Sim vs Seq			#Sim vs PR			#Seq vs PR		
Effect		*b*	SE_b	95% CI_b (*SD*)+	*b*	SE_b	95% CI_b (*SD*)	*b*	SE_b	95% CI_b (*SD*)
Random	Stim			(0.35)			(0.29)			(0.28)
Fixed	Intercept	-0.23			-0.22			-0.91		
	Lineup	-0.70*	0.26	-1.18, -0.21	-0.30	0.24	-0.76, 0.17	0.39	0.25	-0.10, 0.88
	TP	-0.76*	0.25	-1.25, -0.27	-0.76*	0.25	-1.24, -0.27	-0.28	0.26	-0.80, 0.24
	Lineup × TP	0.48	0.36	-0.23, 1.20	0.25	0.35	-0.44, 0.94	-0.23	0.36	-0.95, 0.48

Note: * indicates significant predictors, # indicates the reference lineup condition. + Standard Deviation (*SD*) reflects variability due to the random effect. Number of observations: Sim vs Seq = 600, *n* = 200; Sim vs PR = 606, *n* = 202; Seq vs PR = 606, *n* = 202.

Overall, results were not consistent with an underlying difference in the use of recall-to-reject (Rotello & Heit, 2000) or other recollection monitoring processes between lineups. Rates of positive identifications from target-absent lineups did not significantly differ with lineup type. Similarly, the significantly lower likelihood of positive identification decisions for target-present lineups with the presence-recollection procedure was not due to a significantly lower likelihood of familiarity-based filler identifications than from the simultaneous procedure. In sum, we found no evidence of superior use of available recollection by witnesses in the presence-recollection procedure.

Simultaneous vs sequential. Tables 2 & 3 also show model statistics from our comparison of identification responses for the standard simultaneous and sequential procedures (see Tables 2 & 3). Choosing was significantly lower (OR = 0.46, 95% CI = [0.28, 0.76] for sequential than simultaneous lineups, when filler identifications were included. The odds of a witness following the simultaneous procedure making a positive identification decision were nearly twice as high as those for a witness following the sequential procedure. However, discriminability did not differ significantly between lineup types (the target presence × lineup interaction term was not a significant predictor). Witnesses in the sequential procedure were less willing to identify a lineup member as the offender, but were no better at differentiating target-present from target-absent lineups than simultaneous witnesses. When filler identifications were excluded from the model, a similar pattern of responses was observed. However, lineup was not a significant predictor, demonstrating no significant difference between lineup procedures in the likelihood of identifying the suspect. Filler identifications were significantly less likely from the sequential than the simultaneous procedure (see Tables 4 & 5, Figure 2), and this did not differ with target presence. Thus, the lower response bias demonstrated by witnesses using the sequential procedure was largely driven by a lower likelihood of incorrectly identifying a known-innocent filler as the offender, without a significantly lower likelihood of accurately identifying a guilty offender.

Sequential vs presence-recollection. Consistent with the pairwise comparisons of the simultaneous procedure with sequential and presence-recollection procedures, choosing was also significantly lower with the sequential than the presence-recollection procedure, whether or not filler identifications were included (see Tables 2 & 3, Figure 1). However, neither target presence nor the interaction term significantly predicted the likelihood of witnesses making a positive identification from a lineup. While witnesses in the sequential procedure were less willing to identify a lineup member as the offender than those in the presence-recollection procedure, there was no significant difference between procedures in witnesses' ability to differentiate target-present from target-absent lineups. A model predicting filler identifications showed no significant effects of lineup, target presence or the interaction term (see Tables 4 & 5, Figure 2). In sum, the pattern of identification responses for the sequential procedure differed from that for the presence-recollection procedure only in a slightly lower response bias and provide no evidence for any difference in the evidential basis of witnesses' decisions.

Table 6

Proportion of Identification Responses (Hit rates, Filler ID Rates, Correct Rejection Rates and False ID Rates) by Lineup Type

		Response Proportion						
		Target-present			Target-absent			
Lineup	Stimulus	Hits	Filler IDs (TP)	Number of observations	Correct Rejections	Filler IDs (TA)	False IDs	Number of observations
Sim	Liquor store	.38	.30	50	.46	.50	.04	50
	Park	.52	.19	48	.42	.33	.25	52
	Wallet	.16	.29	51	.59	.31	.10	49
	All	.35	.26	149	.49	.38	.13	151
Seq	Liquor store	.24	.35	51	.59	.35	.06	49
	Park	.61	.13	48	.57	.30	.13	52
	Wallet	.15	.20	51	.64	.22	.13	49
	All	.32	.23	150	.60	.29	.11	150

PR	Liquor store	.33	.44	52	.48	.50	.02	50
	Park	.76	.14	49	.34	.47	.19	53
	Wallet	.38	.27	56	.48	.35	.17	46
	All	.48	.29	157	.43	.44	.13	149

Discussion

An experiment using three crime event stimuli showed no greater discriminability for the presence-recollection procedure than for standard simultaneous or sequential procedures. Specifically, we found no evidence of better use of recollection with the presence-recollection procedure in an identification paradigm. These results raise important questions about the use of familiarity and recollection, and metacognition, in identification tasks. Choosing was more conservative with the sequential than both simultaneous and presence-recollection procedures, which did not differ significantly. Considering filler identification patterns, these effects on choosing are particularly pertinent to discussion about simultaneous-sequential comparisons (see, for example, Amendola & Wixted, 2015; Mickes, 2015; Wells, Dysart, & Steblay, 2015; Wells et al., 2015; Wells, Smith and Smalarz, 2015; Wixted & Mickes, 2014, 2015). Specifically, results provide evidence against Wixted and Mickes' (2014) signal-detection-based diagnostic-feature-detection hypothesis. We consider potential explanations for results, and outline theoretical and applied issues for further investigation.

The ineffectiveness of the presence-recollection lineup in the identification paradigm suggests that the greater discriminability observed in previous mini-lineup experiments might have been due to greater use of recall-to-reject (Rotello & Heit, 2000; see also recollection rejection, Brainerd, Reyna, Wright, & Mojardin, 2003), rather than post-retrieval editing (e.g., Dodson & Schacter, 2002; Gallo et al., 2006; S. A. Guerin, Robbins, Gilmore, & Schacter, 2012). Recall-to-reject dissociates from post-retrieval metacognitive monitoring processes such as the distinctiveness heuristic (Schacter, Israel, & Racine, 1999) when little recollectible detail has been encoded or retained (Gallo et al., 2006), as might have occurred in this experiment. Clearly, recall-to-reject cannot improve performance if scant recollection is available (e.g., Peters et al., 2008). In contrast, the distinctiveness heuristic could improve performance with little available recollection. This heuristic is based on the idea that the amount of recollection at test is expected to be higher for a studied target than an unstudied foil. Therefore, the relative absence of recollected detail would contribute to accurate rejection of foils. Accordingly, the impact on eyewitness performance in the presence-recollection lineup of limited available recollectible detail would differ, depending on the mechanism underlying better use of recollection: concrete recall-to-reject; the distinctive heuristic; or other post-retrieval memory-editing. Our results suggest the possibility that the presence-recollection lineup encourages recall-to-reject or recollection-dependent post-retrieval monitoring, but that insufficient recollection of offenders was available in our task for this to affect discriminability.

Several features of the identification task might have limited the recollection available to eyewitnesses. Encoding recollectible detail of targets' faces viewed in a video crime scenario might have been more difficult than from photos. The additional cognitive load of encoding episodic detail from videos might have led to impoverished face encoding, despite longer exposure. Moreover, matching video stimuli to photos can be difficult for unfamiliar faces (Bruce et al., 1999). Test photos in the mini-lineups were also more consistent with studied photos, than with video stimuli (Bruce et al., 1999; Hancock, Bruce, & Burton, 2000). The longer retention interval in the identification paradigm might also have hindered retention and limited subsequent recollection. If so, our results suggest recall-to-reject was responsible for the presence-first advantage observed with mini-lineups, or that post-retrieval memory editing in identification tasks can only be improved with adequate recollection.

This potential explanation has both theoretical and applied significance. In theoretical terms, it is important to establish whether the active components of the presence-recollection procedure in mini-lineup experiments increase recollection chiefly by disrupting best-match-first decision-making, or by improving eyewitnesses' metacognitive appraisal of already-retrieved evidence. Greater knowledge of the accessibility of retrieval and monitoring processes to metacognitive intervention in lineup tasks is important to establish the limits of our ability to improve eyewitness decision-making. Further, establishing the limits of metacognitive monitoring and control in eyewitness tasks will also inform theory relating to these processes in recognition memory more generally.

In practical terms, if recall-to-reject or better monitoring of recollected detail was largely responsible for the presence-recollection advantage with mini-lineups, the procedure will only be useful when witnesses have access to recollectible detail. In practice, such detail might often be limited. For example, poor exposure conditions are likely to limit encoding of recollective detail more than familiarity (Gardiner & Parkin, 1999). Identification tests can also be conducted when considerable time has elapsed since the crime, when both familiarity and recollection will have degraded (Yonelinas, 2002). However, police frequently use show-ups or lineups soon after the commission of a crime (Wogalter et al., 2004). In these cases, a procedure that improves use of recollection would be valuable to avoid expending limited resources in investigating innocent suspects, and lower the risk of erroneous prosecutions.

Two immediate targets for further investigation of the presence-recollection lineup are suggested. First, given the potential implications of our findings in the mini-lineup paradigm, it is important to rule out the possibility that results were due to a lack of recollectible detail, rather than a failure to translate to an identification paradigm. To that end, the procedure should be tested further in an identification paradigm with a variety of stimuli and test conditions, and explicit variation in available recollection. Second, to isolate mechanisms underlying the presence-recollection advantage in mini-lineups, experimental dissociation of recall-to-reject from post-retrieval editing mechanisms that do not rely on detailed recollection will be informative. For example, by varying test type (e.g., S. A. Guerin et al., 2012) or response lag (e.g., Hunt & Smith, 2014).

Paradigm differences in the conduct of lineup tasks offer two further potential explanations for the lack of effectiveness of the presence-recollection lineup in this identification experiment. First, even in the lab with video stimuli and unbiased instructions (Malpass & Devine, 1981), the identification paradigm might have more strongly cued participants' beliefs about the purpose of a lineup and the desirability of positive identifications, than the face-recognition paradigm. Task expectations can influence decision-making, memory retrieval, and metacognitive appraisal of remembered evidence (e.g., Dodson & Schacter, 2002; Johnson, Hashtroudi, & Lindsay, 1993; Odegard & Lampinen, 2006). Ideas about police lineup procedures loom large in shared understandings of the justice system that are constantly worked over in the public domain. Lay people (e.g., mock-jurors) show high regard for identification evidence and, by extension, the reliability of identification procedures (Lindsay, Wells, & O'Connor, 1989; Semmler, Brewer, & Douglass, 2011). Moreover, eyewitness willingness to choose from a lineup strongly suggests a belief that the task is to assist police to "get the guy" (Malpass & Devine, 1981; Wells, 2014). This imperative might have rendered the novel procedure's inoculation against best-match-first decision-making ineffectual.

We proposed that willingness to pick from the lineup triggers best-match-first decision processes. Therefore, strong cues for choosing would also be expected to activate familiarity-based responding and promote neglect of recollection. In the presence-recollection procedure, stronger activation of eyewitness beliefs might have overridden procedural cues for recollection and encouraged familiarity-based decision-making. For innocent lineup members who closely resemble the offender, familiarity evidence would not provide evidence of mismatch with the offender. Only recollection of specific offender details would offset their misleading familiarity. Therefore, if greater motivation to "get the guy" in the identification paradigm triggered best-match-first decision-making in both simultaneous procedures, this would explain the lack of difference in discriminability.

This explanation is supported by evidence from our previous tests of a presence-first procedure *without metacognitive instructions*, that demonstrated greater discriminability than a simultaneous procedure when used by naïve mock-eyewitnesses (with no experience of a more realistic identification task, N. Guerin et al., 2016a), but not when tested with non-naïve mock-eyewitnesses (N. Guerin et al., 2016b). Even in the less realistic identification context, non-naïve participants' expectations for lineup tasks still appear to have encouraged a best-match-first approach.

Converging evidence also comes from a recent test of a novel grain size (simultaneous) lineup procedure, for which mock eyewitnesses could indicate several potential candidates as matches for the offender, if their confidence for the match of a single lineup member was low (Horry, Brewer, & Weber, 2016). Regardless of confidence and accuracy, participants rarely indicated more than one lineup member, suggesting they held strong preconceptions about what a lineup task requires. This potential explanation for our results suggests that more extensive structural modification might be needed to counter familiarity-based responding.

Second, presence decisions in the presence-recollection procedure in the face recognition paradigm were made for all trials in a block. Then, in a second test block, mini-lineups designated target-present by participants were presented again, for offender-selection responses. Making first-decisions for multiple lineups without needing to point out any lineup members, could have helped to forestall familiarity-based decision-making, even when participants knew they might need to indicate guilty offenders later. In contrast, mock eyewitnesses in the identification experiment we report here made both component decisions for an offender, before being presented with a lineup for the next offender. Interleaving component decisions would have disrupted any facilitating effect of blocking.

However, tests of a similarly blocked presence-first lineup *without recollection instructions*, in which witnesses were not told that the second offender-selection decision might be required (N. Guerin et al., 2016a, 2016b), provide evidence against this explanation. While naïve mock eyewitnesses following the presence-first procedure demonstrated greater discriminability than those making standard simultaneous decisions, non-naïve witnesses did not. Non-naïve participants' expectations of the lineup task appear to have affected their decision-making, while blocking the two-step decisions, in itself, did not. Therefore, interleaving component decisions for offenders in the identification paradigm does not compellingly explain paradigm differences in results.

Our findings also have implications for broader issues in the eyewitness identification field. A more conservative response bias without significantly different discriminability was evident for sequential than simultaneous lineups. Choosing was also more conservative for sequential than presence-recollection lineups. Interestingly, the more conservative response bias for sequential lineups was largely due to a lower likelihood of known-innocent filler identifications from both target-absent and target-present lineups. Therefore, the lower likelihood of positive identifications overall in the sequential procedure had little impact on either the likelihood of a false suspect identification, or an accurate identification of the guilty. Wixted and Mickes' (2014) signal-detection-based diagnostic-feature-detection model predicts that simultaneous presentation of lineup members will enhance discriminability contra sequential presentation. Specifically, simultaneous presentation is proposed to make immediately apparent the misleadingly familiar features that are shared by all lineup members, and which are consequently non-diagnostic. In contrast, sequential presentation does not allow eyewitnesses to rule out as non-diagnostic those features of fillers or innocent suspects that are shared with memory for the offender, but also with other lineup members. Clearly, our results are not consistent with these predictions. We found simultaneous presentation in the standard procedure was no more effective than sequential presentation in allowing eyewitnesses to differentiate a guilty suspect from similar-looking foils, or avoid making a false identification from a target-absent lineup. These results highlight the importance of patterns of filler identification from both target-absent and target-present lineups, for the comparison of eyewitness performance between procedures, but also for developing theoretical understandings of eyewitness memory and decision processes (cf. Mickes, 2015; Wells et al., 2015).

This paper reports the first test of the presence-recollection procedure in an identification paradigm. We found no evidence that the procedure enabled greater discriminability than either standard simultaneous or sequential identification procedures. Replication of this finding would be potentially informative about the type of memorial information available to witnesses making lineup decisions, and their metacognitive monitoring processes. However, a conclusion about the potential usefulness of the procedure would be premature. Further investigation of the mechanisms underlying the presence-recollection advantage found in previous experiments, and testing under a variety of conditions in an identification paradigm, is needed. This experiment contributes to a growing body of theoretically-based research that aims to identify and explain regularities in patterns of eyewitness identification performance with reference to established knowledge of memory and decision processes in recognition. The presence-recollection procedure might prove ineffective in more realistic identification contexts. However, our previous observation and replication of a recollection-driven discriminability advantage for novel lineup tasks in a more basic recognition paradigm should motivate further research into ways to enhance the use of recollection, and eyewitness metacognition more generally, in realistic identification formats.

References

Amendola, K. L., & Wixted, J. T. (2015). Comparing the diagnostic accuracy of suspect identifications made by actual eyewitnesses from simultaneous and sequential lineups in a randomized field trial. *Journal of Experimental Criminology*, *11*, 263-284. doi: 10.1007/s11292-014-9219-2

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390-412. doi: 10.1016/j.jml.2007.12.005

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structures for confirmatory hypothesis testing: Keeping it maximal. *Journal of Memory and Language*, 68, 255-278. doi: 10.1016/j.jml.2012.11.001

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2014). _lme4: Linear mixed-effects models using Eigen and S4_. Retrieved from http://CRAN.R-project.org/package=lme4

Behrman, B. W., & Richards, R. E. (2005). Suspect/foil identification in actual crimes and in the laboratory: a reality monitoring analysis. *Law and Human Behavior*, *29*, 279-301. doi: 10.1007/s10979-005-3617-y

Brainerd, C. J., Reyna, V. F., Wright, R., & Mojardin, A. H. (2003). Recollection rejection: false-memory editing in children and adults. *Psychological Review*, *110*, 762-784. doi: 10.1037/0033-295x.110.4.762

Brewer, N., & Palmer, M. A. (2010). Eyewitness identification tests. *Legal and Criminological Psychology*, *15*, 77-96. doi: 10.1348/135532509x414765

Brewer, N., Weber, N., Wootton, D., & Lindsay, D. S. (2012). Identifying the bad guy in a lineup using confidence judgments under deadline pressure. *Psychological Science*, *23*, 1208-1214. doi: 10.1177/0956797612441217

Brewer, N., & Wells, G. L. (2011). Eyewitness identification. *Current Directions in Psychological Science*, *20*, 24-27. doi: 10.1177/0963721410389169

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J. B., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied, 5,* 339-360. doi: 10.1037/1076-898x.5.4.339

Clark, S. E. (2003). A memory and decision model for eyewitness identification. *Applied Cognitive Psychology*, *17*, 629-654. doi: 10.1002/acp.891

Clark, S. E. (2005). A re-examination of the effects of biased lineup instructions in eyewitness identification. *Law and Human Behavior, 29,* 395-424. doi: 10.1007/s10979-005-5690-7

Clark, S. E., & Davey, S. L. (2005). The Target-to-Foils Shift in Simultaneous and Sequential Lineups. *Law and Human Behavior, 29,* 151-172. doi: 10.1007/s10979-005-2418-7

Clark, S. E., Howell, R. T., & Davey, S. L. (2008). Regularities in eyewitness identification. *Law and Human Behavior, 32*, 187-218. doi: 10.1007/s10979-006-9082-4

Clark, S. E., & Wells, G. L. (2008). On the diagnosticity of multiple witness identifications. *Law and Human Behavior*, *32*, 406–422. doi: 10.1007/s10979-007-9115-7

Cook, G. I., Marsh, R. L., & Hicks, J. L. (2005). Revisiting the role of recollection in item versus forced-choice recognition memory. *Psychonomic Bulletin & Review*, *12*, 720-725. doi: 10.3758/bf03196763

Cutler, B. L., & Penrod, S. D. (1995). *Mistaken identification: The eyewitness, psychology and the law*. Cambridge University Press.

De Carlo, L. T. (1998). Signal detection theory and generalized linear models. *Psychological Methods*, *3*, 186-205. doi: 10.1037//1082-989x.3.2.186

Dobolyi, D. G., & Dodson, C. S. (2013). Eyewitness confidence in simultaneous and sequential lineups: A criterion shift account for sequential mistaken identification overconfidence. *Journal of Experimental Psychology: Applied,* 19, 345-357. doi: 10.1037/a0034596

Doob, A. N., & Kirshenbaum, H. M. (1973). Bias in police lineups – partial remembering. *Journal of Police Science and Administration,* 1, 287-293.

Dodson, C. S., & Schacter, D. L. (2002). When false recognition meets metacognition: The distinctiveness heuristic. *Journal of Memory and Language*, *46*, 782-803. doi: 10.1037/e501882009-058

Gallo, D. A., Bell, D. M., Beier, J. S., & Schacter, D. L. (2006). Two types of recollection-based monitoring in younger and older adults: Recall-to-reject and the distinctiveness heuristic. *Memory, 14*, 730-741. doi: 10.1080/09658210600648506

Gallo, D. A., Roberts, M. J., & Seamon, J. G. (1997). Remembering words not presented in lists: Can we avoid creating false memories? *Psychonomic Bulletin & Review, 4*, 271-276. doi: 10.3758/bf03209405

Gardiner, J. M., & Parkin, A. J. (1990). Attention and recollective experience in recognition memory. *Memory & Cognition*, *18*, 579-583. doi: 10.3758/bf03197100

Gardiner, J. M., & Richardson-Klavehn, A. (2000). Remembering and knowing. In E. Tulving & F. I. M. Craik (Eds.), The Oxford handbook of memory (pp. 229-244). London: Oxford University Press. doi: 10.1002/acp.822.

Garrett, B. (2011). Convicting the innocent: Where criminal prosecutions go wrong. Cambridge, MA: Harvard University Press. doi:10.4159/ harvard.9780674060982

Gelman, A., & Hill, J. (2007). *Data Analysis using Regression and Multilevel/Hierarchical Models.* New York: Cambridge University Press.

Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness, 5*(2), 189-211. doi: 10.1080/19345747.2011.618213

Goldstein, A. G., Chance, J. E., & Schneller, G. R. (1989). Frequency of eyewitness identification in criminal cases: A survey of prosecutors. *Bulletin of the Psychonomic Society*, *27*, 71-74. doi: 10.3758/bf03329902

Gronlund, S. D., Carlson, C. A., Daily, S. B., & Goodsell, C. A. (2009). Robustness of the sequential lineup advantage. *Journal of Experimental Psychology: Applied*, 15, 140-152. doi: 10.1037/a0015082

Gronlund, S. D., Mickes, L., Wixted, J. T., & Clark, S. E. (2015). Conducting an eyewitness lineup: How the research got it wrong. In B. H. Ross (Ed.), The psychology of learning and motivation (Volume 63, pp. 1– 43). Waltham, MA: Academic Press

Gross, S. R., Jacoby, K., Matheson, D. J., Montgomery, N., & Patil, S. (2005). Exonerations in the United States 1989 through 2003. *The Journal of Criminal Law and Criminology*, 95, 523–560.

Guerin, N., Weber, N., & Horry, R. (2016a). Split decision: An account of eyewitness decision-making in simultaneous lineups and a novel presence-first identification procedure. *Under review*

Guerin, N., Weber, N., & Horry, R. (2016b). Only recollect: Improving identification accuracy by reframing eyewitness metacognition and decision structure. *Under review*

Guerin, N., Weber, N., & Horry, R. (2016c). Do eyewitnesses recall-to-reject? Familiarity and recollection in simultaneous lineups. *Under review*

Guerin, N., Weber, N., & Horry, R. (2016d). Deceived by the obvious: Metacognitive instructions target neglect of recollection in eyewitness identification decisions. *Under review*

Guerin, S. A., Robbins, C. A., Gilmore, A. W., & Schacter, D. L. (2012). Retrieval failure contributes to gist-based false recognition. *Journal of Memory and Language*, *66*, 68-78. doi: 10.1016/j.jml.2011.07.002

Hancock, P. J., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences*, *4*(9), 330-337. doi: 10.1016/s1364-6613(00)01519-9

Horry, R., Memon, A., Wright, D. B., & Milne, R. (2011). Predictors of eyewitness identification decisions from video lineups in England: a field study. *Law and Human Behavior*, *36*, 257-265. doi: 1007/s10979-011-9279-z

Horry, R., Palmer, M. A., & Brewer, N. (2012). Backloading in the sequential lineup prevents within-lineup criterion shifts that undermine eyewitness identification performance. *Journal of Experimental Psychology: Applied*, *18*, 346-360. doi: 10.1037/a0029779

Horry, R., Brewer, N., & Weber, N. (2016). The Grain-Size Lineup: A Test of a Novel Eyewitness Identification Procedure. *Law and Human Behavior (in press).*

Hunt, R. R., & Smith, R. E. (2014). How distinctive processing enhances hits and reduces false alarms. *Journal of Memory and Language*, *75*, 45-57. doi: 10.1016/j.jml2014.04.007

Ingram, K. M., Mickes, L., & Wixted, J. T. (2012). Recollection can be weak and familiarity can be strong. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 325-339. doi: 10.1037/e520592012-514

Innocence Project (2016). Innocence project, retrieved January 16, 2016 http://www.innocenceproject

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3-28. doi: 10.1037//0033-2909.114.1.3

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology, 103*(1), 54-69. doi: 10.1037/a0028347

Lane, S. M., Roussel, C. C., Starns, J. J., Villa, D., & Alonzo, J. D. (2008). Providing information about diagnostic features at retrieval reduces false recognition. *Memory, 16*, 836–851. http://doi: 10.1080/0965821080233734

Lindsay, R. C., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology*, *70*, 556. doi: 10.1037/0021-9010.70.3.556

Lindsay, R. C., Wells, G. L., & O'Connor, F. J. (1989). Mock-juror belief of accurate and inaccurate eyewitnesses. *Law and Human Behavior*, *13*, 333-339. doi: 10.1007/bf01067033

Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology*, *57*, 335-384. doi: 10.1016/j.cogpsych.2008.02.004

Malmberg, K. J., & Xu, J. (2007). On the flexibility and the fallibility of associative memory. *Memory & Cognition, 35,* 545-556. doi: 10.3758/bf03193293

Malpass, R. S. (1981). Effective size and defendant bias in eyewitness identification lineups. *Law and Human Behavior,* 5, 299-309. doi: 10.1007/bf01044945

Malpass, R. S., & Devine, P. G. (1981). Eyewitness identification: Lineup instructions and the absence of the offender. *Journal of Applied Psychology*, *66*, 482. doi: 10.1037/0021-9010.66.4.482

Malpass, R. S., & Lindsay, R. C. L. (1999). Measuring line-up fairness. *Applied Cognitive Psychology,* 13, S1-S7. doi: 10.1002/(SICI)1099- 0720(199911)13:1(s$S1::AID-ACP678&3.0.CO;2-9

Malpass, R. S., Tredoux, C. G., & McQuiston-Surrett, D. E. (2007). Lineup construction and lineup fairness. In R. Lindsay, D. Ross, J. D. Read & M. P. Toglia (Eds.), *Handbook of Eyewitness Psychology (Vol. 2): Memory for People.* Lawrence Erlbaum & Associates.

Meissner, C. A., Tredoux, C. G., Parker, J. F., MacLin, O. H. (2005). Eyewitness decisions in simultaneous and sequential lineups: a dual-process signal detection theory analysis. *Memory & Cognition,* 33, 783-792. doi: 10.3758/bf03193074

Mickes, L. (2015). Receiver operating characteristic analysis and confidence-accuracy characteristic analysis in investigations of system variables and estimator variables that affect eyewitness memory. *Journal of Applied Research in Memory and Cognition, 4,* 83-102. doi: 10.1016/j.jarmac.2015.01.003

Murayama, K., Sakaki, M., Yan, V. X., & Smith, G. M. (2014). Type I error inflation in the traditional by-participant analysis to metamemory accuracy: A generalized mixed-effects model perspective. *Journal of Experimental Psychology: Learning Memory and Cognition, 40*(5), 1287-1306. doi: 10.1037/a0036914

National Research Council. (2014). I*dentifying the Culprit: Assessing Eyewitness Identification.* Washington, DC: The National Academies Press. doi: 10.17226/18891

National VIPER Bureau. (2009). Viper: Video Identification Parade Electronic Recording, retrieved January 16, 2016, http://www.viper.police.uk/

Odegard, T. N., & Lampinen, T. N. (2006). Memory editing: Knowledge, criteria, and alignment. *Memory, 14,* 777-787. doi: 10.1080/09658210600648589

Palmer, M. A., & Brewer, N. (2012). Sequential lineup presentation promotes less-biased criterion setting but does not improve discriminability. *Law and Human Behavior*, *36*, 247-255. doi: 10.1037/h0093923

Palmer, M. A., Brewer, N., McKinnon, A. C., & Weber, N. (2010). Phenomenological reports diagnose accuracy of eyewitness identification decisions. *Acta Psychologica*, *133*, 137-145. doi: 10.1016/j.actpsy.2009.11.002

Peters, M. J., Jelicic, M., Gorski, B., Sijstermans, K., Giesbrecht, T., & Merckelbach, H. (2008). The corrective effects of warning on false memories in the DRM paradigm are limited to full attention conditions. *Acta Psychologica*, *129*, 308-314. doi: 10.1016/j.actpsy.2008.08.007

Pozzulo, J. D., & Lindsay, R. C. L. (1999). Elimination lineups: An improved identification procedure for child eyewitnesses. *Journal of Applied Psychology*, *84*, 167-176. doi: 10.1037/0021-9010.84.2.167

Pryke, S., Lindsay, R. C. L., Dysart, J. E., & Dupuis, P. (2004). Multiple independent identification decisions: a method of calibrating eyewitness identifications. *Journal of Applied Psychology*, *89*, 73-84. doi: 10.1037/0021-9010.89.1.73

Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modelling with crossed random effects and with binomial data. *Journal of Memory and Language*, *59, 413-425.* doi: 10.1016/j.jml.2008.02.002

R Core Team. (2014). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org.

Rotello, C. M., & Heit, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory & Cognition*, *28*, 907-922. doi: 10.3758/bf03209339

Sauer, J. D., Brewer, N., & Weber, N. (2008). Multiple Confidence Estimates as Indices of Eyewitness Memory. *Journal of Experimental Psychology: General*, 137, 528-547. doi: 10.1037/a0012712

Sauerland, M., & Sporer, S. L. (2008). The application of multiple lineups in a field study. *Psychology, Crime & Law, 14*, 549-564. doi: 10.1080/10683160801972519

Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language*, *40*, 1-24. doi: 10.1006/jmla.1998.2611

Schooler, J. W., Gerhard, D., & Loftus, E. F. (1986). Qualities of the unreal. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 171-181. doi: 10.1037/0278-7393.12.2.171

Semmler, C., Brewer, N., & Douglass, A. B. (2011). Jurors believe eyewitnesses. In B. L. Cutler (Ed.), *Conviction of the innocent: Lessons from psychological research* (pp. 185–209). Washington, DC: American Psychological Association. doi: 10.1037/13085-009

Starns, J. J., Lane, S. M., Alonzo, J. D., & Roussel, C. C. (2007). Metamnemonic control over the discriminability of memory evidence: A signal detection analysis of warning effects in the associative list paradigm. *Journal of Memory and Language*, *56*, 592-607. doi: 10.1016/j.jml.2006.08.013

Steblay, N. K., Dysart, J. E., & Wells, G. L. (2011). Seventy-two tests of the sequential lineup superiority effect: A meta-analysis and policy discussion. *Psychology, Public Policy, and Law*, *17*, 99-139. doi: 10.1037/a0021650

Tredoux, C. G. (1999). Statistical considerations when determining measures of lineup size and lineup bias. *Applied Cognitive Psychology,* 13, S9-S26. doi: 10.1002/(SICI)1099-0720(199911)13:1($S9::AIDACP634&3.0.CO;2-1

Tulving, E. (1985). Memory and Consciousness. *Canadian Psychology/Psychologie Canadienne*, *26*, 1. doi: 10.1037/h0080017

Tulving, E. (1989). Memory: Performance, Knowledge, and Experience. *European Journal of Cognitive Psychology, 1*, 3-26. doi: 10.1080/09541448908403069

Weber, N., & Perfect, T. J. (2012). Improving eyewitness identification accuracy by screening out those who say they don't know. *Law and Human Behavior, 36*, 28-36. doi: 10.1037/h0093976

Wells, G. L. (1984). The Psychology of Lineup Identifications1. *Journal of Applied Social Psychology, 14*, 89-103. doi: 10.1111/j.1559-1816.1984.tb02223.x

Wells, G. L. (1993). What do we know about eyewitness identification? *American Psychologist*, *48*, 553-571. doi: 10.1037//0003-066x.48.5.553

Wells, G. L. (2008). Field experiments on eyewitness identification: Towards a better understanding of pitfalls and prospects. *Law and Human Behavior, 32*, 6-10. doi: 10.1007/s10979-007-9098-4

Wells, G. L. (2014). Eyewitness Identification Probative Value, Criterion Shifts, and Policy Regarding the Sequential Lineup. *Current Directions in Psychological Science*, *23*, 11-16. doi: 10.1177/0963721413504781

Wells, G. L., & Bradfield, A. L. (1999). Distortions in eyewitnesses' recollections: Can the postidentification-feedback effect be moderated? *Psychological Science, 10*, 138-144. doi: 10.1111/1467-9280.00121

Wells, G. L., Dysart, J. E., & Steblay, N. K. (2015). The flaw in Amendola and Wixted's conclusion on simultaneous versus sequential lineups. *Journal of Experimental Criminology*, *11*, 285-289. doi: 10.1007/s11292-014-9225-4

Wells, G. L., & Olson, E. A. (2002). Eyewitness identification: Information gain from incriminating and exonerating behaviors. *Journal of Experimental Psychology: Applied*, *8*, 155–167. doi: 10.1037/1076-898x.8.3.155

Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M., & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photo spreads. *Law and Human Behavior*, *22*, 603–647. doi: 10.1023/a:1025750605807

Wells, G. L., Smith, A. M., & Smalarz, L. (2015). ROC analysis of lineups obscures information that is critical for both theoretical understanding and applied purposes. *Journal of Applied Research in Memory and Cognition*, *4*, 324-328.doi: 10.1016/j.jarmac.2015.08.010

Wells, G. L., Steblay, N. K., & Dysart, J. E. (2015). Double-blind photo lineups using actual eyewitnesses: An experimental test of a sequential versus simultaneous lineup procedure. *Law and Human Behavior*, *39*, 1-14. doi: 10.1037/lhb0000096

Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review*, *117*, 1025-1054. doi: 10.1037/a0020874

Wixted, J. T., & Mickes, L. (2014). A signal-detection-based diagnostic-feature-detection model of eyewitness identification. *Psychological Review*, *121*, 262-276. doi: 10.1037/a0035940

Wixted, J. T., & Mickes, L. (2015). ROC analysis measures objective discriminability for any eyewitness identification procedure. *Journal of Applied Research in Memory and Cognition*, *4*, 329-334. doi: 10.1016/j.jarmac.2015.08.007

Wogalter, M. S., Malpass, R. S., & McQuiston, D. E. (2004). A national survey of US police on preparation and conduct of identification lineups. *Psychology, Crime and Law*, *10*, 69-82. doi: 10.1080/10683160410001641873

Wright, D.B., & London, K. (2009). Multilevel modelling: Beyond the basic applications. *British Journal of Mathematical and Statistical Psychology,* 62, 439-456. doi: 10.1348/000711008X327632

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*, 441-517. doi: 10.1006/jmla.2002.2864

Zajac, R., & Karageorge, A. (2009). The wildcard: A simple technique for improving children's target-absent lineup performance. *Applied Cognitive Psychology, 23*, 358-368. doi: 10.1002/acp.1511

CHAPTER 8

General Discussion

Overview

In this thesis I proposed a novel best-match-first dual-process analytical framework to guide investigation of eyewitnesses' memory and decision processes in simultaneous lineup procedures. I reported a series of nine experiments[1] designed to investigate two major research questions: 1) could the best-match-first dual-process framework explain patterns of response latency and identification performance in simultaneous lineup tasks? and 2) would modifications to simultaneous lineup procedures designed around framework predictions improve the memorial basis of eyewitnesses' identification responses, and thus improve identification accuracy overall? For the first eight experiments (Chapters 3-6), I used a face-recognition mini-lineup paradigm to collect substantial amounts of data from samples of participants in response to multiple stimuli. This approach allowed the evaluation of regularities in responding across individual and stimulus variability to rigorously test the novel framework and lineup tasks, and increase confidence in the generalisation of conclusions. In the final experiment, I used an eyewitness identification paradigm for three mock offenders, to test framework predictions and the efficacy of trialled lineup modifications in a more veridical task and context.

My first focus of investigation was testing predictions of the best-match-first dual-process framework for the standard simultaneous lineup task. Theory of eyewitness identification decision-making remains relatively under-developed, with sometimes limited

---

[1] As noted previously, a tenth experiment (Experiment 1, Chapter 5) was submitted for examination toward BA (Hons.) Psychology, Flinders University, 2011, and is not examinable with work submitted in this thesis for the award of PhD.

reference to established understandings from basic recognition memory research (Brewer & Wells, 2011; Clark & Gronlund, 2015; Gronlund, Mickes, Wixted, & Clark, 2015; Lane & Meissner, 2008). More firmly theoretically-grounded research questions are critical to direct future investigation of lineup procedures that will be informative about underlying memory and decision processes. In particular, theory is needed to guide advances in our understanding of the comparative reliability and usefulness of both currently available and innovative identification test formats and procedure, under diverse witnessing and test conditions.

Identification tasks are complex, real-world recognition judgments that require eyewitnesses to discriminate a relatively unfamiliar offender from a lineup of other similar-looking and unfamiliar faces, and to accurately discern the absence of the offender when the lineup has been composed for an innocent suspect. The extension of knowledge developed with simple laboratory tests to tasks with this greater complexity throws up questions not previously considered in basic theory. Thus, findings feed back to potentially expand and refine basic theory. The work related to this first major focus of research in this project, contributes to the advancement of both of these ongoing endeavours. First, I found clear evidence in support of the two theoretical foundations of the best-match-first dual-process framework: i) a best-match-first account of simultaneous lineup decision-making (e.g., Wells, 1984, 1993; see also Clark, 2003, 2008); and, ii) extension of a continuous dual-process account of the use of global familiarity and specific recollection in recognition memory (e.g., Wixted, 2007; Wixted & Mickes, 2010) to identification lineup tasks to explain patterns of identification performance (see also, Gronlund, 2005; Meissner, Tredoux, Parker, & Maclin, 2005; Palmer, Brewer, McKinnon & Weber, 2010; Mickes, 2015). And, second, I demonstrated that the use of familiarity and recollection in response to a complex test of recognition will be influenced by i) the specific decisions people are asked to make about the stimuli presented; ii) the temporal sequence of these decisions.

My second major focus of investigation was to investigate and develop methods that would effectively disrupt eyewitnesses' error-prone decision-making processes for simultaneous lineups, and promote better use of available memory. This research concentrated on eyewitness metacognition and task structure as two key elements of simultaneous lineup procedure that play a part in encouraging sub-optimal recognition. I targeted these mechanisms, respectively, by providing mock eyewitnesses with metacognitive instructions, and requiring them to follow a non-standard, modified decision structure to make identification judgments about simultaneous lineups.

Research with other recognition tasks (e.g., Odegard & Lampinen, 2006), including face recognition (e.g., Young, Hugenberg, Bernstein, & Sacco, 2012), suggests that a critical issue in understanding eyewitnesses' identification responses is the effect of their metacognitive predispositions to task performance on the (conscious or unconscious) adoption of a decision strategy. Investigation of the effects of biased instructions (Clark, 2005; Malpass & Devine, 1981) and administrator bias (Greathouse & Kovera, 2009) to increase eyewitness choosing from a lineup, have established the importance of eyewitnesses' metacognitive understandings for recognition decision-making. An eyewitness' metacognitive approach operates within the constraints of the procedure they follow. Thus, the interaction of metacognitive approach and task structure is likely to affect the accessibility and activation of potentially useful memory processes (e.g., Alban & Kelley, 2012; Cook, Marsh, & Hicks, 2005; Jacoby, Kelley, & McElree, 1999; Johnson, Hashtroudi, & Lindsay, 1993). The experiments reported here provided evidence about the effects of task structure, and the alignment of metacognitive understandings with task demands, on the accessibility of both memory and decision processes to metacognitive control in lineup tasks. Moreover, results were informative for the future development of more basic theory of metacognitive monitoring and control in recognition. For example, theoretical development is needed to delineate the role of early selection (i.e., at retrieval) and late correction (i.e., as part of post-retrieval monitoring) processes (Jacoby et al., 1999), and their accessibility to conscious control, in difficult recognition tasks involving episodic memory and face recognition (e.g., Blank & Launay, 2014; Johnson et al., 1993). Here, I discuss major findings of the experiments reported in this thesis, and consider their most important theoretical implications for continuing research, as well as applied implications for the collection and evaluation of eyewitness identification evidence in practice.

A Typically Ordered Sequence of Component Lineup Decisions

In the early experimental work reported in Chapter 3, I aimed to test the basic propositions of the best-match-first dual-process framework and evaluate its potential usefulness for analysis of simultaneous lineup decision-making. To first test the order of component recognition judgments proposed to contribute to an overall identification decision, I evaluated direct evidence from patterns of response latency. General ideas about a likely best-match-first order of decision processes for simultaneous lineups, in which relative match to memory for the offender can be assessed comparatively between lineup members, have been assumed in a great deal of identification research since Wells (1984, 1993) first proposed a distinction between relative and absolute judgments. However, there has been little further elaboration of these ideas (cf., Charman & Wells, 2014; Pozzulo & Lindsay, 1999). Clark's (2003, 2008) formalisation of Wells' relative judgment theory in the WITNESS computational model was the important exception. In addition, there have been few tests, and no reliable evidence, of the order or time-course of component recognition judgments assumed by best-match-first conceptions of simultaneous lineup decision-making (cf. Pozzulo, Crescini, & Lemieux, 2008; Weber & Varga, 2012). Thus, the results of the response latency analyses I report here provide an important first step in the development of dynamic models of identification decisions that account for decision-making in real time.

In WITNESS, Clark (2003) proposed that when a lineup member represents a good-enough match to memory for the offender to be selected as best-match, the eyewitness will either positively identify the best-match as the offender, or provide a *don't know* response. However, I proposed that the outcome of an eyewitness' second binary offender-presence judgment about a best-match candidate (i.e., lineup rejection, or positive identification) would crucially depend on their use of underlying familiarity and recollection. Specifically, a neglect of recollection will raise the risk of familiarity-based false identification of an innocent suspect with a strong resemblance to the culprit. Conversely, appropriately weighted use of familiarity and recollection will allow eyewitnesses to correctly reject this innocent suspect as the culprit, following their familiarity-based selection as the best-match. Therefore, I adapted Clark's model of best-match-first decision processes to include the possibility of post-best-match lineup rejection, and further specify the relative time-course of positive identifications and lineup rejections following best-match-selection. This adaptation thus allowed me to test not only the best-match-first idea of the order of lineup decisions, but also dual-process predictions for eyewitnesses' use of familiarity and recollection in these tasks.

In Chapter 3 (Experiment 1), I compared patterns of first-decision response latency for a standard simultaneous lineup procedure and a two-step best-match-first modified procedure in which eyewitnesses made the two component judgments separately: best-match selection, then offender-presence. Evidence supported predictions of the best-match-first dual-process framework for the order of component judgments in standard and modified simultaneous lineup procedures. Specifically, I found that eyewitnesses made best-match judgments faster than standard identification decisions when a positive identification would be made. For lineup rejections, best-match judgments were likely to be made more slowly than standard identification decisions. This finding is consistent with eyewitnesses in the standard condition having rejected substantial numbers of target-absent mini-lineups relatively quickly due to the lack of a plausible best-match candidate. For these lineups, from which no lineup member would reach criterion for best-match selection, forced best-match responses in the modified best-match-first condition were made after an initial lineup rejection judgment. These responses thus took longer to provide than when eyewitnesses in the standard condition were simply able to provide their initial lineup rejection as their response. Thus, overall findings were consistent with framework predictions that eyewitnesses were likely to first choose a best-match, if at least one lineup member was a plausible-enough match to memory for the offender, before making a final judgment about whether the best-match was the offender.

I further compared first decision response latency for standard and best-match-first procedures with a modified two-step presence-first procedure in which component decisions were separated, and made in reverse order (i.e., a binary offender-presence judgment, before offender selection, if present). Participants in the presence-first procedure did not make their initial offender-presence judgments for positive identifications more slowly than best-match-first participants' initial best-match judgments. Thus, I found no evidence that presence-first participants adopted a best-match-first approach to arrive at the offender-presence judgment. That is, participants did not inevitably adopt a best-match-first approach to decision-making. This finding clearly demonstrated the important general point that eyewitnesses do not spontaneously adopt a single approach to simultaneous lineup decision-making irrespective of the response requirements of the task. Rather, congruent with evidence of variability in decision-making strategies in a range of basic (Gallo, 2010; Malmberg, 2008; Odegard & Lampinen, 2006; Yonelinas, 2002) and applied (e.g., Blank & Launay, 2014; Johnson et al., 1993) recognition tasks, lineup decision processes depended on task context and purpose. Response latency results also thus provide the first direct evidence for any conceptualisation of the time-course of ordered recognition judgments with best-match-first or other decision-making approaches in simultaneous lineups.

In more basic theoretical terms, these results are an important first step toward the development of more formally specified, dynamic models of recognition decision-making for tasks that include the simultaneous presentation of an array of multiple stimuli. Models that provide a finer-grained account of the sequence of component decision-making in response to a stimulus array would have diverse practical applications. For lineup tasks, dynamic modelling would usefully inform innovative task design by enabling investigation of potentially more nuanced effects of decision type, sequence, and timing on eyewitnesses' use of available memory (e.g., Alban & Kelley, 2012). By extension, results should also encourage more detailed inquiry into the component judgments made within overall identification decisions in sequential and other alternative lineup formats (e.g., Brewer, Weber, Wootton, & Lindsay, 2012; Sauer, Brewer, & Weber, 2008), and their effects on underlying memory processes. More generally, these findings highlight the usefulness of modelling recognition response latency to generate testable predictions about decision-making in applied tasks with greater complexity than those on which simpler recognition models have been based (e.g., Ratcliff & Rouder, 1998; Ratcliff & Starns, 2009; Voskuilen, & Ratcliff, 2016). Further research following this approach is vital to help bridge the basic-applied gap in recognition memory research and increasing understanding of how basic memory processes unfold in real time in everyday contexts.

Best-Match-First Decision-Making Neglects Recollection

The second basic proposition of the best-match-first dual-process framework is that continuous dual-process models of recognition memory (e.g., Wixted, 2007; Wixted & Mickes, 2010) could explain patterns of identification responses in simultaneous lineup tasks (cf. Gronlund, 2005; Meissner et al., 2005). Most specifically, that best-match-first decision-making would encourage an over-reliance on familiarity, with a corresponding neglect of recollection. This neglect would leave eyewitnesses susceptible to mistaken identification of an innocent suspect with a strong resemblance to the real culprit. I initially tested this idea by comparing standard procedures with procedures modified to disrupt best-match-first decision processes and familiarity-based responding (Experiments 1 & 2, Chapter 3). I gained further evidence from similar comparisons in several later experiments (Chapter 4 and 6), and from experimental dissociation of the contributions of familiarity and recollection to standard (Chapters 5 & 6) and modified (Chapter 6) simultaneous procedures. The experiment reported in Chapter 7 provided a test of this proposition in a more realistic identification paradigm.

I found greater discriminability for a presence-first procedure, in which component recognition judgments were made in reverse order, than for a standard simultaneous procedure, when tested with novice mock eyewitnesses (Experiment 2, Chapter 3). This finding supported framework predictions of suboptimal use of recollection in the standard task. Greater discriminability was substantially underwritten by a lower likelihood of false identification, consistent with better use of recall-to-reject (Rotello & Heit, 2000; cf. recollection rejection, Brainerd, Reyna, Wright & Mojardin, 2003; Reyna & Brainerd, 1995). Though, other post-retrieval monitoring (e.g., Koriat & Goldsmith, 1996; Johnson et al., 1993; Lane, Roussel, Starns, Villa, & Alonzo, 2008) or memory editing mechanisms (e.g., the distinctiveness heuristic, Dodson & Schacter, 2002; Gallo, Bell, Beier & Schacter, 2006; Schacter, Israel, & Racine, 1999) might also have contributed to enhanced use of recollective information.

Comparisons of a standard simultaneous procedure with a novel presence-recollection lineup that combined the presence-first decision structure and metacognitive instructions about the diagnostic phenomenology of accurate memories, also demonstrated greater discriminability for the modified procedure (Experiment 3, Chapter 4). Again, greater accuracy was driven by the lower likelihood of false identification characteristic of greater use of recollection (Malmberg, 2008; Yonelinas, 2002) via recall-to-reject (Rotello & Heit, 2000). However, orthogonal manipulations of objectively-available familiarity and recollection revealed that participants did make some use of recollection, when available, for standard simultaneous lineups (Chapter 5, Experiments 1 & 2, Chapter 6). But, when recollection was not available, participants were willing to make familiarity-based responses, despite the higher likelihood of false identification. When the separate contributions of familiarity and recollection were evaluated, participants following the presence-recollection procedure showed more considered weighting of available familiarity and recollection than those using the standard procedure (Chapter 6). However, the detection of differences in identification response patterns between the two procedures was limited by a restricted range of potential responses to target-absent lineups in this experiment. Specifically, the scope for improvement in false identification rates with greater availability of recollection was more limited for the presence-recollection lineup than for the standard procedure, as target-absent performance was already better, and close to ceiling.

Overall, results demonstrated that while eyewitnesses made use of recollection in standard lineup procedures, they tended to neglect its full potential for both avoiding false identification, and strengthening evidence for the accurate identification of guilty suspects. Herein, these investigations revealed clear evidence of scope for improvement in eyewitnesses' use of available memory. Indeed, results served to support my identification of two prime locations for procedural interventions aimed at enhancing eyewitnesses' use of recollection. Corresponding to these locations, the combination of two specific interventions was found to be effective in the face-recognition mini-lineup paradigm: a presence-first decision structure, and metacognitive instructions.

Decision Structure and Metacognitive Alignment in a Novel Presence-Recollection Lineup

A great deal of eyewitness identification research has focused on investigating ways to reduce the likelihood of false identification without also reducing the likelihood of accurate identifications of guilty suspects. Therefore, the findings outlined above are important in two ways. First, they provide evidence for the basic proposition that best-match-first decision-making encourages error-prone familiarity-based responding. But second, they also provide evidence for the efficacy of two specific procedural interventions in standard simultaneous tasks. And for the conditions under which these interventions promoted greater use of recollection to improve discriminability with negligible reduction in accurate identifications. This is a rare and important finding, given that lineup modifications shown to reduce the risk of false identification generally do so with an accompanying reduction in hits (Clark, 2012; cf., Brewer et al., 2012; Perfect & Weber, 2012; Weber & Perfect, 2012; Zajac & Karageorge, 2009). The presence-recollection lineup has potential practical application, though extensive testing under diverse conditions is needed before a confident assessment of its usefulness can be made. But beyond the presence-recollection lineup itself, its effectiveness highlights the importance of attending to the role of decision structure (i.e., the recognition responses eyewitnesses are asked to provide) and eyewitness metacognition, and to the mechanisms by which they might be accessible and responsive to cognitive control. These factors are central to understanding patterns of identification responses under varying conditions, and in the consideration of procedural modifications for tests that might better diagnose suspect guilt or innocence than those now used by police.

The development of these lineup modifications that might reliably improve eyewitnesses' use of recollection in simultaneous procedures formed the second major research focus in work I conducted for this project. The observation of a presence-first advantage to discriminability for naïve participants (i.e., without previous experience with identification lineup-type tasks, Experiment 2, Chapter 3) was not replicated with non-naïve participants (Experiments 1 & 2, Chapter 4). While alternative explanations are possible, the preponderance of evidence suggests that this difference resulted from the greater salience for non-naïve participants of expectations about lineups (e.g., Clark, 2005; Malpass & Devine, 1981). That is, preconceived expectations about lineups were reinforced and made more salient for non-naïve participants than novices, by their previous experience with standard lineup tasks in the laboratory.

The effect of expectations on retrieval and monitoring has been demonstrated in various recognition tasks (e.g., Dodson & Schacter, 2002; Johnson et al., 1993; Lampinen, Odegard, & Bullington, 2003). I argued that, when salient, the common expectations about lineup tasks that predisposed eyewitnesses to best-match-first decision-making, also offset the efficacy of the modified presence-first decision structure to disrupt best-match-first processes. Therefore, I sought to achieve the discriminability advantage originally observed for the presence-first procedure by alternative means. Metacognitive instructions have been shown to effectively encourage participants to make better use of recollection in similar recognition tasks in which the familiarity of unstudied stimuli is likely to be misleading (e.g., Schooler, Gerhard, & Loftus, 1986; Starns, Lane, Alonzo & Roussel, 2007), but had never previously been tested with an identification task. Therefore, I tested instructions designed to align task expectations more closely with the requirements of the task (e.g., Gallo, Roberts, & Seamon, 1997; McCabe & Smith, 2002; McDermott & Roediger, 1998) and the known characteristics of memory processes (e.g., Lane, Roussel, Villa, & Morita, 2007; Lane et al., 2008; Mather, Henkel, & Johnson, 1997).

Results from the four experiments reported in Chapter 4 provided evidence for several important conclusions. First, metacognitive instructions about the diagnostic phenomenology of accurate memories (i.e., those featuring accurate retrieval of recollection) effectively improved discriminability when provided with the presence-first decision structure designed to disrupt a best-match-first decision sequence (Experiment 3). This result represents the first demonstration of enhanced use of recollection in a lineup task following the provision of metacognitive information. Further, the finding provides evidence to support the idea that eyewitnesses' metacognitive orientation, or inadequate knowledge of relevant memory and decision processes, contributed to identification errors in the standard procedure. This result thus contributes to the broader body of research showing that metacognitive shortcomings contribute to recognition errors in multiple paradigms (e.g., Blank & Launay, 2014; Odegard & Lampinen, 2006; Lane et al., 2007). Moreover, as noted, previous demonstrations of improved performance in eyewitness identification tasks with the use of instructions or other procedural modifications (e.g., blind administration; sequential presentation; increased lineup fairness) have been associated with more conservative choosing, but not significant improvement in discriminability itself (e.g., Clark, 2012). Therefore, it is particularly notable that the modified presence-recollection lineup, combining phenomenological instructions with the presence-first decision structure, produced greater discriminability than a standard simultaneous procedure, with negligible effect on hits.

Second, my results suggest that not all accurate information that increased metacognitive understanding of the lineup task was equally effective in improving participants' use of recollection. Indeed, the provision of some kinds of task-related information might even be counterproductive. I tested instructions that included task-related information about commonly observed errors in identification tasks and the risks of false identification, as well as the phenomenological elements that featured in the presence-recollection lineup, with the presence-first decision structure (Experiments 1 & 2, Chapter 4). While patterns of identification responses were in the direction of improved discriminability, these effects were not significant, and response bias was more conservative than for the standard procedure. In addition, when instructions featuring task-related elements were provided with a standard procedure, choosing was also more conservative, with no tendency toward greater discriminability (Experiment 2, Chapter 4). One explanation for these results is that task-related elements tapped into best-match-first processes to some extent. For example, by highlighting the need to avoid false identification of the most familiar lineup member, the focus remained on first finding the best-match. This focus might have prevented eyewitnesses from directing limited attentional resources toward effortful recollection, and thus from making effective use of the phenomenological information provided. However, controlled investigation of the separate effects of these instructions elements is needed to identify the likely mechanisms by which different types of metacognitive information can have differential effects on decision-making, and eyewitnesses' use of available memory. By extension, more refined categorisation (e.g., Blank & Launay, 2014) and controlled investigation of elements of warnings and instructions in other recognition tasks is likely to increase understanding of the mechanisms underlying observed variability in their effect.

Third, greater discriminability was only observed with phenomenological instructions when they were provided with a presence-first decision structure. I found no evidence of greater use of recollection when phenomenological instructions were provided with either standard (Experiment 3, Chapter 4) or modified best-match-first procedures (Experiment 4, Chapter 4). These results suggest that the over-reliance on familiarity characteristic of best-match-first decision-making is difficult to disrupt and override in favour of better use of recollection. When best-match-first processes were facilitated by task structure, participants were not able to make as effective use of phenomenological information as when the task structure did not cue familiarity-based responding. This finding has implications for attempts to enhance metacognitive knowledge and awareness in other recognition tasks. For example, changes to the questions participants are asked about previously-experienced stimuli, or the structure of the task (e.g., enforcing delayed responding, Malmberg & Xu, 2007), might increase the efficacy of metacognitive instructions to enhance recollection retrieval and monitoring.

Fourth, results of experiments reported in Chapter 4 did not allow specification of the retrieval or post-retrieval memory editing mechanisms responsible for the greater discriminability observed for the presence-recollection procedure. However, greater discriminability was driven by lower rates of false identification, consistent with the use of recall-to-reject (Rotello & Heit, 2000). This represents a potential limitation for application of the presence-recollection procedure, as recall-to-reject is not useful when limited recollectible detail has been encoded, or retained. This is frequently likely to be the case for eyewitness memory for a culprit (Brewer & Palmer, 2010). In such cases, improved identification accuracy will rely on eyewitnesses' ability to weight familiarity-based information appropriately, in the absence of recollection (e.g., the distinctiveness heuristic, Schacter et al., 1999, cf. S. A. Guerin, Robbins, Gilmore, & Schacter, 2012).

Evidence from studies of prospective judgments of learning (JOLs) in recognition (Sacher, Taconnat, Souchay, & Isingrini, 2009; Souchay, Moulin, Clarys, Taconnat, & Isingrini, 2007) and eyewitness memory (Perfect & Hollins, 1996, 1999), have shown that participants' JOLs are poorer predictors of accuracy in the absence than the presence of recollection. That is, participants do not adequately consider the absence of recollection when making these judgments, resulting in over-confidence. If the presence-recollection procedure enhanced eyewitnesses' use of post-retrieval monitoring mechanisms (e.g., Koriat & Goldsmith, 1996; Schacter et al., 1999) so as to improve their consideration of a relative absence of recollection, identification performance might be enhanced in the absence of recollection. If not, its usefulness will be seriously constrained. My test of the presence-recollection lineup in an identification paradigm (Chapter 7) provided no evidence of improved monitoring in this way. When viewed alongside the efficacy of the presence-recollection lineup in the mini-lineup paradigm, this finding highlights the need to identify the specific mechanisms underlying the effect of the modified procedure with mini-lineups, to properly evaluate its potential translation to more realistic test contexts.

One reason that it is particularly important to clarify an answer to this question, is that while eyewitnesses could experience strong familiarity in the absence of recollection (e.g., Ingram, Mickes & Wixted, 2012; Mickes, Wais, & Wixted, 2009), task context is likely to affect this subjective experience (Bodner & Lindsay, 2003; Tousignant, Bodner, & Arnold, 2015). Eyewitnesses could potentially misperceive familiarity-based memories as being based on recollection. While subjective experience broadly corresponds to these underlying processes, it does not map directly onto them (Selmeczy & Dobbins, 2014; Tulving, 1985, 1989). Meta-analyses and reviews of the effects of warnings and instructions in specific paradigms suggest that just as multiple mechanisms are likely to contribute to recognition errors (e.g., Clark & Godfrey, 2009; Gallo, 2010; Odegard & Lampinen, 2006), there are likely to be multiple mechanisms potentially accessible to reduce susceptibility to these errors (e.g., Blank & Launay, 2014; Hunt & Smith, 2014; Jacoby et al., 1999). However, as their efficacy is likely to vary between tasks, contexts, and even individuals (e.g., Alban & Kelley, 2002; Gray & Gallo, 2015), further evidence of the specific retrieval or post-retrieval mechanisms implicated in the effectiveness of the presence-recollection lineup is needed.

Eyewitness Identification and Metacognitive Control in Context

In practice, implementation of the presence-recollection procedure in real policing contexts would be straightforward. However, extensive testing across witnessing conditions is needed before the procedure should be trialled in this way. In Chapter 7, comparison of the presence-recollection procedure with standard simultaneous and sequential procedures in an identification paradigm revealed no evidence of significantly better discriminability. The greater salience and immediacy of eyewitness predispositions to best-match-first decision-making might have negated the effectiveness of the procedure in this more veridical format. Alternatively, mock witnesses might not have encoded or retained sufficient recollectible detail to make use of recall-to-reject or monitoring mechanisms that relied on the retrieval of at least some baseline recollection. Replication of this result in an identification paradigm with different stimuli and witnessing conditions will be important, but this result also highlights the potential effect of extrinsic factors, such as administrator or co-witness bias, in counter-cueing best-match-first decision-making and rendering the presence-recollection procedure ineffective. If the effects of metacognitive enhancement and structural cues to memory processes underlying the observed presence-recollection advantage with mini-lineups is not robust enough to withstand such influences, the procedure will be of limited usefulness.

Summary

This project has provided clear evidence that simultaneous lineup decision-making is characterised by a best-match-first approach that encourages familiarity-based responding, with suboptimal use of potentially beneficial recollection. Failing to consider recollected detail that would also strengthen accurate positive identification decisions, leaves eyewitnesses vulnerable to falsely identifying misleadingly familiar innocent suspects who closely resemble the culprit. These results indicate important areas for future research into the underlying memory and decision processes that shape patterns of identification responses in simultaneous and other lineup procedures. In particular, these results highlighted the impact of decision structure (within a specific lineup format), and metacognitive orientation and knowledge, on eyewitness use of familiarity and recollection, and identified these factors as focal points for procedural reform.

Investigation of the combined efficacy of presence-first decision structure and metacognitive instructions in the presence-recollection lineup to enhance eyewitness use of recollection yielded results that were consistent with predictions of the best-match-first dual-process account of simultaneous lineup decision-making. Findings from basic recognition memory research into methods for improving participants' use of available memory were successfully extended to the simultaneous lineup task (e.g., Blank & Launay, 2014; Gallo, 2010; Johnson et al., 1993; Odegard & Lampinen, 2006). Further investigation of the mechanisms underpinning improved use of recollection will be important to establish limits to the effectiveness of interventions in identification tasks, and refine their implementation. In addition, I demonstrated that the usefulness of providing participants with accurate metacognitive information before test depended on its content. Specifically, efficacy might depend on how the content of instructions relates to participants' pre-existing understandings of the task, or independently triggers potentially error-prone decision processes. Results clearly showed that while neither a presence-first structure nor phenomenological instructions alone could produce greater discriminability than for a standard lineup, their dual implementation improved the memorial basis of identification responses. Thus, decision structure and metacognition interact in their effects on eyewitness identification performance, and its potential improvement. This finding identifies an important focus for future theoretical investigation of the mechanisms that affect the retrieval and monitoring of familiarity and recollection in recognition decision-making, and their operation in eyewitness identification and other complex real-world recognition tasks.

References

Alban, M. W., & Kelley, C. M. (2012). Variations in constrained retrieval. *Memory & Cognition*, *40*, 681-692. doi: 10.3758/s13421-012-0185-5

Blank, H., & Launay, C. (2014). How to protect eyewitness memory against the misinformation effect: A meta-analysis of post-warning studies. *Journal of Applied Research in Memory and Cognition*, *3*, 77-88. doi: 10.1016/j.jarmac.2014.03.005

Bodner, G. E., & Lindsay, D. S. (2003). Remembering and knowing in context. *Journal of Memory and Language*, *48*, 563-580. doi: 10.1016/s0749-596x(02)00502-8

Brainerd, C. J., Reyna, V. F., Wright, R., & Mojardin, A. H. (2003). Recollection rejection: False-memory editing in children and adults. *Psychological Review*, *110*, 762-784. doi: 10.1037/0033-295x.110.4.762

Brewer, N., & Palmer, M. A. (2010). Eyewitness identification tests. *Legal and Criminological Psychology*, *15*, 77-96. doi: 10.1348/135532509x414765

Brewer, N., Weber, N., Wootton, D., & Lindsay, D. S. (2012). Identifying the bad guy in a lineup using confidence judgments under deadline pressure. *Psychological Science*, *23*, 1208-1214. doi: 10.1177/0956797612441217

Brewer, N., & Wells, G. L. (2011). Eyewitness identification. *Current Directions in Psychological Science*, *20*, 24-27. doi: 10.1177/0963721410389169

Charman, S., & Wells, G. L. (2014). Applied Lineup Theory. *Handbook of Eyewitness Psychology 2 Volume Set*, 219-254. doi: 10.4324/9781315805535

Clark, S. E. (2003). A memory and decision model for eyewitness identification. *Applied Cognitive Psychology*, *17*, 629-654. doi: 10.1002/acp.891

Clark, S. E. (2005). A re-examination of the effects of biased lineup instructions in eyewitness identification. *Law and Human Behavior*, *29*, 575-604. doi: 10.1007/s10979-005-7121-1

Clark, S. E. (2008). The importance (necessity) of computational modelling for eyewitness identification research. *Applied Cognitive Psychology*, *22*, 803-813. doi: 10.1002/acp.1484

Clark, S. E. (2012). Costs and benefits of eyewitness identification reform psychological science and public policy. *Perspectives on Psychological Science*, *7*, 238-259. doi: 10.1177/1745691612439584

Clark, S. E., & Godfrey, R. D. (2009). Eyewitness identification evidence and innocence risk. *Psychonomic Bulletin & Review*, *16*, 22-42. doi: 10.3758/pbr.16.1.22

Clark, S. E., & Gronlund, S. D. (2015). The development and evolution of theory in eyewitness identification. In J. G. W., Raajmakers, A. H. Criss, R. L. Goldstone, R, M. Nosofsky, & M. Steyvers (Eds.) *Cognitive Modeling in Perception and Memory: A Festschrift for Richard M. Shiffrin*, Psychology Press: New York. doi: 10.4324/9781315885582

Cook, G. I., Marsh, R. L., & Hicks, J. L. (2005). Revisiting the role of recollection in item versus forced-choice recognition memory. *Psychonomic Bulletin & Review*, *12*, 720-725. doi: 10.3758/bf03196763

Dodson, C. S., & Schacter, D. L. (2002). When false recognition meets metacognition: The distinctiveness heuristic. *Journal of Memory and Language*, *46*, 782-803. doi: 10.1037/e501882009-058

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition*, *38*, 833-848. doi: 10.3758/mc.38.7.833

Gallo, D. A., Bell, D. M., Beier, J. S., & Schacter, D. L. (2006). Two types of recollection-based monitoring in younger and older adults: Recall-to-reject and the distinctiveness heuristic. *Memory*, *14*, 730-741. doi: 10.1080/09658210600648506

Gallo, D. A., Roberts, M. J., & Seamon, J. G. (1997). Remembering words not presented in lists: Can we avoid creating false memories? *Psychonomic Bulletin & Review*, *4*, 271-276. doi: 10.3758/bf03209405

Gray, S. J., & Gallo, D. A. (2015). Disregarding familiarity during recollection attempts: Content-specific recapitulation as a retrieval orientation strategy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 134-147. doi: 10.1037/a0038363

Greathouse, S. M., & Kovera, M. B. (2009). Instruction bias and lineup presentation moderate the effects of administrator knowledge on eyewitness identification. *Law and Human Behavior, 33*, 70-82. doi: 10.1007/s10979-008-9136-x

Gronlund, S. D. (2005). Sequential lineup advantage: Contributions of distinctiveness and recollection. *Applied Cognitive Psychology,* 19, 23-37. doi: 10.1002/acp.1047

Gronlund, S. D., Mickes, L., Wixted, J. T., & Clark, S. E. (2015). Conducting an eyewitness lineup: How the research got it wrong. In B. H. Ross (Ed.), The psychology of learning and motivation (Volume 63, pp. 1– 43). Waltham, MA: Academic Press. doi: 10.1016/bs.plm.2015.03003

Guerin, S. A., Robbins, C. A., Gilmore, A. W., & Schacter, D. L. (2012). Retrieval failure contributes to gist-based false recognition. *Journal of Memory and Language*, *66*, 68-78. doi: 10.1016/j.jml.2011.07.002

Hunt, R. R., & Smith, R. E. (2014). How distinctive processing enhances hits and reduces false alarms. *Journal of Memory and Language*, *75*, 45-57. doi: 10.1016/j.jml.2014.04.007

Ingram, K. M., Mickes, L., & Wixted, J. T. (2012). Recollection can be weak and familiarity can be strong. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 325-339. doi: 10.1037/e520592012-514

Jacoby, L. L., Kelley, C. M., & McElree, B. D. (1999). The role of cognitive control: Early selection versus late correction. In. S. Chaiken & Y. Trope (Eds.), Dual-process theories in social psychology (pp. 383– 400). New York, NY: Guilford Press.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3-28. doi: 10.1037//0033-2909.114.1.3

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*, 490-517. doi: 10.1037/0033-295x.103.3.490

Lampinen, J. M., Odegard, T. N., & Bullington, J. L. (2003). Qualities of memories for performed and imagined actions. *Applied Cognitive Psychology*, *17*, 881-893. doi: 10.1002/acp.916

Lane, S. M., & Meissner, C. A. (2008). A 'middle road' approach to bridging the basic–applied divide in eyewitness identification research. *Applied Cognitive Psychology*, *22,* 779-787. doi: 10.1002/acp.1482

Lane, S. M., Roussel, C. C., Starns, J. J., Villa, D., & Alonzo, J. D. (2008). Providing information about diagnostic features at retrieval reduces false recognition. *Memory*, *16*, 836–851. http://doi: 10.1080/09658210802337347

Lane, S. M., Roussel, C. C., Villa, D., & Morita, S. K. (2007). Features and feedback: Enhancing metamnemonic knowledge at retrieval reduces source-monitoring errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33,* 1131-1142. doi: 10.1037/0278-7393.33.6.1131

Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology*, *57*, 335-384. doi: 10.1016/j.cogpsych.2008.02.004

Malmberg. K. J., & Xu, J. (2007). On the flexibility and the fallibility of associative memory. *Memory & Cognition, 35,* 545-556. doi: 10.3758/bf03193293

Malpass, R. S., & Devine, P. G. (1981). Eyewitness identification: Lineup instructions and the absence of the offender. *Journal of Applied Psychology*, *66,* 482. doi: 10.1037/0021-9010.66.4.482

Mather, M., Henkel, L. A., & Johnson, M. K. (1997). Evaluating characteristics of false memories: Remember/know judgments and memory characteristics questionnaire compared. *Memory & Cognition*, *25*(6), 826-837. doi: 10.3758/bf03211327

McCabe, D. P., & Smith, A. D. (2002). The effect of warnings on false memories in young and older adults. *Memory & Cognition*, *30*, 1065-1077. doi: 10.3758/bf03194324

McDermott, K. B., & Roediger, H. L. (1998). Attempting to avoid illusory memories: Robust false recognition of associates persists under conditions of explicit warnings and immediate testing. *Journal of Memory and Language*, *39*, 508-520. doi: 10.1006/jmla.1998.2582

Meissner, C. A., Tredoux, C. G., Parker, J. F., MacLin, O. H. (2005). Eyewitness decisions in simultaneous and sequential lineups: A dual-process signal detection theory analysis. *Memory & Cognition*, 33, 783-792. doi: 10.3758/bf03193074

Mickes, L., Wais, P. E., & Wixted, J. T. (2009). Recollection is a continuous process: Implications for dual-process theories of recognition memory. *Psychological Science, 20*, 509-515. doi: 10.1111/j.1467-9280.2009.02324.x

Mickes, L. (2015). Receiver operating characteristic analysis and confidence-accuracy characteristic analysis in investigations of system variables and estimator variables that affect eyewitness memory. *Journal of Applied Research in Memory and Cognition, 4,* 83-102. doi: 10.1016/j.jarmac.2015.01.003

Odegard, T. N., & Lampinen, J. M. (2006). Memory editing: Knowledge, criteria, and alignment. *Memory, 14*, 777-787. doi: 10.1080/09658210600648589

Palmer, M. A., Brewer, N., McKinnon, A. C., & Weber, N. (2010). Phenomenological reports diagnose accuracy of eyewitness identification decisions. *Acta Psychologica*, *133*, 137-145. doi: 10.1016/j.actpsy.2009.11.002

Perfect, T. J., & Hollins, T. S. (1996). Predictive feeling of knowing judgements and postdictive confidence judgements in eyewitness memory and general knowledge. *Applied Cognitive Psychology*, *10*, 371-382. doi: 10.1002/(sici)1099-0720(199610)10:5<371::aid-acp389>3.0.co;2-o

Perfect, T. J., & Hollins, T. S. (1999). Feeling-of-knowing judgments do not predict subsequent recognition performance for eyewitness memory. *Journal of Experimental Psychology: Applied*, *5*, 250-264. doi: 10.1037/1076-898x.5.3.250

Perfect, T. J., & Weber, N. (2012). How should witnesses regulate the accuracy of their identification decisions: One step forward, two steps back? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1810-1818. doi: 10.1037/a0028461

Pozzulo, J. D., Crescini, C., & Lemieux, J. M. (2008). Are accurate witnesses more likely to make absolute judgments? *International Journal of Law and Psychiatry*, *31*, 495-501. doi: 10.1016/j.ijlp.2008.09.006

Pozzulo, J. D., & Lindsay, R. C. L. (1999). Elimination lineups: An improved identification procedure for child eyewitnesses. *Journal of Applied Psychology*, *84*, 167-176. doi: 10.1037/0021-9010.84.2.167

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*, 347-356. doi: 10.1111/1467-9280.00067

Ratcliff, R., & Starns, J. J. (2009). Modeling confidence and response time in recognition memory. *Psychological Review*, *116*, 59. doi: 10.1037/a0014086

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and individual Differences, 7*(1), 1-75. doi: 10.1016/1041-6080(95)90031-4

Rotello, C. M., & Heit, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory & Cognition*, *28*, 907-922. doi: 10.3758/bf03209339

Sacher, M., Taconnat, L., Souchay, C., & Isingrini, M. (2009). Divided attention at encoding: Effect on feeling-of-knowing. *Consciousness and Cognition*, *18*, 754-761. doi: 10.1016/j.concog.2009.04.001

Sauer, J. D., Brewer, N., & Weber, N. (2008). Multiple confidence estimates as indices of eyewitness memory. *Journal of Experimental Psychology: General,* 137, 528-547. doi: 10.1037/a0012712

Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language*, *40*, 1-24. doi: 10.1006/jmla.1998.2611

Schooler, J. W., Gerhard, D., & Loftus, E. F. (1986). Qualities of the unreal. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 171-181. doi: 10.1037/0278-7393.12.2.171

Selmeczy, D., & Dobbins, I. G. (2014). Relating the content and confidence of recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 66-85. doi: 10.1037/a0034059

Souchay, C., Moulin, C. J., Clarys, D., Taconnat, L., & Isingrini, M. (2007). Diminished episodic memory awareness in older adults: Evidence from feeling-of-knowing and recollection. *Consciousness and Cognition*, *16*, 769-784. doi: 10.1016/j.concog.2006.11.002

Starns, J. J., Lane, S. M., Alonzo, J. D., & Roussel, C. C. (2007). Metamnemonic control over the discriminability of memory evidence: A signal detection analysis of warning effects in the associative list paradigm. *Journal of Memory and Language*, *56*, 592-607. doi: 10.1016/j.jml.2006.08.013

Tousignant, C., Bodner, G. E., & Arnold, M. M. (2015). Effects of context on recollection and familiarity experiences are task dependent. *Consciousness and cognition, 33*, 78-89. doi: 10.1016/j.concog.2014.11.011

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie Canadienne*, *26*, 1. doi: 10.1037/h0080017

Tulving, E. (1989). Memory: Performance, knowledge, and experience. *European Journal of Cognitive Psychology, 1*, 3-26. doi: 10.1080/09541448908403069

Voskuilen, C., & Ratcliff, R. (2016). Modeling confidence and response time in associative recognition. *Journal of Memory and Language*, *86*, 60-96. doi: 10.1016/j.jml.2015.09.006

Weber, N., & Perfect, T. J. (2012). Improving eyewitness identification accuracy by screening out those who say they don't know. *Law and Human Behavior, 36*, 28-36. doi: 10.1037/h0093976

Weber, N., & Varga, M. (2012). Can a modified lineup procedure improve the usefulness of confidence? *Journal of Applied Research in Memory and Cognition*, *1*, 152-157. doi: 10.1016/j.jarmac.2012.06.007

Wells, G. L. (1984). The Psychology of Lineup Identifications1. *Journal of Applied Social Psychology*, *14*, 89-103. doi: 10.1111/j.1559-1816.1984.tb02223.x

Wells, G. L. (1993). What do we know about eyewitness identification? *American Psychologist*, *48,* 553-571. doi: 10.1037//0003-066x.48.5.553

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review, 114*, 152-176. doi: 10.1037/0033-295x.114.1.152

Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review*, *117*, 1025-1054. doi: 10.1037/a0020874

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*, 441-517. doi: 10.1006/jmla.2002.2864

Young, S. G., Hugenberg, K., Bernstein, M. J., & Sacco, D. F. (2012). Perception and motivation in face recognition: A critical review of theories of the cross-race effect. *Personality and Social Psychology Review*, *16*, 116-142. doi: 10.1177/1088868311418987

Zajac, R., & Karageorge, A. (2009). The wildcard: A simple technique for improving children's target-absent lineup performance. *Applied Cognitive Psychology*, *23*, 358-368. doi: 10.1002/acp.1511

Appendix 1

Chapter 3 Supplemental Materials

Here, we report the random effects standard deviations for mixed-effects models reported in the manuscript (see Tables 1-8). Standard deviations provide an estimate of the variability of the random effect.

Table 10

Random Effects Standard Deviations (SDs) for Mixed-Effects Model Predicting First-decision Response Latency (lnrl) by Lineup Type (Standard; Best-match-first; Presence-first) in Experiments 1 & 2

	+*SD*					
	#Standard vs Best-match-first		#Standard vs Presence-first		#Best-match-first vs Presence-first	
Random Effect	Expt 1	Expt 2	Expt 1	Expt 2	Expt 1	Expt 2
Participant (P)	0.32	0.34	0.29	0.31	0.31	0.32
TP|P	0.11	0.11	0.12	0.09	0.06	0.06
Stimulus (Stim)		0.07		0.05	0.10	0.10
TP|Stim		0.13		0.06	0.10	0.11

Note. # indicates the reference lineup condition, + *SD* reflects variability due to the random effect. Expt 1 models that included the standard lineup procedure did not include random effects by stimulus due to a recording error during the experiment.

Table 11

Random Effects Standard Deviations (SDs) for Mixed-Effects Models Predicting Decision Type (Lineup Rejection; Positive Identification) by Lineup Type (Standard; Best-match-first; Presence-first) in Experiments 1 & 2

		+*SD*					
		#Standard vs Best-match-first		#Standard vs Presence-first		#Best-match-first vs Presence-first	
Filler Identifications	Random Effect	Expt 1	Expt 2	Expt 1	Expt 2	Expt 1	Expt 2
Included	Participant	0.86	0.87	0.50	0.82	0.93	0.59
	TP|Participant	1.05	0.52	0.74	0.78	0.45	0.22
	Stim		0.75		0.65	0.61	0.26
	TP|Stim		1.09		1.02	1.25	0.38
	Lineup|Stim		0.60		0.29	0.39	0.45
	Lineup*TP|Stim		0.97		0.44	0.23	0.53
Excluded	Participant	0.86	0.65	0.44	0.67	0.92	0.46
	TP|Participant	1.01	0.57	0.60	0.81	0.72	0.62
	Stim		0.99		0.98	0.69	0.70
	TP|Stim		1.29		1.24	1.32	0.40
	Lineup|Stim		0.71		0.20	0.41	0.49
	Lineup*TP|Stim		1.13		0.66	0.61	

Note: # indicates the reference lineup condition. + Standard Deviation (*SD)* reflects variability due to the random effect. Expt 1 models that included the standard lineup procedure did not include random effects by stimulus due to a recording error during the experiment.

Appendix 2

Chapter 4 Supplemental Materials

Here, we report:

1. Random effects standard deviations for mixed-effects models reported in the manuscript (see Tables 1-8). Standard deviations provide an estimate of the variability of the random effect.
2. Metacognitive instructions scripts for Experiments 1-4
3. Instructions manipulation check questionnaire, Experiments 1-4.
4. Instructions manipulation check questionnaire results.

1. Random Effects for Mixed-Effects Models Reported for Experiments 1-4

Table 10

Random Effects Standard Deviations for Mixed-Effects Model Predicting Decision Type (Lineup Rejection; Positive Identification) by Lineup Type in Experiment 1

	Lineup Comparison					
	#Standard vs PF		#Standard vs PFI		#PF vs PFI	
Random Effect	Filler Identifications Included	Filler Identifications Excluded	Filler Identifications Included	Filler Identifications Excluded	Filler Identifications Included	Filler Identifications Excluded
Participant	1.01	1.00	1.01	1.00	0.87	0.87
TP|Participant	1.08	1.27	1.10	1.26	1.05	1.25
Stim	0.29	0.27	0.32	0.31	0.18	0.16
TP|Stim	0.17	0.00	0.26	0.32	0.49	0.38
Lineup|Stim	0.46	0.03	0.57	0.58	0.40	0.42
Lineup*TP|Stim	0.62		0.60	0.69	0.72	0.62

Note: # indicates the reference lineup condition. Number of observations, fillers included (fillers excluded): Standard vs PF = 2400 (2220), *n* = 60 (60); Standard vs PFI = 2520 (2384), *n* = 63 (63); PF vs PFI = 2520 (2344), *n* = 63 (63).

Table 11

Random Effects Standard Deviations for Mixed-Effects Models Predicting Decision Type (Lineup Rejection; Positive Identification) by Lineup Type in Experiment 2

	Lineup Comparisons							
	[#]Standard vs PF		[#]Standard vs PFI		[#]Standard vs SI		[#]PF vs PFI	
Random Effect	Filler Identifications Included	Filler Identifications Excluded	Filler Identifications Included	Filler Identifications Excluded	Filler Identifications Included	Filler Identifications Excluded	Filler Identifications Included	Filler Identifications Excluded
Participant (ID)	0.78	0.82	0.93	0.86	0.96	0.98	0.65	0.55
TP|ID	0.61	0.58	0.71	0.53	0.84	0.78	0.51	0.50
Lineup|ID					0.80	0.43	0.39	0.66
Lineup*TP|ID					0.79	0.29		0.22
Stim	0.47	0.84	0.46	0.93	0.46	0.69	0.43	0.94
TP|Stim	0.52	1.02	0.55	0.93	0.51	0.60	0.65	0.98
Lineup|Stim	0.02	0.59	0.08	0.43	0.06			
Lineup*TP|Stim	0.22		0.08	0.34	0.13			

Note: # indicates the reference lineup condition. Number of observations, fillers included (fillers excluded): Standard vs PF = 2560 (2086), *n* = 64 (64); Standard vs PFI = 2560 (2131), *n* = 64 (64); Standard vs SI = 2560 (1870), *n* = 32 (32); PF vs PFI = 2560 (2179), *n* = 32 (32).

Table 12

Random Effects Standard Deviations for Mixed Effects Model Predicting Decision Type (Lineup Rejection; Positive Identification) by Lineup Type in Experiment 3

	Lineup Comparisons			
	#Standard vs PR		#Standard vs SI	
Random Effect	Filler Identifications Included	Filler Identifications Excluded	Filler Identifications Included	Filler Identifications Excluded
Participant	0.74	0.74	0.72	0.70
TP|Participant	0.59	0.74	0.48	0.65
Stim	0.43	0.42	0.42	0.00
TP|Stim	0.77	0.92	0.74	0.68
Lineup|Stim	0.05		0.14	0.43
Lineup*TP|Stim			0.19	0.25

Note: # indicates the reference lineup condition. Number of observations, fillers included (fillers excluded): Standard vs PR = 5120 (4727), *n* = 64 (64); Standard vs SI = 5120 (4669), *n* = 64 (64).

Table 13

Random Effects Standard Deviations for Mixed Effects Model Predicting Decision Type (Lineup Rejection; Positive Identification) by Lineup Type in Experiment 4

	Lineup Comparisons					
	#Standard vs BMFI		#BMF vs BMFI		#Standard vs BMF	
Random Effect	Filler Identifications Included	Filler Identifications Excluded	Filler Identifications Included	Filler Identifications Excluded	Filler Identifications Included	Filler Identifications Excluded
ID	0.98	0.94	0.98	0.96	0.94	0.99
TP|ID	1.01	1.41	0.80	1.12	0.61	1.06
Stim	0.09	0.35	0.29	0.99	0.10	0.35
TP|Stim	0.53	0.98	0.47	0.80	0.38	1.14
Lineup|Stim	0.29	0.10	0.45	0.79	0.47	0.86
Lineup*TP|Stim	0.33	0.09	0.66	0.63	0.85	1.17

Note: # indicates the reference lineup condition. Number of observations, fillers included (fillers excluded): Standard vs BMF = 2220 (1809), *n* = 78 (78); Standard vs BMFI = 2260 (1831), *n* = 79 (79); BMF vs BMFI = 2680 (2204), *n* = 67 (67).

2. Metacognitive Instructions in Experiments 1-4

Experiment 1: Task-Related and Phenomenological Instructions

Before you begin this first series of recognition tests, we would like to give you some more information about how to do the task. Each group of photographs in this recognition test is made up of faces that match the same description as the studied face in gender, face shape, hair colour and style, eye colour, skin colour and complexion, and age. However, some faces in the group will look more similar to the studied face than others. Sometimes the studied face is present in the group of photographs and sometimes it is not.

Because some of the faces in the group are very similar to the studied face, people sometimes mistakenly decide that the studied face is present when all four photographs show faces that have not been studied. When people focus on the similarity between test faces and the studied face, they pay less attention to useful evidence that the test faces are not the studied face.

Research has shown that accurate decisions about the studied face being present are likely to be accompanied by more vivid and clear memories of studying the face than when an unstudied test face is inaccurately and mistakenly recognised. These can include memories of how the studied face looked, memories of how the [name or occupation] and the studied face are related, and other thoughts and feelings that were experienced when the face was studied.

We would like you to use this information to accurately decide when a studied face is present in the test group of photographs and when it is not. We would like you to avoid mistakenly and inaccurately recognising a test face that is similar to the studied face, but was not studied. When you think that the face you studied with the [name or occupation] is present, think about how clear, vivid, and detailed your memory is. Can you remember the thoughts and feelings you experienced when you studied the face?

If you have a clear and vivid memory of the studied face, you should indicate that the studied face is present. If you think the test photograph is similar to your memory for the studied face, but your memory is not clear, vivid or detailed, you should respond ‘Not Present’.

Experiment 2: Task-Related and Phenomenological Instructions

Before you begin this first series of recognition tests, we would like to give you some more information about how memory works in recognition tasks. Each group of photographs in this recognition test is made up of faces that match the same description as the studied face in gender, face shape, hair colour and style, eye colour, skin colour and complexion, and age. Sometimes the studied face is present in the group of photographs and sometimes it is not.

Research has shown that accurate decisions about the studied face being present are likely to be accompanied by more vivid and clear memories of studying the face than when an unstudied test face is inaccurately and mistakenly recognised. These can include memories of how the studied face looked, memories of how the [name or occupation] and the studied face are related, and other thoughts and feelings that were experienced when the face was studied.

We would like you to use this information to accurately decide when a studied face is present in the test group of photographs and when it is not. When you think that the face you studied with the [name or occupation] is present, think about how clear, vivid, and detailed your memory is. Can you remember the thoughts and feelings you experienced when you studied the face?

If you have a clear and vivid memory of the studied face, you should indicate that the studied face is present. If your memory is not clear, vivid, or detailed, you should respond ‘Not Present’.

Experiments 3 & 4: Phenomenological Instructions

Before you begin this first series of recognition tests, we would like to give you some more information about how memory works in recognition tasks. Each group of photographs in this recognition test is made up of faces that match the same description as the studied face in gender, face shape, hair colour and style, eye colour, skin colour and complexion, and age. Sometimes the studied face is present in the group of photographs and sometimes it is not.

Research has shown that accurate decisions about the studied face being present are likely to be accompanied by more vivid and clear memories of studying the face than when an unstudied test face is inaccurately and mistakenly recognised. These can include memories of how the studied face looked, memories of how the [name or occupation] and the studied face are related, and other thoughts and feelings that were experienced when the face was studied.

We would like you to use this information to accurately decide when a studied face is present in the test group of photographs and when it is not. When you think that the face you studied with the [name or occupation] is present, think about how clear, vivid, and detailed your memory is. Can you remember the thoughts and feelings you experienced when you studied the face?

3. Instructions Manipulation Check Questionnaire, Experiments 1-4

Instructions

You will now be asked to complete a series of questions about how you used your memory when recognising faces. Each question will be presented on a separate screen. For some questions, you will be asked to type your responses in a text box provided at the bottom of the screen. After you have typed your answer for these questions, click the 'Next' button to proceed to the following screen. For other questions, you will be asked to respond on a rating scale with several response options. Your task will be to click on the option that best describes your response to the question. Once you have clicked on a response option, a 'Next' button will appear. Click the 'Next' button to proceed to the following screen. If you have any questions, please ask the experimenter now. Click 'Next' to proceed to the questions.

Open-ended Questions

Item 1. When one of the four photographs in the test group seemed very similar to your memory of a studied face, what do you think it was important to consider when deciding whether the studied face was present? Please type your answers into the text box presented below. You may type as much as you like. When you have finished typing your answer, click the 'Next' button to proceed to the following question.

Item 2. What is likely to be different between the experience of accurately recognising a studied face and the experience of mistakenly recognising a test photograph that is similar to the studied face? Please type your answers into the text box presented below. You may type as much as you like. When you have finished typing your answer, click the 'Next' button to proceed to the following question.

Item 3. Please describe how you think the test groups of 4 photographs were made up by the experimenters. That is, how do you think the photographs were selected for each group? Please type your answers into the text box presented below. You may type as much as you like. When you have typed every answer, click the 'Next' button to proceed to the following question

Ranking Question

Item 1. Please use the drop-down lists to select 6 features in any order that you think were used to match the 4 faces selected by the experimenters to be in each test group.

Features. Gender; ethnicity; face shape; facial expression; similarity to studied face; hair colour; hair style; eye colour; skin colour; facial hair; nose shape; age; distinctiveness; same description.

Ratings Questions

Instructions. You will now be presented with a series of statements about how likely it is that different kinds of memories are correct. For these statements, you will be presented with 7 response options ranging from "Very unlikely" to "Very likely" to be correct. Your task will be to click on the option that best describes how correct a memory is likely to be when it has this characteristic. If you have any questions, please ask the experimenter now. When you are ready to proceed to the questions, click 'Next' to proceed to the questions.

Item 1. When my memory of a studied face is moderately vivid and the test photograph matches the general description of the studied face, it is___________to be correct.

Item 2. When my memory of a recognised face includes recollection of specific details, it is___________to be correct.

Item 3. When my memory of the studied face is not clear, but one photograph in the test group is much more similar to my memory than the other three, it is___________to be correct.

Item 4.

When I decide the studied face is present by reasoning about my memory for the face and the group of test photographs, I am___________to be correct.

Item 5. When I can remember what I thought or felt when I studied a face, my memory is___________to be correct.

Item 6. When I can remember what I thought about how the displayed name or occupation was connected to the studied face, my memory is___________to be correct.

Item 7. When I can remember how early or late in the study phase a recognised face was studied, my memory is___________to be correct.

Instructions. You will now be presented with a series of statements about how much you relied on different kinds of information when you were deciding whether you had accurately recognised a face or whether it had not been studied. For these statements, you will be presented with 7 response options for how much you relied on the kind of information mentioned in the statement, ranging from “Not at all” to “Always”. Your task will be to click on the option that best describes how much you relied on that kind of information. If you have any questions, please ask the experimenter now. When you are ready to proceed to the questions, click 'Next' to proceed to the questions.

Item 8. (Instructions conditions only) When deciding that a studied face was present, how much did you use the information about memory provided in the instructions?

Item 9. (Instructions conditions only) When deciding that a studied face was not present, how much did you use the information about memory provided in the instructions?

Item 10. When deciding that a studied face was present, how much did you rely on one face being much more similar to the studied face than the other three faces?

Item 11. When deciding that a studied face was not-present, how much did you rely on the overall vividness of your memory of the studied face?

Item 12. When making a decision about whether the studied face was present or not, how much did you rely on how clear your memory of the studied face was?

Item 13. When making a decision about whether the studied face was present or not, how much did you rely on how detailed your memory of the studied face was?

Item 14. When making a decision about whether the studied face was present or not, how much did you rely on remembering thoughts and feelings you had when the face was studied?

Item 15. When deciding that a familiar test face was not the studied face, how much did you pay attention to not having a detailed memory of the studied face?

Item 16. When deciding that a familiar test face had been studied, how much did you rely on reasoning about whether the studied face was likely to be present in that test group?

Item 17. When deciding whether a studied face was present in the test group or not, how much did you rely on finding the face in the test group that was most similar to your memory for the studied face?

Note. For questions with response scales, we used 7-point Likert scales to allow us to capture relatively fine-grained differences in responding (Cook, Heath, Thompson, & Thompson, 2001; Matell & Jacoby, 1972).

References

Cook C., Heath F., Thompson R. L., Thompson B. (2001). Score reliability in web- or Internet-based surveys: Unnumbered graphic rating scales versus Likert-type scales. *Educational and Psychological Measurement*, 61, 697-706. doi: 10.1177/00131640121971356

Matell, M. S., & Jacoby, J. (1972). Is there an optimal number of alternatives for Likert-scale items? Effects of testing time and scale properties. *Journal of Applied Psychology*, *56*, 506. doi: 10.1037/h0033601

4. Instructions Manipulation Check Questionnaire Results

Table 14

Descriptive Statistics and Effect Sizes for Responses to Instructions Manipulation Check Questionnaire Items by Lineup Type and Pairwise Comparison for Experiment 1

			Item Response					
			Lineup Type			Pairwise Comparison		
Item Type	Item	Statistic	Standard $n = 30$	PF $n = 30$	PFI $n = 33$	Statistic	Standard vs PFI	PF vs PFI
Open[#]	1	*M*	0.83	0.77	1.12	*d*	0.41*	0.49*
		SD	0.70	0.73	0.74			
		95% CI	0.58, 1.08	0.50, 1.02	0.87, 1.37	95% CI	0.23, 0.58	0.31, 0.67
	2	*M*	0.07	0.3	058	*d*	0.71*	0.39*
		SD	0.25	0.65	0.79			
		95% CI	-0.02, 0.16	0.07, 0.53	0.31, 0.85	95% CI	0.56. 0.85	0.21, 0.57
	3	*M*	0.87	0.77	1.55	*d*	0.40	0.42
		SD	1.22	1.14	1.75			
		95% CI	0.43, 1.31	0.36, 1.18	0.95, 2.15	95% CI	-0.21, 1.01	0.06, 0.79
Rating	1[^]	*M*	4.30	4.03	4.06	*d*	0.41*	0.03*
		SD	0.99	1.29	0.61			
		95% CI	3.95, 4.65	3.57, 4.49	3.84, 4.28	95% CI	0.21, 0.62	-0.20, 0.26
	2[^]	*M*	4.87	4.40	4.64	*d*	0.16*	0.17
		SD	0.68	1.35	1.48			
		95% CI	4.66, 5.14	3.92, 4.88	4.14, 5.14	95% CI	-0.34, 0.67	-0.18, 0.52
	3[^]	*M*	3.60	3.57	3.06	*d*	0.51*	0.44*

	SD	1.07	1.22	1.09			
	95% CI	3.22, 3.98	3.13, 4.01	2.69, 3.43	95% CI	0.14, 0.89	0.15, 0.73
4^	*M*	3.57	3.27	3.64	*d*	0.06	0.29
	SD	1.25	1.34	1.25			
	95% CI	3.12, 4.02	2.79, 3.75	3.21, 4.07	95% CI	-0.37, 0.49	-0.01, 0.61
5^	*M*	4.57	4.37	4.88	*d*	0.39*	0.48*
	SD	1.55	1.27	0.86			
	95% CI	4.02, 5.12	3.92, 4.82	4.59, 5.17	95% CI	0.10, 0.68	0.21, 0.75
6^	*M*	4.63	4.50	4.85	*d*	0.21*	0.31*
	SD	1.25	1.20	1.06			
	95% CI	4.18, 5.08	4.07, 4.93	4.49, 5.21	95% CI	-0.16, 0.58	0.03, 0.60
7^	*M*	3.43	3.37	3.36	*d*	0.06*	0.00*
	SD	1.38	1.45	1.22			
	95% CI	2.94, 3.92	2.85, 3.89	2.94, 3.78	95% CI	-0.36, 0.48	-0.33, 0.34
8+	*M*	-	-	3.03		-	-
	SD	-	-	1.21			
	95% CI	-	-	2.62, 3.44		-	-
9+	*M*	-	-	3.09		-	-
	SD	-	-	1.38			
	95% CI	-	-	2.62, 3.56		-	-
10+	*M*	3.47	3.47	3.67	*d*	0.15*	0.14*
	SD	1.48	1.50	1.43			
	95% CI	2.94, 4.00	2.93, 4.01	3.18, 4.16	95% CI	-0.35, 0.64	-0.23, 0.50
11+	*M*	3.63	3.77	3.97	*d*	0.57*	0.14*
	SD	1.22	1.17	1.21			

	95% CI	3.19, 4.07	3.35, 4.19	3.56, 4.38	95% CI	0.15, 0.99	-0.22, 0.50
12+	*M*	4.40	3.87	4.39	*d*	0.01	0.50
	SD	1.07	1.31	0.79			
	95% CI	4.02, 4.78	3.40, 4.34	4.12, 4.66	95% CI	-0.26, 0.28	0.23, 0.77
13+	*M*	4.43	4.23	4.12	*d*	0.38*	0.10*
	SD	1.07	1.17	1.17			
	95% CI	4.05, 4.81	3.81, 4.65	3.72, 4.52	95% CI	-0.13, 0.68	-0.19, 0.39
14+	*M*	3.50	3.40	3.91	*d*	0.26*	0.30*
	SD	1.76	1.79	1.63			
	95% CI	2.87, 4.13	2.76, 4.04	3.35, 4.47	95% CI	-0.30, 0.82	-0.12, 0.73
15+	*M*	3.47	2.97	3.33	*d*	0.11	0.27*
	SD	1.46	1.54	1.22			
	95% CI	2.95, 3.99	2.42, 3.52	2.91, 3.75	95% CI	-0.31, 0.53	-0.08, 0.61
16+	*M*	3.27	3.03	3.21	*d*	0.21	0.14*
	SD	1.39	1.16	1.49			
	95% CI	2.87, 3.87	2.61, 3.45	2.70, 3.72	95% CI	-0.20, 0.61	-0.20, 0.47
17+	*M*	3.90	4.07	3.36	*d*	0.38*	0.57*
	SD	1.21	1.01	1.48			
	95% CI	3.47, 4.33	3.71, 4.43	2.86, 3.86	95% CI	-0.14, 0.89	0.25, 0.88

Note. * indicates a difference in the direction predicted if participants attended to instructions. # indicates open-ended questions, mean number of instructions elements mentioned in response is reported. ^ indicates questions measured on a scale from 1 (Very unlikely) to 7 (Very likely). + indicates questions measured on a scale from 1 (Not a lot) to 7 (Always).

Table 15

Descriptive Statistics and Effect Sizes for Responses to Instructions Manipulation Check Questionnaire Items by Lineup Type and Pairwise Comparison for Experiment 2

			Item Response						
			Lineup Type					Pairwise Comparison	
Item Type	Item	Statistic	Standard *n* = 32	SI *n* = 32	PF *n* = 32	PFI *n* = 32	Statistic	Standard vs SI	PF vs PFI
Open[#]	1	*M*	0.65	1.03	1.00	0.94	*d*	0.64*	-0.10
		SD	0.49	0.71	0.71	0.56			
		95% CI	0.48, 0.82	0.78, 1.28	0.76, 1.24	0.75, 1.13	95% CI	0.13, 1.13	-0.58, 0.39
	2	*M*	0.31	0.22	0.39	0.53	*d*	-0.19	0.22*
		SD	0.47	0.49	0.50	0.72			
		95% CI	0.15, 0.47	0.05, 0.39	0.22, 0.56	0.28, 0.78	95% CI	-0.68, 0.30	-0.26, 0.70
	3	*M*	2.31	2.63	2.33	3.13	*d*	0.18*	0.45*
		SD	1.69	1.74	1.81	1.70			
		95% CI	1.72, 2.90	2.02, 3.22	1.71, 2.95	2.55, 3.71	95% CI	-0.30, 0.67	-0.04, 0.93
Rating	1[^]	*M*	3.84	3.88	3.91	4.00	*d*	0.03	0.07
		SD	0.72	1.10	1.13	1.30			
		95% CI	3.49, 4.09	3.50, 4.26	3.52, 4.30	3.56, 4.44	95% CI	-0.46, 0.52	-0.41, 0.56
	2[^]	*M*	4.53	4.53	4.39	4.69	*d*	0.00	0.20
		SD	1.24	1.08	1.66	1.23			
		95% CI	4.10, 4.96	4.16, 4.90	3.82, 4.96	4.27, 5.11	95% CI	-0.49, 0.49	-0.28, 0.68
	3[^]	*M*	3.16	2.97	3.30	3.53	*d*	-0.20*	0.24
		SD	0.85	1.00	0.95	0.95			

	95% CI	2.87, 3.46	2.62, 3.32	3.01, 3.65	3.21, 3.85	95% CI	-0.69, 0.29	-0.25, 0.72
4^	*M*	3.19	3.31	3.33	3.47	*d*	0.11	0.14
	SD	1.28	0.93	0.95	1.02			
	95% CI	2.75, 3.63	2.99, 3.63	3.01, 3.65	3.12, 3.82	95% CI	-0.38, 0.6	-0.35, 0.62
5^	*M*	3.97	4.03	4.64	4.19	*d*	0.04*	-0.40
	SD	1.45	1.38	0.96	1.49			
	95% CI	3.47, 4.47	3.55, 4.51	4.31, 4.97	3.68, 4.70	95% CI	-0.45, 0.53	-0.84, 0.13
6^	*M*	4.16	3.94	4.21	4.44	*d*	-0.15	0.19*
	SD	1.53	1.41	1.17	1.19			
	95% CI	3.63, 4.69	3.45, 4.43	3.81, 4.61	4.03, 4.85	95% CI	-0.64, 0.34	-0.29, 0.67
7^	*M*	2.88	2.94	2.91	2.72	*d*	0.05	-0.10*
	SD	1.45	1.24	1.42	1.42			
	95% CI	2.38, 3.38	2.51, 3.37	2.43, 3.39	2.24, 3.20	95% CI	-0.44, 0.54	-0.62, 0.35
8+	*M*	-	3.00	-	3.47		-	-
	SD	-	1.50	-	1.19			
	95% CI	-	2.48, 3.52	-	3.06, 3.88		-	-
9+	*M*	-	3.00	-	3.41		-	-
	SD	-	1.52	-	1.13			
	95% CI		2.47, 3.53	-	3.02, 3.80		-	-
10+	*M*	3.28	3.69	3.88	4.28	*d*	0.34*	0.35*
	SD	1.22	1.18	1.27	1.02			
	95% CI	2.86, 3.70	3.28, 4.10	3.45, 4.31	3.92, 4.63	95% CI	-0.16, 0.83	-0.14, 0.83
11+	*M*	3.38	4.22	3.88	3.91	*d*	0.68*	0.02*
	SD	1.26	1.21	0.99	1.28			
	95% CI	2.94, 3.82	3.80, 4.63	3.54, 4.22	3.47, 4.35	95% CI	0.17, 1.18	-0.46, 0.51

12+	*M*	4.03	4.09	4.21	4.16	*d*	0.05	-0.1*
	SD	1.23	1.20	0.78	0.95			
	95% CI	3.60, 4.46	3.67, 4.51	3.94, 4.48	3.84, 4.48	95% CI	-0.44, 0.54	-0.55, 0.42
13+	*M*	3.72	3.66	3.79	4.03	*d*	-0.05*	0.19
	SD	1.20	1.38	1.41	1.15			
	95% CI	3.30, 4.14	3.18, 4.14	3.31, 4.27	3.91, 4.69	95% CI	-0.54, 0.44	-0.30, 0.67
14+	*M*	3.53	3.94	3.09	3.53	*d*	0.25*	0.23*
	SD	1.65	1.56	1.97	1.80			
	95% CI	2.96, 4.10	3.40, 4.48	2.42, 3.76	2.92, 4.14	95% CI	-0.24, 0.74	-0.25, 0.71
15+	*M*	2.94	3.00	3.18	3.59	*d*	0.04*	0.32*
	SD	1.44	1.57	1.36	1.19			
	95% CI	2.44, 3.44	2.46, 3.54	2.72, 3.64	3.18, 4.00	95% CI	-0.45, 0.53	-0.17, 0.80
16+	*M*	2.94	2.84	3.30	3.31	*d*	-0.07	0.01*
	SD	1.44	1.11	1.36	1.12			
	95% CI	2.44, 3.44	2.46, 3.22	2.84, 3.76	3.31, 4.07	95% CI	-0.56, 0.42	-0.47, 0.49
17+	*M*	4.00	3.81	4.24	3.69	*d*	-0.15*	-0.5*
	SD	1.14	1.31	1.20	1.12			
	95% CI	3.61, 4.40	3.36, 4.26	3.83, 4.65	3.31, 4.07	95% CI	-0.64, 0.34	-0.96, 0.02

Note. * indicates a difference in the direction predicted if participants attended to instructions. # indicates open-ended questions, mean number of instructions elements mentioned in response is reported. ^ indicates questions measured on a scale from 1 (Very unlikely) to 7 (Very likely). + indicates questions measured on a scale from 1 (Not a lot) to 7 (Always).

Table 16

Descriptive Statistics and Effect Sizes for Responses to Instructions Manipulation Check Questionnaire Items by Lineup Type and Pairwise Comparison for Experiment 3

			Item Response					
			Lineup Type			Pairwise Comparison		
Item Type	Item	Statistic	Standard n = 32	SI n = 32	PFI n = 32	Statistic	Standard vs SI	Standard vs PFI
Open[#]	1	*M*	0.92	0.95	0.89	*d*	0.05*	-0.06
		SD	0.48	0.62	0.50			
		95% CI	0.72, 1.11	0.76, 1.14	0.70, 1.09	95% CI	-0.44, 0.54	-0.59, 0.47
	2	*M*	0.12	0.19	0.50	*d*	0.20*	0.89*
		SD	0.33	0.40	0.51			
		95% CI	-0.01, 0.25	0.07, 0.31	0.30, 0.70	95% CI	-0.29, 0.69	0.32, 1.44
	3	*M*	1.88	2.00	1.82	*d*	0.06*	-0.04
		SD	1.75	1.90	1.44			
		95% CI	1.17, 2.59	1.43, 2.58	1.26, 2.38	95% CI	-0.40, 0.55	-0.57, 0.49
Rating	1[^]	*M*	3.54	3.86	4.14	*d*	0.29*	0.58*
		SD	1.07	1.09	1.01			
		95% CI	3.11, 3.97	3.53, 4.19	3.75, 4.53	95% CI	-0.20. 0.78	0.03, 1.12
	2[^]	*M*	4.58	4.95	5.32	*d*	0.39*	0.82*
		SD	1.03	0.91	0.77			
		95% CI	4.15, 5.00	4.69, 5.23	5.02, 5.62	95% CI	-0.10, 0.88	0.26, 1.37
	3[^]	*M*	3.23	3.31	3.25	*d*	0.07*	0.02*
		SD	1.42	0.95	1.00			

	95% CI	2.66, 3.80	3.02, 3.60	2.86, 3.64	95% CI	-0.42, 0.55	-0.52, 0.55
4^	*M*	3.42	3.17	4.18	*d*	-0.25*	0.71
	SD	1.10	1.01	1.02			
	95% CI	2.98, 3.86	2.87, 3.48	3.78, 4.58	95% CI	-0.73, 0.25	0.15, 1.25
5^	*M*	4.35	4.10	4.68	*d*	-0.19	0.30*
	SD	1.29	1.28	0.90			
	95% CI	3.83, 4.87	3.71, 4.49	4.33, 5.03	95% CI	-0.68, 0.30	-0.24, 0.83
6^	*M*	4.08	4.26	4.68	*d*	0.14*	0.49*
	SD	1.41	1.31	1.02			
	95% CI	3.51, 4.65	3.86, 4.66	4.28, 5.08	95% CI	-0.35, 0.63	-0.06, 1.03
7^	*M*	2.62	2.86	3.43	*d*	0.18	0.64
	SD	1.10	1.51	1.43			
	95% CI	2.18, 3.06	2.40, 3.32	2.87, 3.99	95% CI	-0.32, 0.67	0.08, 1.17
8+	*M*	-	2.57	2.86		-	-
	SD	-	1.25	1.30			
	95% CI	-	2.19, 2.95	2.35, 3.37		-	-
9+	*M*	-	2.45	2.57		-	-
	SD	-	1.27	1.45			
	95% CI	-	2.07, 2.83	2.01, 3.13		-	-
10+	*M*	3.54	3.81	3.96	*d*	0.21*	0.35*
	SD	1.27	1.25	1.17			
	95% CI	3.03, 4.05	3.43, 4.19	3.51, 4.42	95% CI	-0.28, 0.70	-0.19, 0.88
11+	*M*	3.35	3.86	3.71	*d*	0.42*	0.26*
	SD	1.62	0.90	1.18			
	95% CI	2.70, 4.00	3.59, 4.13	3.25, 4.17	95% CI	-0.08, 0.91	-0.28, 0.79

12+	*M*	3.54	3.79	4.00	*d*	0.19	0.35
	SD	1.53	1.16	1.09			
	95% CI	2.92, 4.16	3.44, 4.14	3.58, 4.42	95% CI	-0.30, 0.68	-0.19, 0.88
13+	*M*	3.85	3.50	4.36	*d*	-0.28*	0.47
	SD	1.22	1.23	0.95			
	95% CI	3.36, 4.34	3.13, 3.87	3.99, 4.73	95% CI	-0.77, 0.21	-0.08, 1.00
14+	*M*	2.81	3.62	3.96	*d*	0.46*	0.67*
	SD	1.94	1.64	1.53			
	95% CI	2.03, 3.59	3.12, 4.12	3.36, 4.56	95% CI	-0.04, 0.95	0.11, 1.20
15+	*M*	2.73	3.10	3.11	*d*	0.25*	0.24*
	SD	1.59	1.41	1.50			
	95% CI	2.09, 3.37	2.67, 3.53	2.53, 3.69	95% CI	-0.25, 0.73	-0.30, 0.78
16+	*M*	3.00	3.17	3.36	*d*	0.13*	0.25*
	SD	1.50	1.21	1.42			
	95% CI	2.39, 3.61	2.80, 3.54	2.81, 3.91	95% CI	-0.37, 0.61	-0.29, 0.78
17+	*M*	3.88	3.95	4.18	*d*	0.05	0.25
	SD	1.42	1.25	0.94			
	95% CI	3.24, 4.38	3.57, 4.33	3.81, 4.55	95% CI	-0.44, 0.54	-0.29, 0.78

Note. * indicates a difference in the direction predicted if participants attended to instructions. # indicates open-ended questions, mean number of instructions elements mentioned in response is reported. ^ indicates questions measured on a scale from 1 (Very unlikely) to 7 (Very likely). + indicates questions measured on a scale from 1 (Not a lot) to 7 (Always).

Table 17

Descriptive Statistics and Effect Sizes for Responses to Instructions Manipulation Check Questionnaire Items by Lineup Type and Pairwise Comparison for Experiment 4

			Item Response					
			Lineup Type				Pairwise Comparison	
Item Type	Item	Statistic	Standard n = 45	BMF n = 33	BMFI n = 34	Statistic	Standard vs BMFI	BMF vs BMFI
Open[#]	1	*M*	1.00	1.04	1.05	*d*	0.33*	0.04*
		SD	0.00	0.20	0.22			
		95% CI	1.00, 1.00	0.97, 1.11	0.98, 1.12	95% CI	-0.12, 0.78	-0.44, 0.52
	2	*M*	1.00	1.09	1.00	*d*	0.00	-0.43
		SD	0.00	0.30	0.00			
		95% CI	1.00, 1.00	0.99, 1.19	1.00, 1.00	95% CI		-0.91, 0.06
	3	*M*	2.87	2.30	3.15	*d*	0.22*	0.65*
		SD	1.36	1.47	1.13			
		95% CI	2.47, 3.27	1.80, 2.80	2.77, 3.53	95% CI	-0.23, 0.67	0.15, 1.13
Rating[^]	1[^]	*M*	3.51	3.67	4.00	*d*	0.41	0.344
		SD	1.27	0.85	1.07			
		95% CI	3.14, 3.88	3.38, 3.96	3.64, 4.36	95% CI	-0.04, 0.86	-0.14, 0.82
	2[^]	*M*	4.51	4.33	4.59	*d*	0.08	0.27
		SD	1.19	1.02	0.89			
		95% CI	4.19, 4.83	3.98, 4.68	4.29, 4.89	95% CI	-0.37, 0.52	-0.22, 0.74
	3[^]	*M*	2.89	3.24	3.12	*d*	0.19*	-0.11
		SD	1.28	1.20	1.07			

	95% CI	2.52, 3.26	2.83, 3.65	2.76, 3.48	95% CI	-0.26, 0.64	-0.59, 0.37
4^	*M*	3.16	3.30	3.41	*d*	0.24	0.10
	SD	1.04	1.13	1.13			
	95% CI	2.86, 3.46	2.91, 3.69	3.03, 3.79	95% CI	-0.21, 0.68	-0.38, 0.57
5^	*M*	3.40	4.06	3.88	*d*	0.36*	-0.14
	SD	1.42	1.32	1.25			
	95% CI	2.99, 3.82	3.61, 4.51	3.46, 4.30	95% CI	-0.1, 0.8	-0.62, 0.34
6^	*M*	3.82	4.03	3.91	*d*	0.07*	-0.10
	SD	1.34	1.21	1.22			
	95% CI	3.43, 4.21	3.62, 4.44	3.50, 4.32	95% CI	-0.38, 0.54	-0.58, 0.38
7^	*M*	2.91	3.21	3.03	*d*	0.10	-0.15*
	SD	1.18	1.17	1.19			
	95% CI	2.57, 3.24	2.81, 3.61	2.63, 3.43	95% CI	-0.35, 0.54	-0.63, 0.33
8+	*M*	-	-	3.09		-	-
	SD	-	-	1.03			
	95% CI	-	-	2.74, 3.44		-	-
9+	*M*	-	-	3.03		-	-
	SD	-	-	1.06			
	95% CI			2.67, 3.39		-	-
10+	*M*	3.89	3.48	3.59	*d*	-0.23	0.07*
	SD	1.13	1.50	1.50			
	95% CI	3.56, 4.22	2.97, 3.99	3.09, 4.09	95% CI	-0.68, 0.22	-0.41, 0.55
11+	*M*	4.16	3.85	3.38	*d*	-0.69	-0.40
	SD	1.02	1.09	1.26			
	95% CI	3.86, 4.46	3.48, 4.22	2.96, 3.80	95% CI	-1.14, -0.22	-0.87, 0.09

12+	*M*	4.40	3.97	3.85	*d*	-0.55*	-0.11*
	SD	0.99	1.19	0.99			
	95% CI	4.11, 4.69	3.56, 4.38	3.52, 4.18	95% CI	-1.00, -0.10	-0.59, 0.37
13+	*M*	4.13	3.85	4.15	*d*	0.01	0.25
	SD	1.01	1.20	1.21			
	95% CI	3.84, 4.43	3.44, 4.26	3.74, 4.56	95% CI	-0.43, 0.46	-0.24, 0.73
14+	*M*	2.71	3.39	3.26	*d*	0.29*	-0.08
	SD	1.91	1.37	1.83			
	95% CI	2.15, 3.27	2.92, 3.86	2.65, 3.88	95% CI	-0.16, 0.74	-0.56, 0.40
15+	*M*	3.56	3.30	3.24	*d*	-0.23	-0.05
	SD	1.42	1.29	1.42			
	95% CI	3.15, 3.98	2.86, 3.74	2.76, 3.72	95% CI	0.67, 0.22	-0.53, 0.43
16+	*M*	2.87	2.30	3.15	*d*	0.22*	0.65*
	SD	1.36	1.47	1.13			
	95% CI	2.47, 3.27	1.80, 2.80	2.77, 3.53	95% CI	-0.23, 0.67	0.15, 1.13
17+	*M*	3.47	2.94	3.06	*d*	-0.35*	0.09
	SD	1.22	1.46	1.07			
	95% CI	3.11, 3.83	2.44, 3.44	2.70, 3.42	95% CI	-0.80, 0.10	-0.39, 0.57

Note. * indicates a difference in the direction predicted if participants attended to instructions. # indicates open-ended questions, mean number of instructions elements mentioned in response is reported. ^ indicates questions measured on a scale from 1 (Very unlikely) to 7 (Very likely). + indicates questions measured on a scale from 1 (Not a lot) to 7 (Always).

Appendix 3

Chapter 5 Supplemental Materials

Here, we report the random effects standard deviations for mixed-effects models reported in the manuscript (see Tables 1-4). Standard deviations provide an estimate of the variability of the random effect.

Table 6

Random Effects Standard Deviations for Mixed-Effects Model Predicting Decision Type (Lineup Rejection: Positive Identification) by Repetitions (Two; Six) and Attention (Divided; Full) in Experiments 1 & 2

	Expt 1		Expt 2	
Random Effect	Filler IDs Included	Filler IDs Excluded	Filler IDs Included	Filler IDs Excluded
ID	1.03	0.99	1.18	0.95
TP|ID	1.00	1.22	0.66	0.79
Reps|ID	-	-	0.16	-
Stim	0.39	0.70	0.29	0.60
TP|Stim	0.71	0.90	0.29	0.69
Attn|Stim	0.51	0.34	0.25	0.35
Attn*TP|Stim	0.49	0.72	0.29	-

Note: Number of observations, filler identification included (excluded): *Experiment 1* = 3680 (2946), *n* = 92 (92); *Experiment 2* = 4000 (3172), *n* = 100 (100).

Table 7

Random Effects Standard Deviations for Mixed Effects Model Predicting Decision Accuracy by Attention, Decision Type and Confidence in Experiment 1

	Filler IDs	
Random Effect	Included	Excluded
ID	1.08	1.80
TP|ID	1.27	2.00
Stim	0.39	2.04
TP|Stim	0.46	2.15
Attn|Stim	0.59	1.68
Attn*TP|Stim	0.77	1.71

Note: Number of observations, filler identifications included (excluded): 3680 (2946), n = 92 (92).

Appendix 4

Chapter 6 Supplemental Materials

Here, we report:

1. Text of phenomenological instructions provided in the presence-recollection lineup
2. Manipulation checks for the effects of repetitions on target familiarity, and divided attention on available recollection.
3. Random effects standard deviations for mixed-effects models reported in the manuscript (see Tables 1&2). Standard deviations provide an estimate of the variability of the random effect.

Text of Phenomenological Instructions Provided in the Presence-Recollection Lineup

Research has shown that accurate decisions about the studied face being present are likely to be accompanied by more vivid and clear memories of studying the face than when an unstudied test face is inaccurately and mistakenly recognised. These can include memories of how the studied face looked, memories of how the name/occupation and the studied face are related, and other thoughts and feelings that were experienced when the face was studied.

We would like you to use this information to accurately decide when a studied face is present in the test group of photographs and when it is not. When you think that the face you studied with the [cue word] is present, think about how clear, vivid, and detailed your memory is. Can you remember the thoughts and feelings you experienced when you studied the face?

Manipulation Checks

Separate models predicting decision-type for each lineup procedure were used to check the manipulation of familiarity (see Tables 1 & 2, Figure 1). To dissociate the effects of target recollection and target familiarity on identification performance, we aimed to evaluate the effect of recollection at two levels of familiarity. If the repetitions manipulation created two levels of familiarity, we expected more identifications of targets presented six times at study than those presented twice. Replicating previous results (Guerin et al., 2016c), both target presence and the target presence × repetitions interaction term were significant predictors of choosing in the standard lineup model. For target-present lineups, identifications were more likely with six repetitions than two (Figure 1, Panels A, B, E & F). We checked whether the attention manipulation had confounded the familiarity manipulation, and found no significant difference in the effect of repetitions on target familiarity between attention conditions for target-present lineups (indexed by the non-significant interaction terms for attention with repetitions). Similarly, for the presence-recollection lineup condition, target presence significantly predicted choosing, and the target presence × repetitions interaction term was a meaningful (but not significant) predictor with filler identifications included (Figure 1, Panels C & D). As shown in Table 2, the coefficient CIs for the interaction term largely extend above zero. With filler identifications excluded, the effect was significant (Figure 1, Panels G & H). There was no significant difference in effects between levels of attention. Thus the repetitions manipulation produced two levels of familiarity, with no evidence of confounding by attention, for each procedure.

We expected the tone-monitoring task in the divided-attention condition to limit attention available to encode recollectible target detail, and thereby create two levels of recollection within each lineup condition. For both standard, $t = 26.42$, $df = 82$, $p < .001$ ($M = .79$, $SD = .10$) and presence-recollection procedures, $t = 30.42$, $df = 77$, $p < .001$ ($M = .81$, *SD* = .09), the proportion of tones in each test block for which participants responded correctly before the next tone onset was significantly above chance, consistent with participants having attended to the tones.

Table 4

Random Effects Standard Deviations for Mixed-Effects Models Predicting Decision Type (Lineup Rejection: Positive Identification) by Repetitions (Two; Six) and Attention (Divided; Full) separately for Standard and Presence-Recollection Lineups

	Standard		Presence-Recollection	
Random Effect	Filler IDs Included	Filler IDs Excluded	Filler IDs Included	Filler IDs Excluded
ID	0.94	0.79	0.64	0.55
TP|ID	0.85	1.03	0.93	1.09
Stim	0.52	0.92	0.64	0.96
TP|Stim	0.59	1.06	0.80	1.16
Attn|Stim	0.01	0.16	0.54	0.70
Attn*TP|Stim	-	-	0.57	0.60

Note: Number of observations, fillers included (fillers excluded): Standard = 6640 (5309), *n* = 83 (83); PR = 6240 (5281), *n* = 78 (78).

Table 5

Random Effects Standard Deviations for Mixed-Effects Model Predicting Decision Type (Lineup Rejection: Positive Identification) by Attention (Divided; Full), and Lineup Type (Standard; Presence-Recollection)

Random Effect	Filler IDs Included	Filler IDs Excluded
ID	(0.80)	(0.69)
TP|ID	(0.89)	(1.07)
Stim	(0.54)	(0.84)
TP|Stim	(0.64)	(0.96)
Attn|Stim	(0.28)	(0.22)
Attn*TP|Stim	(0.29)	NA

Note: Number of observations, fillers included (fillers excluded): Standard = 6640 (5309), *n* = 83 (83); PR = 6240 (5281), *n* = 78 (78).