Effects of lineup size and pre-lineup instructions, crime-lineup delays and culprit descriptions on the cognitive processes underlying eyewitness responses to lineups

Inaugural-Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Amelie Frederike Therre

aus Worms

Düsseldorf, Juni 2025

aus dem Institut für Experimentelle Psychologie

der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der

Mathematisch-Naturwissenschaftlichen Fakultät der

Heinrich-Heine-Universität Düsseldorf

Berichterstatter:

1. Prof. Dr. Axel Buchner

2. Prof. Dr. Raoul Bell

Tag der mündlichen Prüfung: 25.7.2025

Table of contents

Abstract

Eyewitness testimonies obtained through lineups can provide valuable evidence in criminal prosecutions. However, they also present legal professionals—including investigating officers, attorneys, judges and jurors—with a fundamental dilemma: while correct culprit identifications can support legitimate convictions and are therefore desirable, false identifications of innocent suspects may contribute to wrongful convictions and should therefore be avoided. However, despite an extensive body of research, many forensically relevant factors that influence eyewitness responses to lineups and the cognitive processes underlying them remain insufficiently understood. In this dissertation, the well-validated two-high threshold eyewitness identification model was used to examine several of these factors. In Experiments 1a and 1b, the combined effects of *lineup size* and *pre-lineup instructions* were investigated. It was demonstrated in both experiments that the probability of culprit-presence detection was higher in three-person lineups than in six-person lineups. Furthermore, the probability of guessing-based selection was lower when pre-lineup instructions suggesting a low probability of culprit presence than when neutral pre-lineup instructions not implying a low probability of culprit presence were provided. Ultimately, culprit identification rates were higher in three-person lineups combined with instructions suggesting a low probability of culprit presence compared to six-person lineups with neutral instructions, while the innocent-suspect-identification rates were statistically indistinguishable. In Experiment 2, the effect of varying delays between the crime and the associated lineup was examined. The probability of culprit-presence detection decreased progressively across four time points (no delay, one day, one week and one month), with the decline best described by a power function. Notably, the probability of guessing-based selection remained stable, indicating that eyewitnesses did not compensate for declining culprit-presence detection by increasing their tendency to guess. In Experiment 3, the effect of providing culprit descriptions prior to the lineups was investigated. Both the probability of culprit-presence detection and the probability of guessing-based selection were significantly lower when culprit descriptions were provided compared to when no descriptions were provided. Together, these findings underscore the utility of the two-high threshold eyewitness identification model for addressing applied legal questions. This dissertation contributes to the field by offering new insights into how several forensically relevant factors influence the cognitive processes underlying eyewitness responses to lineups.

Introduction

Eyewitness testimony plays a critical role in criminal investigations and prosecutions (e. g., Wells & Olson, 2003). A common method for gathering evidence for or against a suspect's guilt is the photo lineup, in which a suspect's photograph is presented alongside photographs of known-to-be-innocent persons (so-called fillers). When presented with a lineup, eyewitnesses are asked to identify the culprit if the culprit is present or to reject the lineup if the culprit is absent. While this might seem straightforward at first glance, eyewitness memory is known to be fallible (Pezdek, 2012; Rose & Beck, 2016). As of 2024, false identifications of innocent suspects made by eyewitnesses contributed to 63 % of DNA exoneration cases that were handled by the Innocence Project, a non-profit organization dedicated to exonerating wrongfully convicted individuals (Innocence Project, 2024). This highlights an inherent dilemma when interpreting the evidentiary value of eyewitness responses to lineups: while a correct identification of a culprit can support a legitimate conviction and is therefore desirable, a false identification of an innocent suspect can lead to a wrongful conviction and is therefore undesirable (Clark, 2012). Addressing this dilemma, efforts have been undertaken for decades to identify and investigate potential forensically relevant factors that influence eyewitness responses to lineups, with the ultimate goal to „[minimize] the likelihood that an innocent suspect will be (falsely) identified and [maximize] the likelihood that a guilty suspect will be (accurately) identified“ (Wells et al., 1993, p. 835).

There are numerous potential forensically relevant factors influencing eyewitness responses to lineups. Some of these relate to the environmental conditions present at the time of the crime (i. e., situational factors), while others concern characteristics of the eyewitness or the culprit (i. e., individual factors). These are entirely beyond the control of legal professionals and include factors such as the number of culprits involved in a crime (Lockamyeir et al., 2021; Tupper et al., 2018), the eyewitness's age (Erickson et al., 2016; Fitzgerald & Price, 2015), the ethnicity of both the culprit and the eyewitness (Brigham & Ready, 1985; Meissner & Brigham, 2001b) or the duration of exposure to the culprit's face during the crime (e. g., Memon et al., 2003; Palmer et al., 2013). Other forensically relevant factors (i. e., procedural factors) pertain either to the broader investigative process that precedes the lineup and serves to build evidence-based suspicion (i. e., pre-lineup investigative factors) or to the administration of the lineup itself (i. e., lineup-specific factors). Pre-lineup investigative factors include decisions and circumstances associated with the collection and assessment of evidence. While some of these investigative steps are guided by protocol and are therefore under partial control of the investigating officers, such as exposure of the eyewitness to mugshots (Deffenbacher et al., 2006), the construction of facial composites (Sporer et al., 2020; Tredoux et al., 2021) or the techniques used when interviewing eyewitnesses (Gabbert & Brown, 2015), others are shaped by external influences, such as co-witness misinformation (Levett, 2013; Zajac & Henderson, 2009). As such, these pre-lineup investigative factors may or may not fall fully within the control of legal professionals, yet they can still exert significant effects on eyewitness responses to the eventual lineup. Unlike situational, individual or pre-lineup investigative factors, lineup-specific factors are entirely under the control of legal professionals and include decisions such as whether lineups are presented sequentially—where each photo is shown individually—or simultaneously—where all photos are shown at once (Lindsay & Wells, 1985; Steblay et al., 2001), whether the procedure is conducted live or via photographs (Egan et al., 1977; Rubínová et al., 2021), which criteria are used to select filler photographs (Fitzgerald et al., 2013; McKinley & Peterson, 2022) and whether eyewitnesses are asked to report their confidence level after having given a response (Li et al., 2024). These aspects of the procedure can be systematically planned and standardized. Additionally, researchers have investigated combined effects of situational or individual factors with procedural factors. For example, specific lineup procedures have been developed for children (Karageorge & Zajac, 2011; Pozzulo & Lindsay, 1999) and older adults (Wilcock & Bull, 2010) or adapted in cross-racial cases (e. g., Evans et al., 2009) or cases of masked culprits (e. g., Guerra et al., 2022; Manley et al., 2022). In light of these extensive research efforts, numerous recommendations designed to optimize lineups have been proposed (Wells, 2001; Wells et al., 2020; Wells et al., 1998).

However, many forensically relevant factors influencing eyewitness responses to lineups remain insufficiently understood. One reason for this is that in many studies, conclusions are based on the investigation of raw eyewitness response rates to lineups (e. g., the rates of culprit-identifications or innocent-suspect-identifications). While this is a straightforward and accessible approach, these eyewitness responses can result from a variety of underlying cognitive processes, which remain insufficiently investigated. Taking into consideration the forensically relevant factor of the eyewitness's age as an example, older adults are less likely than younger adults to correctly identify a culprit in a lineup (Erickson et al., 2016; Fitzgerald & Price, 2015). However, this difference in culprit identifications between older and younger eyewitnesses could be caused by multiple underlying cognitive processes: first, it is possible that the eyewitness's ability to detect the presence of the culprit in the lineup declines with increasing age. Second, older eyewitnesses may be less sensitive to subtle differences between the culprit's photo and those of the fillers—differences that, if noticed, could make the culprit stand out and render the lineup unfair. As a result, characteristics of the culprit's appearance or photo that might lead younger eyewitnesses to make a biased suspect selection may go unnoticed by older adults. Third, it is possible that older adults are less inclined to select someone based on guessing than younger adults.

Although all of these processes lead to a decrease in correct culprit identifications, the underlying cognitive processes differ considerably. Distinguishing between these processes is critical, because if the observed decrease is due to a reduction in non-detection-based processes—such as a reduced susceptibility to unfairness in lineups or a reduced inclination to make a guessing-based selection—the decrease in culprit identifications is a desirable outcome. That is because non-detection-based processes are inherently undesirable—even when they occasionally result in correctly identifying the culprit, because in the moment of the lineup, the suspect's guilt is unknown. Therefore, these non-detection based processes might equally lead to a wrongful identification of an innocent suspect. Conversely, if the decrement in culprit identifications arises because the eyewitness's ability to detect the presence of the culprit in the lineup decreases with increasing age, this is clearly an undesirable outcome, given that culprit identifications based on culprit-presence detection are the only ones that are actually diagnostic of a suspect's guilt. This illustrates that by investigating raw response rates alone, important information is disregarded. Without taking into account the latent cognitive processes that drive these responses, the evidentiary value of the responses cannot be determined. Therefore, to thoroughly assess the effects of forensically relevant factors (and examine potential remedies if the factor increases non-detection-based processes), it is necessary to apply measurements that go beyond raw response rates.

A classic measure that goes beyond raw response rates is the *diagnosticity ratio* (Wells & Lindsay, 1980). This ratio expresses the rate of culprit identifications (numerator) relative to the rate of innocent-suspect identifications (denominator). If a forensically relevant factor is associated with a higher diagnosticity ratio, then an identification made by a single eyewitness when influenced by this factor provides stronger evidence of guilt—i. e., increases the probability that the identified suspect is actually the culprit (Wells, 2014)—compared to a suspect identification obtained under the influence of a forensically relevant factor associated with a lower diagnosticity ratio. As an alternative to the diagnosticity ratio, Mickes et al.

(2012) introduced *Receiver Operating Characteristic* (ROC) *analyses* to lineup research. Here, the relationship between culprit identifications (on the y-axis) and innocent-suspect identifications (on the x-axis) is plotted in a graph across different levels of response criteria (as assessed by subjective confidence). When connecting the values of culprit and innocent-suspect identifications at different levels of subjective confidence, an ROC curve is created. The area under the ROC curve serves as an index of how well a forensically relevant factor discriminates innocent from guilty suspects in a lineup, with larger areas indicating better discriminability.

However, both measures—diagnosticity ratio and ROC analyses—are based on only a subset of the response categories that can occur when eyewitnesses respond to lineups, namely culprit identifications and innocent-suspect identifications. In the case of the diagnosticity ratio, filler identifications are excluded entirely. ROC analyses, which are based on signal detection theory (Green & Swets, 1966), do not treat filler identifications as a separate category and instead require a reclassification of the data. Whereas the original eyewitness lineup response structure forms a 2 × 3 matrix—comprising culprit identifications, filler identifications, and incorrect rejections in culprit-present lineups, as well as innocent-suspect identifications, filler identifications, and correct rejections in culprit-absent lineups—ROC analyses necessitate collapsing this structure into a 2 × 2 matrix, including only culprit identifications, innocent-suspect identifications, correct rejections and incorrect rejections. This is typically done by classifying filler identifications as false rejections in culprit-present lineups and as correct rejections in culprit-absent lineups (Menne et al., 2022; Wells et al., 2015a; Wells et al., 2015c; Winter et al., 2022). . While it has been argued that this treatment of filler identifications is unproblematic, because identifying a filler has the same legal consequences for the suspect as rejecting the lineup because both responses do not provide evidence of the suspect's guilt (Penrod & Cutler, 1995; Wixted & Mickes, 2015), this 2 × 2 classification reduces the data's informative value. That is because filler identifications and lineup rejections likely involve distinct cognitive processes: rejecting a culprit-absent lineup is a correct response, whereas mistakenly identifying a filler is an error (Menne et al., 2022; Wells et al., 2015a; Wells et al., 2015b; Winter et al., 2022). Whether ignored or reclassified, not considering filler identifications as a distinct category leads to the loss of potentially meaningful information..

Moreover, both the diagnosticity ratio and ROC analyses yield only a single measure each. For instance, assessing response biases (i. e., conservative or liberal responding) requires reverting to raw response rates or calculating the response-bias measure *c* taken from signal detection theory (Hautus et al., 2021). Determining lineup fairness demands additional experimental approaches, (Doob & Kirshenbaum, 1973; Lee et al., 2022) which have also been criticized (e. g., Menne et al., 2023b). Importantly, neither the diagnosticity ratio nor ROC analyses enable the measurement of the cognitive processes that underlie eyewitness responses to lineups. However, understanding these latent cognitive processes is crucial in order to fully evaluate the influences of forensically relevant factors. Finally and critically, neither the diagnosticity ratio nor ROC analyses have been validated specifically for use with eyewitness response data. To date, in no empirical validation studies has it been demonstrated that either measure reliably and sensitively captures what it was designed to assess. In light of these limitations, it is advantageous to use a measurement model that does not partially disregard relevant data categories, takes into account all available responses—including filler identifications—and allows for the direct, unconfounded, and simultaneous measurement of multiple cognitive processes underlying eyewitness responses to lineups. Ideally, such a model should also have undergone empirical validation confirming that it accurately captures the cognitive processes it was intended to capture.

The thoroughly-validated *two-high threshold eyewitness identification model* (Menne et al., 2022; Winter et al., 2022; illustrated in Figure 1) fulfills all these criteria. It takes into account all possible observable responses that might occur when eyewitnesses are presented with a lineup. In culprit-present lineups (see rounded rectangle on the left side of the upper tree of Figure 1), the possible responses are culprit identifications, filler identifications and lineup rejections (see rectangles on the right side of the upper tree of Figure 1). In culprit-absent lineups (see rounded rectangles on the left side of the lower tree of Figure 1), the possible responses are innocent-suspect identifications, filler identifications and lineup rejections (see rectangles on the right side of the lower tree of Figure 1). Based on these observable responses, the two-high threshold eyewitness identification model permits concurrent and direct measurement of the probabilities of four distinct latent cognitive processes that underlie eyewitness responses to lineups: two detection-based processes (culprit-presence detection and culprit-absence detection) and two non-detection-based processes (biased suspect selection and guessing-based selection). The two-high threshold eyewitness identification model is part of the multinomial processing tree (MPT) model family (Batchelder & Riefer, 1999; Erdfelder et al., 2009; Riefer & Batchelder, 1988). MPT models are straightforward and popular measurement models widely applied across various domains of cognitive research (Alexander et al., 2016; Bayen & Murnane, 1996; Berres et al., 2025; Buchner et al., 1995; Erdfelder et al., 2007; Kroneisen et al., 2021; Lee et al., 2020). Within MPT models, it is assumed that observable responses (here eyewitness responses to lineups) stem from underlying cognitive processes that occur with specific probabilities. These probabilities are precisely defined by the model equations and represented by model parameters (letters denoted alongside the branches of the tree depicted in Figure 1), which can be statistically compared using accessible software such as multiTree (Moshagen, 2010).

With probability *dP*, the presence of the culprit is detected in a culprit-present lineup, which leads to the identification of the culprit. With the probability 1 – *dP*, the presence of the culprit is not detected. However, two non-detection-based processes can still lead to the identification of the culprit. With conditional probability *b*, biased suspect selection occurs, which also leads to the identification of the culprit. Specifically, if the culprit stands out from the lineup noticeably to the eyewitness, for example due to conspicuous characteristics of the culprit's photograph or appearance relative to the filler photographs, biased suspect selection may occur with the conditional probability $b > 0$. With the conditional probability 1 – *b*, no biased suspect selection occurs. In this case, with the conditional probability *g*, eyewitnesses may select someone from the lineup based on guessing. Conditional upon a guessing-based selection, the probability of the identification of the culprit is given by the reciprocal of the number of individuals presented in the lineup ($1 \div n$). For example, in a six-person lineup, the random sampling probability of the culprit is $1 \div 6$ (approximately .17). With the counter-probability of $5 \div 6$ (approximating .83), a filler is selected. If no culprit-presence detection (1 – *dP*), no biased suspect selection (1 – *b*) and no guessing-based selection (1 – *g*) occurs, the lineup is falsely rejected.

With probability *dA*, the absence of the culprit is detected in a culprit-absent lineup, which leads to the correct rejection of the lineup. If the absence of the culprit is not detected (1 – *dA*), the same non-detection-based processes as in a culprit-present lineup occur. That is because in the absence of detection-based processes, there is no reason to assume that non-detection-based processes differ depending on the suspect's guilt. However, here it is important to note that selection of the suspect in a culprit-absent lineup—whether due to biased or guessing-based selection—results in the identification of an innocent person.

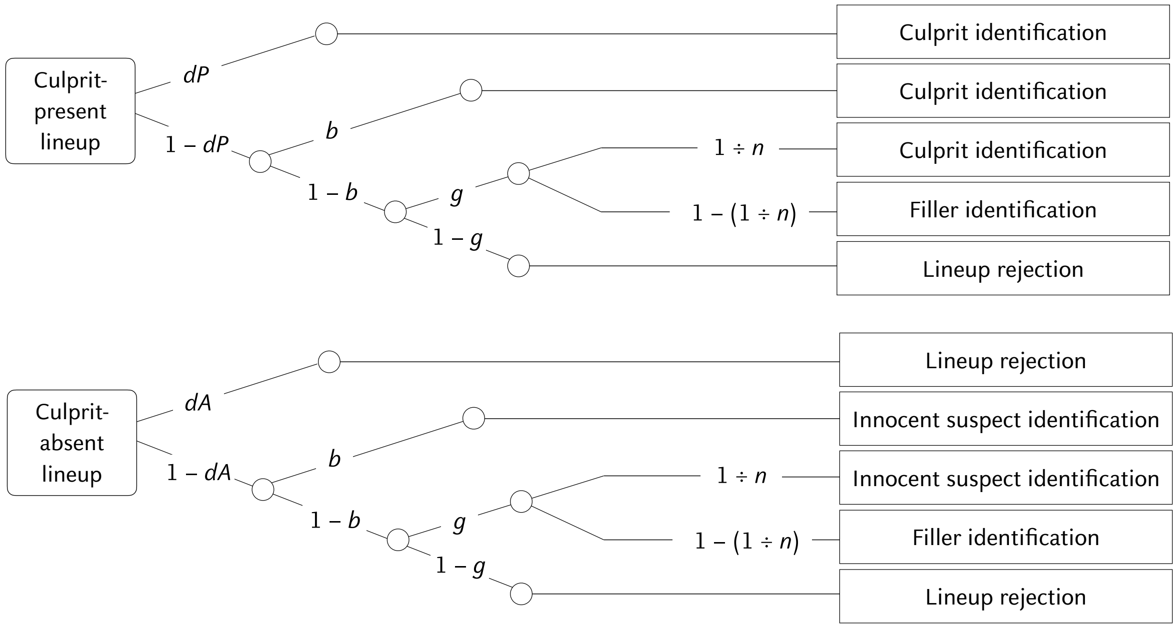

Figure 1. Representation of the two-high threshold eyewitness identification model (Menne et al., 2022; Winter et al., 2022). Rounded rectangles on the left depict lineup types: culprit-present (upper tree) and culprit-absent (lower tree) lineups. Rectangles on the right represent observable response categories (culprit identifications, filler identifications and lineup rejections in culprit-present lineups, as well as lineup rejections, filler identifications and innocent-suspect identifications in culprit-absent lineups). Letters along the branches indicate probabilities of latent cognitive processes: *dP* (culprit-presence detection), *dA* (culprit-absence detection), *b* (biased suspect selection) and *g* (guessing-based selection). If guessing-based selection occurs, the probability of identifying the suspect—whether the actual culprit or an innocent suspect—is given by $1 \div n$, where *n* is the lineup size (number of individuals present in the lineup).

The two-high threshold eyewitness identification model was extensively validated by reanalyses of previously published data sets from various laboratories, with different stimulus material and experimental designs (Menne et al., 2022), as well as by a series of dedicated validation experiments (Winter et al., 2022), providing evidence that the parameters of the model directly and sensitively reflect the cognitive processes they were designed to measure.

Following its successful validation, the two-high threshold eyewitness identification model has been used as a valuable tool for examining the effects of various prominent forensically relevant factors on the cognitive processes underlying eyewitness responses to lineups. By applying of the two-high threshold eyewitness identification model it was demonstrated that utilizing morphed filler photographs does not increase the probability of biased suspect selection (parameter *b*) relative to non-morphed filler photographs when the contents of the pre-lineup instructions were comparable in both conditions (Menne et al., 2023b). Also, it was shown that utilizing AI-generated filler photographs even reduced the probability of biased suspect selection relative to database-derived filler photographs (Bell et al., 2024). Furthermore, administering pre-lineup instructions suggesting that only the first identification in a sequential lineup would be counted (so-called “first-yes-counts” instructions) selectively reduced the probability of guessing-based selection relative to pre-lineup instructions not suggesting that only the first identification in a sequential lineup would be counted (Winter et al., 2023). Reducing lineup sizes (the number of individuals present in a lineup) to smaller, two- or three-person lineups resulted in an increased probability of culprit-presence detection (parameter *dP*) and a reduced probability of guessing-based selection relative to larger, five- or six-person lineups (Menne et al., 2023a). Lineup position selectively affects guessing-based selection, because a person in the first position within a lineup is more likely to be selected based on guessing than are those in other positions (Mayer et al., 2024). Culprit-presence detection and, to some extent, culprit-absence detection decrease with age, while guessing-based selection increases (Mayer et al., in press). And finally, disclosing the number of lineups that would be shown to an eyewitness after a multiple-culprit crime increases the probability of guessing-based selection compared to when no number or a false, high number was announced beforehand (Menne et al., 2025).

These examples illustrate the utility of the two-high threshold eyewitness identification model in investigating the effects of various forensically relevant factors on the cognitive processes underlying eyewitness responses to lineups. In the present dissertation, the two-high-threshold eyewitness identification model was applied to investigate how various important forensically relevant factors—beyond those illustrated in the preceding examples—affect the cognitive processes underlying eyewitness responses. Specifically, three key research questions are addressed, focusing on the effects of (1) lineup size and pre-lineup instructions, (2) delays between the witnessed event and the lineup and (3) the provision of culprit descriptions prior to the lineup.

On the combined effect of lineup size and pre-lineup instructions

One key lineup-specific factor influencing eyewitness responses to lineups is the *lineup size*—that is, the number of individuals included in the lineup. In order to assess effects of lineup size, larger (including more fillers) lineups are compared to smaller (including fewer fillers) lineups (Akan et al., 2021; Juncu & Fitzgerald, 2021; Menne et al., 2023a; Nosworthy & Lindsay, 1990; Wooten et al., 2020). The main rationale for using larger lineups is that they “protect” innocent suspects from being randomly selected by spreading guesses across a greater number of lineup members (Levi & Lindsay, 2001; Wells, 2001; Wells et al., 2020). Given that the random sampling probability of the suspect is determined by the reciprocal of the number of lineup members ($1 \div n$, where n is lineup size), the probability of a random selection of a suspect decreases as lineup size increases. While this is an advantage of larger over smaller lineups, it is important to note that this mathematical fact pertains to all suspects, innocent *and* guilty. Consequentially, a cost-benefit tradeoff has often been reported: while larger lineups reduce innocent-suspect identifications, they also reduce culprit identifications. Conversely, smaller lineups increase culprit identifications, but at the cost of more innocent-suspect identifications (Juncu & Fitzgerald, 2021). Given this tradeoff, lineup size recommendations vary across jurisdictions: in some jurisdictions, for example, in Portugal, Italy and Russia, three-person lineups are recommended, whereas in others, for example, in England, Wales and Scotland, up 12-person lineups are recommended for photo lineups (Fitzgerald et al., 2021).

However, beyond the sampling probability—which is only relevant if responses are made based on guessing in the first place—observable lineup responses (including the reported cost-benefit tradeoff) can be stimulated by detection-based and non-detection based cognitive processes. It is therefore critical to determine how variations of lineup size affect these processes. In a previous study in which the two-high threshold eyewitness identification model was applied (Menne et al., 2023a), it was found that reducing lineup size (e. g., from six to three or from five to two individuals) increased the probability of culprit-presence detection (parameter dP) while it decreased the probability of guessing-based selection (parameter g). While this suggests a strong advantage of smaller over larger lineups, there were still more innocent-suspect identifications in smaller compared to larger lineups. The culprit-absent decision tree (lower tree of Figure 1) illustrates why this was the case: Menne et al. (2023a) reported no differences in biased suspect selection (parameter *b*) and culprit-absence detection (parameter *dA*) depending on lineup size. Therefore, the difference of the number of innocent-suspect identifications between the two lineup sizes was determined by the term $g \times (1 \div n)$. Even though the probability of guessing-based selection (parameter *g*) occurring in the first place was significantly lower in smaller than in larger lineups, the considerably higher sampling probability ($1 \div n$) outweighed this effect. For instance, in three-person lineups ($n = 3$), $1 \div n$ is $1 \div 3$ (approximating .33), whereas in six-person lineups ($n = 6$), $1 \div n$ is $1 \div 6$ (approximating .17)—therefore, the sampling probability ($1 \div n$) is twice as high in three-person lineups than in six-person lineups. In two-person lineups, $1 \div n$ is even two-and-a-half times higher (.50) than in five-person lineups (.20). Given this, there is a need for strategies to counteract this particular disadvantage of smaller compared to larger lineups. Figure 1 illustrates a potential countermeasure: it is inevitable that a decrease in lineup size *n* always leads to an increase of its reciprocal $1 \div n$. However, if the probability of guessing-based selection occurring initially can be reduced to a level even below that achieved by decreasing lineup size alone, it may be possible to reduce innocent-suspect identifications—preferably to a level lower than, or comparable to, those in larger lineups. At the same time, due to the higher probability of culprit-presence detection in smaller lineups, culprit identifications should not be reduced.

A lineup-specific factor that specifically affects the probability of guessing-based selection is the instructions given to eyewitnesses before a lineup (Menne et al., 2022; Winter et al., 2022). It has been shown in two reanalyses of published data conducted with the two-high threshold eyewitness identification model (Lampinen et al., 2020; Malpass & Devine, 1981) that when eyewitnesses receive pre-lineup instructions suggesting a high probability that the culprit is present in the lineup, the probability of guessing-based selection tends to increase compared to neutral pre-lineup instructions that do not specify such a probability—henceforth neutral instructions (Menne et al., 2022). Conversely, pre-lineup instructions suggesting a low probability that the culprit is present in the lineup reduce the probability of guessing-based selection (Winter et al., 2022).

Taking these considerations into account, two novel hypotheses concerning observable responses were tested in Experiments 1a and 1b. First, if pre-lineup instructions insinuating that the culprit is unlikely to be in the lineup (henceforth low-culprit-probability instructions) effectively reduce the probability of guessing-based selection, then combining smaller lineups with such pre-lineup instructions should result in comparable or lower innocent-suspect-identification rates relative to larger lineups with neutral pre-lineup instructions (not insinuating a low probability of culprit presence). Second, while the decrease in the probability of guessing-based selection also affects culprit identification rates, the increase in $1 \div n$—which also pertains to culprit-present lineups—should outweigh this effect. In addition to these non-diagnostic culprit identifications, the probability of culprit-presence detection (parameter *dP*) in smaller compared to larger lineups was reported to be higher (Menne et al., 2023a). Taken together, both factors—the increase in $1 \div n$ and the higher probability of culprit-presence detection—should therefore lead to higher culprit identification rates in smaller lineups, even despite the reduction of non-diagnostic, guessing-based culprit identifications that will occur by pairing the smaller lineups with low-culprit-probability instructions.

In line with the previous study the present experiments were built on (Menne et al., 2023a), three-person lineups were compared to six-person lineups. Within each lineup-size condition, participants received either low-culprit-probability or neutral instructions. Given that six-person lineups with neutral instructions represent standard police practice in the United States of America (Police Executive Research Forum, 2013; Wells et al., 2020), this combination was considered to be the standard of comparison. For the experimental manipulations to manifest in effects on observable responses, first, they had to influence the cognitive processes underlying these responses. Thus, the first aim was to replicate prior findings regarding the following effects on cognitive processes: three-person lineups should effectuate a higher probability of culprit-presence detection (parameter *dP*) and a lower probability of guessing-based selection (parameter *g*) relative to six-person lineups (Menne et al., 2023a); and low-culprit-probability instructions should effectuate a lower probability of guessing-based selection (parameter *g*) relative to neutral instructions (Menne et al., 2022; Winter et al., 2022). At the level of observable responses, it was expected that the combination of three-person lineups with low-culprit-probability instructions would result in innocent-suspect identification rates to be lower or comparable to those in six-person lineups with neutral instructions, while significantly higher culprit identification rates would be maintained in three-person lineups with low-culprit-probability instructions.

Given that both sequential and simultaneous lineup procedures may be used worldwide (Fitzgerald et al., 2021) and to ensure robustness and replicability (Open Science Collaboration, 2015), these predictions were tested in two experiments using sequential (Experiment 1a) and simultaneous (Experiment 1b) lineup formats.

Experiment 1a

The goal of Experiment 1a was to investigate the combined effect of lineup size and pre-lineup instructions on culprit identifications and innocent-suspect identifications, as well as their underlying cognitive processes, using sequential lineups. Participants watched one of two parallel staged-crime videos of four football hooligans (henceforth the culprits) verbally and physically attacking a fan of a rivaling club. Subsequently, half of them received low-culprit-probability instructions intended to discourage guessing-based selection, while the other half received neutral instructions. Afterwards, they were presented with four photographic lineups containing either six (one suspect and five fillers) or three individuals (one suspect and a random subset of two of the five fillers presented in six-person lineups). Two of the lineups were culprit-present lineups (including a culprit from the previously-seen video), while the other two were culprit-absent lineups (including an innocent suspect). Innocent suspects were portrayed by photographs of the culprits from the parallel video the participants had not previously seen. Lineups were presented in sequential format. Therefore, each photograph was shown individually and participants had to either identify the presented person as one of the culprits or reject the lineup by indicating that the presented person was not one of the culprits.

At the level of the latent cognitive processes, the results from prior studies were replicated. Three-persons lineups yielded a higher probability of culprit-presence detection (parameter *dP*) and a lower probability of guessing-based selection (parameter *g*) compared to six-person lineups. Furthermore, providing low-culprit-probability instructions resulted in a decreased probability of guessing-based selection (parameter *g*) relative to neutral instructions. At the level of observable responses, two novel predictions that were derived from the structure of the two-high threshold eyewitness identification model were tested. It was confirmed that combining three-person lineups with low-culprit-probability instructions resulted in innocent-suspect-identification rates that were statistically indistinguishable from those in six-person lineups with neutral instructions, while culprit-identification rates were significantly higher in three-person lineups with low-culprit-probability instructions than in six-person lineups with neutral instructions.

Experiment 1b

To test the replicability and generalizability of the results—considering that recommended lineup formats vary across jurisdictions (Juncu & Fitzgerald, 2021)—Experiment 1b served as a conceptual replication of Experiment 1a, differing only in that the simultaneous lineup format instead of the sequential one was applied. Photographs of all individuals in the lineup were thus shown at once. Participants had to either identify one of the persons displayed on the photographs as one of the culprits or reject the lineup by indicating that none of the persons were one of the culprits. As in Experiment 1a, at the level of latent cognitive processes, it was found that three-person lineups yielded a higher probability of culprit-presence detection (parameter *dP*) than six-person lineups. While there was no statistically significant difference in the probability of guessing-based selection (parameter *g*) between three-person lineups and six-person lineups, low-culprit-probability instructions led to a reduced probability of guessing-based selection (parameter *g*) relative to neutral instructions. Experiment 1b was also aimed at replicating the predictions of Experiment 1a regarding observable responses. It was confirmed that combining three-person lineups with low-culprit-probability instructions resulted in innocent-suspect-identification rates that were statistically indistinguishable from those in six-person lineups with neutral instructions, while culprit-identification rates were significantly higher in three-person lineups with low-culprit-probability instructions than in six-person lineups with neutral instructions.

Discussion

Across both experiments, it was demonstrated that smaller, three-person lineups led to an increased probability of culprit-presence detection relative to six-person lineups. Additionally, three-person sequential lineups led to a decreased probability of guessing-based selection relative to six-person lineups. In line with previous research (Winter et al., 2022), it was also demonstrated that low-culprit-probability instructions led to a decreased probability of guessing-based selection relative to neutral instructions. Importantly, when combining three-person lineups with low-culprit-probability instructions, innocent-suspect-identification rates were effectively reduced to a level comparable to those in six-person lineups combined with neutral instructions, while culprit-identification rates were significantly higher in three-person lineups with low-culprit-probability instructions than in six-person lineups with neutral instructions. The two-high threshold eyewitness identification model provided unique and valuable insights—not only for measuring latent cognitive processes but also for generating hypotheses to potentially improve lineup procedures.

Previous research had already shown beneficial effects of small lineups on culprit-presence detection and guessing-based selection (Menne et al., 2023a). However, the unique structure of the model allowed for generating further hypotheses not only about underlying cognitive processes but also about observable responses. It is important to emphasize that these insights were only possible through the use of the two-high threshold eyewitness identification model, which allowed for the differentiation between culprit-presence detection, guessing-based selection, and the suspect's sampling probability as determined by lineup size ($1 \div n$). By focusing on the interplay between these distinct factors that underlie eyewitness responses to lineups, it was possible to identify a particularly effective strategy to balance the necessity of achieving high rates of culprit identifications while minimizing the risk of innocent-suspect identifications. This was achieved by combining three-person lineups with low-culprit-probability instructions known to discourage guessing-based selection (Winter et al., 2022). This approach successfully resulted in increased culprit identifications without a corresponding increase in innocent-suspect identifications, relative to the standard six-person lineup with neutral instructions. While it is necessary to extend procedures to more naturalistic settings before deriving recommendations from the present results, the findings from the present experiments may represent a meaningful step toward resolving a central dilemma in lineups: maintaining high culprit identification rates while not increasing innocent-suspect identification rates.

On the effect of crime-lineup delays

Before a lineup can be presented to an eyewitness and lineup-specific factors like lineup size and pre-lineup instructions need to be taken into account, evidence must be accumulated until there is sufficient justification for including a suspect in a lineup. This is necessary because it increases the probability that the actual culprit and not an innocent suspect is present in the lineup (Wells et al., 2020). However, the necessary investigations to establish evidence-based suspicion inevitably require time, which can lead to considerable delays between a crime and an associated lineup—especially given that resource limitations of investigators have often been reported (His Majesty's Inspectorate of Constabulary and Fire &

Rescue Services, 2015). Indeed, in a number of British archival studies, it has been reported that delays of one to three months between the crime and the lineup are most common (Horry et al., 2014; Horry et al., 2012; Memon et al., 2011; Pike et al., 2002). Given that it is a well-established fact that human memory deteriorates over time (Ebbinghaus, 1885; Radvansky et al., 2022; Rubin & Wenzel, 1996), it has been recommended that a lineup needs to be presented to an eyewitness as soon as possible after a crime (Wells et al., 2020). This highlights a dilemma faced by legal professionals during criminal investigations: pre-lineup investigative factors are often interdependent. While it is necessary to accumulate sufficient evidence-based suspicion before a lineup, it is also important to account for the need that the lineup is presented to an eyewitness as soon as possible after a crime.

The effects of delays between witnessing a crime and completing a lineup on eyewitness responses to lineups have been investigated in numerous studies. However, previous research has been limited in several ways. First, most studies compared only short or no delays with long delays (e. g., Palmer et al., 2013; Sauer et al., 2010; Wetmore et al., 2015). This limitation to two time points does not allow for a more comprehensive understanding of delay effects across the passage of time. Second, most prior analyses were based solely on observable response rates (Egan et al., 1977; Gettleman et al., 2021; Giacona et al., 2021; Karageorge & Zajac, 2011; Krafka & Penrod, 1985; Lin et al., 2019). As explicated in the Introduction of this dissertation, while also informative, the exclusive investigation of raw response rates is insufficient because observable responses to lineups can result from a variety of latent cognitive processes, some of which are detection-based and are therefore diagnostic of the suspect's guilt (i. e., culprit-presence and culprit-absence detection), while others are not detection-based and are therefore be not diagnostic of the suspect's guilt (i. e., guessing-based selection or biased suspect selection). Third, findings from prior studies were mixed, given that in some, a decline in culprit identifications with longer delays was reported (e. g., Cutler et al., 1986; Lin et al., 2019; Shepherd, 1983), while in others, no effect of delay was reported (e. g., Egan et al., 1977; Karageorge & Zajac, 2011; Krafka & Penrod, 1985; Wetmore et al., 2015). However, given that raw response rates are insufficient in order to draw conclusions about underlying cognitive processes, it is unclear what these stable rates of culprit identifications reflect. It is possible that they reflect a true lack of delay effects on culprit-presence detection or a compensatory increase in guessing-based selection that mask detection-based declines in culprit identifications. Furthermore, small sample sizes may have resulted in insufficient statistical power.

Given these ambiguities, it remains unclear whether crime-lineup delays have negative effects on the cognitive processes underlying eyewitness responses to lineups. Using the two-high threshold eyewitness identification model, in Experiment 2, limitations of prior research were addressed (1) by examining effects across four different delay intervals (no delay, one day, one week and one month), (2) by parallel and direct measurement of delay effects on culprit-presence detection and guessing-based selection separately and (3) by using a large sample size (N = 2108) to ensure adequate statistical sensitivity.

Here it was expected that culprit-presence detection (parameter *dP*) would decrease with increasing delay, consistent with established forgetting functions (Rubin & Wenzel, 1996; Wixted & Ebbesen, 1991). Hypotheses regarding guessing-based selection (parameter *g*) were less clear-cut. It was deemed possible that guessing-based-selection remains stable, because factors known to affect the parameter—for instance, pre-lineup instructions—were identical across conditions. Alternatively, it was deemed possible that guessing-based-selection would increase with decreasing parameter *dP* due to *compensatory guessing* (Küppers & Bayen, 2014; Meiser et al., 2007)—that is, an eyewitness' attempt to counterbalance the decline in culprit-presence detection by exhibiting a higher probability of guessing-based selection. No strong predictions were made regarding culprit-absence detection (parameter *dA*), as low baseline values have been reported even in experiments without delay (e. g., Mayer et al., 2024; Menne et al., 2023b), making further reductions unlikely.

Experiment 2

As in Experiments 1a and 1b, participants watched a staged-crime video including four culprits. Subsequently, participants assigned to the no-delay condition immediately completed the four simultaneous lineups, two of which were culprit-present (including a culprit from the previously-seen video), while the other two were culprit-absent (including a designated innocent suspect). Participants assigned to either of the delay conditions were instead required to complete the simultaneous lineups after a delay of either one day, one week, or one month.

As in previous studies, culprit-identification rates decreased with increasing delay (Cutler et al., 1986; Lin et al., 2019; Shepherd, 1983), while innocent-suspect identification rates remained stable (Lucas et al., 2021; Wetmore et al., 2015). Furthermore, the probability of culprit-presence detection (*dP*) steadily and significantly declined with increasing delay, with the most rapid decline occurring at a short crime-lineup delay (between no-delay condition and the one-day-delay condition). Moreover, *dP* approximated zero after a delay of one month. In line with previous research on forgetting functions (Rubin & Wenzel, 1996; Wixted & Ebbesen, 1991), this decline in *dP* was best described by a power function. The probability of guessing-based selection (*g*) did not vary as a function of delay. Therefore, there was no evidence of compensatory guessing (Küppers & Bayen, 2014; Meiser et al., 2007)—participants did not attempt to counterbalance the decline in culprit-presence detection by exhibiting a higher probability of guessing-based selection. Finally, in line with previous studies in which no explicit manipulation of culprit-absence detection was applied (Mayer et al., 2024; Menne et al., 2023b; Winter et al., 2022), the probability of culprit-absence detection was close to zero even in the no-delay condition and remained equally low across all delays.

Discussion

The results of Experiment 2 demonstrate that the memory-based process of culprit-presence detection (parameter *dP*) declines systematically with increasing delay. By applying an experimental design with four different delay intervals (no delay, one day, one week, and one month), it was possible to show that in line with well-established forgetting functions—including human face recognition (Wixted & Ebbesen, 1991)—this decline was best described by a power function. In contrast to the decline in memory-based culprit-presence detection, the process of guessing-based selection (parameter *g*) remained stable across all delays, thus providing no evidence for compensatory guessing. Likewise, culprit-absence detection (parameter *dA*) was unaffected by delay and remained near zero across all conditions. This pattern is consistent with previous literature (Mayer et al., 2024; Menne et al., 2023b; Winter et al., 2022) and is likely attributable to the inherent difficulty of ruling out each individual in a lineup, as opposed to detecting the presence of a single face (the culprit's). Taken together, the findings indicate that the observed decrease in culprit identifications with increasing delay that has also been reported by numerous researchers priorly (e. g., Cutler et al., 1986; Lin et al., 2019; Shepherd, 1983) is driven specifically by a decline in memory-based culprit-presence detection, rather than by changes in guessing-based selection.

These results also underscore the importance of the recommendation to minimize crime-lineup delays as much as possible (Wells et al., 2020) and of informing legal professionals about the effects of delay on eyewitness memory. Consistent with prior findings (Cutler et al., 1986; Egan et al., 1977; Lin et al., 2019; Shapiro & Penrod, 1986), here primarily culprit-present lineups and not culprit-absent lineups were affected by delay: while memory-based culprit-presence detection declined substantially with increasing delay and approximated zero after a delay of one month, both guessing-based selection and culprit-absence detection remained unaffected. This pattern has important implications regarding culprit identifications—given that culprit-presence detection approached zero after a one-month delay, any culprit identification at this stage is likely attributable to guessing-based selection.

However, these delays are often inevitable and real-world delays between a crime and a lineup have often been reported to range from one to three months (Horry et al., 2014; Horry et al., 2012; Memon et al., 2011; Pike et al., 2002). One factor contributing to these delays are the thorough investigations necessary to provide strong evidence-based grounds against potential suspects before presenting them in a lineup. Given that these delays may often be unavoidable, it is important to educate legal professionals about the detrimental effects these delays may have on eyewitness memory. Jurors and judges should be aware of the steep decline in memory-based culprit-presence detection over time and the potential risks associated with suspect identifications following extended delays.

On the effect of culprit descriptions

Importantly, crime-lineup delays are not the only pre-lineup investigative factor that is linked to thorough investigations. In Experiment 3 of this dissertation, a similar dilemma regarding a direct consequence of an often necessary pre-lineup investigative factor is examined: while a detailed initial description of a culprit's appearance is often crucial in order to narrow the pool of potential suspects (Brown et al., 2008; Dodson et al., 2024; Wells et al., 2020), it has been priorly reported studies that such descriptions can have negative effects on eyewitness responses to subsequent lineups. This phenomenon, known as the *verbal overshadowing effect* (Schooler & Engstler-Schooler, 1990), refers to the finding that providing verbal description of a culprit can result in decreased culprit-identification rates in culprit-present lineups (Alogna et al., 2014; Holdstock et al., 2022; Schooler & Engstler-Schooler, 1990; Wilson et al., 2018). However, given that in both the seminal study (Schooler & Engstler-Schooler, 1990) as well as its large-scale replication (Alogna et al., 2014), exclusively culprit-present lineups were used, conclusions regarding the reasons behind these decrements remain impossible. That is because without culprit-absent lineups, not all data categories from lineups can be analyzed, which prevents conclusions about the origins of the effect.

One possible explanation for the decrease in culprit identifications is that verbalizing a face interferes with holistic face processing (Cheung & Gauthier, 2010; Richler & Gauthier, 2014) and instead fosters feature-based processing that has been discussed to impair facial recognition (Chin & Schooler, 2008; Wickham & Lander, 2008), thereby complicating culprit-presence detection. Another possibility is that the difficulty of describing a face increases caution, thereby reducing the willingness to make a guessing-based selection. Accordingly, in studies in which culprit-present and culprit-absent lineups were used, reductions of both culprit and innocent-suspect identifications were reported after a description had been provided (Clare & Lewandowsky, 2004; Dodson et al., 2024; Holdstock et al., 2022; Mickes & Wixted, 2015; Smith & Flowe, 2015; Wilson et al., 2018).

This illustrates that the analysis of observable response rates alone is insufficient in order to determine the origins of the verbal overshadowing effect, because the reduction in culprit identifications can be driven by several underlying cognitive processes. Therefore, in order to determine the cognitive processes underlying the verbal overshadowing effect, the two-high threshold eyewitness identification model was applied. If facial recognition is impaired due to an interference of feature-based description with holistic face processing, then the probability of culprit-presence detection (parameter *dP*) should be lower when a culprit description was provided before a lineup compared to when no culprit description was provided beforehand. If eyewitnesses are more reluctant to make a guessing-based selection due to the experienced difficulty of describing a culprit, then the probability of guessing-based selection (parameter *g*) should be lower when a culprit description was provided before a lineup compared to when no culprit description was provided beforehand. It is also possible that both aspects contribute to the verbal overshadowing effect, in which case both culprit-presence detection (parameter *dP*) and guessing-based selection (parameter *g*) should be lower when a culprit description was provided before a lineup compared to when no culprit description was provided beforehand.

Experiment 3

Given that the size of the verbal overshadowing effect varies (Chin & Schooler, 2008; Meissner & Brigham, 2001a), here manipulations were implemented which are known to reliably elicit the effect. Participants were randomly assigned to either a culprit-description condition or a no-culprit-description condition. After having been shown a staged-crime video in which four culprits attack a victim, participants were required to play Tetris for 14 minutes. This delay aligns with typical response times for crimes of comparable severity (e. g., Houston Police Department Office of Planning & Data Governance, 2023; Salt Lake City Police Department, 2022) and has been discussed to increase the verbal overshadowing effect relative to longer or shorter crime-description delays (Protzko et al., 2023). Afterwards, for five minutes, participants in the culprit-description condition were instructed to provide feature-based face descriptions entailing everything they could remember about the four culprits' appearances. This was done because this has been found to increase the verbal overshadowing effect compared to when participants are warned to describe only what they are certain of (Clare & Lewandowsky, 2004; Dodson et al., 1997) or when participants are asked to provide holistic descriptions (Wickham & Lander, 2008). Participants in the no-culprit-description condition were instructed to list countries and their capitals for five minutes. Subsequently, participants were presented with four lineups, two of which were culprit-present (that is ,they included a culprit from the previously-seen video), while the other two were culprit-absent (that is, they included a designated innocent suspect). It was found that providing a culprit description prior to a lineup resulted in both a significant reduction of culprit-presence detection (parameter *dP*) and guessing-based selection (parameter *g*) compared to when no culprit description was provided beforehand.

Discussion

The goal of Experiment 3 was to investigate the cognitive processes underlying the verbal overshadowing effect—the finding that providing a description of the culprit prior to a lineup leads to reduced culprit identifications in culprit-present lineups (Schooler & Engstler-Schooler, 1990). Researchers priorly reported reduced culprit identification rates (Alogna et al., 2014; Holdstock et al., 2022; Schooler & Engstler-Schooler, 1990; Wilson et al., 2018) and reduced innocent-suspect-identification rates (Clare & Lewandowsky, 2004; Dodson et al., 2024; Holdstock et al., 2022; Mickes & Wixted, 2015; Smith & Flowe, 2015; Wilson et al., 2018) if a culprit description was given before the lineup was shown. This pattern of observable responses allows for multiple explanations: it could be explained by a reduction in culprit-presence detection, guessing-based selection or both. Here the two-high threshold eyewitness identification model was applied in oder to examine the underlying cognitive processes of the verbal overshadowing effect. Providing culprit descriptions instead of naming countries and capitals during the same time period resulted in both a reduction of culprit-presence detection (parameter *dP*) and guessing-based selection (parameter *g*). These findings support both possible explanations regarding the origins of the verbal overshadowing effect: they suggest that the effect is at least partly caused by culprit descriptions interfering with face recognition, reducing the ability to detect the presence of the culprit in the lineup (e. g., Chin & Schooler, 2008; Wickham & Lander, 2008) and by a greater reluctance of participants to make a guessing-based selection (e. g., Clare & Lewandowsky, 2004).

General discussion

In the present dissertation, effects of several forensically relevant factors on the cognitive processes underlying eyewitness responses to lineups were investigated. Specifically, effects of the lineup-specific factors *lineup size* and *pre-lineup instructions* (Experiment 1a and 1b), as well as effects of the pre-lineup investigative factors *crime-lineup delay* (Experiment 2) and *culprit descriptions* (Experiment 3) were examined. By applying the well-validated two-high threshold eyewitness identification model, it was possible to move beyond observable responses and to directly assess the effects of these forensically relevant factors on four distinct cognitive processes underlying eyewitness responses to lineups—specifically, two detection-based processes (culprit-presence detection and culprit-absence detection) and two non-detection-based processes (biased suspect selection and guessing-based selection).

Effects of such forensically relevant factors are complex. For example, because smaller lineups have been reported to increase culprit-presence detection and reduce guessing-based selection (Menne et al., 2023a), they can be considered beneficial regarding eyewitness's latent cognitive processes. However, smaller lineups also inherently increase the random sampling probability of each lineup member—including an innocent suspect—and are therefore sometimes viewed critically (Juncu & Fitzgerald, 2021). Here, Experiments 1a and

1b demonstrated that this disadvantage can be offset. Specifically, pre-lineup instructions implying a low probability that the culprit would be present led to a lower probability of guessing-based selection compared to neutral instructions. As a consequence, the combination of three-person lineups and low-culprit-probability instructions resulted in innocent-suspect-identification rates that were statistically indistinguishable from those in standard six-person lineups with neutral instructions. Importantly, culprit-presence detection was higher in three-person lineups than in six-person lineups and low-culprit-probability instructions did not have a negative effect on culprit-presence detection. As a consequence, the combination of three-person lineups and low-culprit-probability instructions resulted in higher culprit-identification rates compared to six-person lineups with neutral instructions. These effects on the observable response rates were consistently observed in both sequential and simultaneous formats. Only by taking into account the unique structure of the two-high threshold eyewitness identification model was it possible to identify the beneficial strategy of considering the two forensically relevant lineup-specific factors lineup size and pre-lineup instructions concurrently. The two-high threshold eyewitness identification model not only allowed for predictions about how these factors would interact, but also for interpreting the resulting distinct patterns of latent cognitive processes and observable responses together.

Furthermore, in the present dissertation, effects of pre-lineup investigative factors were taken into account. In Experiment 2, the effects of delays between the witnessed crime and the lineup were examined. One factor contributing to these crime-lineup delays is the time required to accumulate sufficient evidence-based suspicion to ethically and legally justify including a suspect in a lineup. Among other factors, this contributes to the commonly observed delays of one to three months between the crime and the lineup that has often been reported in archival studies (Horry et al., 2014; Horry et al., 2012; Memon et al., 2011; Pike et al., 2002). The results of Experiment 2 indicated a sharp, power-function-shaped decline in memory-based culprit-presence detection as a function of increasing crime-lineup delay, with culprit-presence detection approaching zero after one month. This decline was not accompanied by changes in the probability of guessing-based selection, indicating that eyewitnesses did not compensate for their declining memory by increasing their tendency to guess. The results are consistent with established forgetting functions in cognitive psychology (Rubin & Wenzel, 1996; Wixted & Ebbesen, 1991) and provide strong support for longstanding recommendations that lineups should be administered as soon as possible following the crime (Wells et al., 2020; Wells et al., 1998). From an applied perspective, these findings also carry critical implications. If culprit-presence detection approaches zero after extended delays, any identifications that occur under such conditions are more likely to result from a non-detection-based process that holds no evidentiary value: guessing-based selection. This underscores the risk of overestimating the evidentiary value of suspect identifications obtained long after the witnessed event. While the need for legal professionals to be informed about the effects of crime-lineup delays on cognitive processes underlying eyewitness lineup responses is underscored by these findings, it remains unresolved how well these results generalize to field conditions. However, given that field conditions have been reported to elicit either comparable (Pozzulo et al., 2008) or even lower (Eisen et al., 2022) accuracy levels compared to laboratory conditions when eyewitnesses were presented with lineups (with a high accuracy being defined by fewer innocent-suspect-identifications and more culprit identifications), caution is warranted also in real-world scenarios for legal professionals when interpreting the evidentiary value of a suspect identification after long crime-lineup delays.

Experiment 3 addressed another pre-lineup investigative factor: the provision of a culprit description. While obtaining culprit descriptions is often vital for narrowing down suspect pools, researchers have priorly reported that providing verbal descriptions of a culprit's features can result in reduced rates of culprit identifications in subsequent lineups. This so-called verbal overshadowing effect (Schooler & Engstler-Schooler, 1990) illustrates a fundamental dilemma faced by legal professionals: While culprit descriptions are needed to locate possible suspects, these very descriptions can jeopardize an eyewitness's ability to identify the culprit in a lineup. Within the present experiment, it was shown that when participants provided a culprit description shortly after viewing a staged-crime video, both the probability of culprit-presence detection and the probability of guessing-based selection were reduced in subsequent lineups. One possible explanation for this is that describing a face may interfere with holistic face processing (e. g., Cheung & Gauthier, 2010; Richler et al., 2011), thereby reducing culprit-presence detection, while simultaneously increasing caution (Clare & Lewandowsky, 2004; Horry et al., 2012), leading to a reduced probability of guessing-based selection. The application of the two-high threshold eyewitness identification model allowed for a precise attribution of the reduced culprit identifications to changes in distinct latent cognitive processes.

Taken together, the results of the present experiments demonstrate that the effects of forensically relevant factors on eyewitness latent cognitive processes and their resulting responses to lineups are complex. Legal professionals—whether investigating officers, judges, jurors or policy makers—are faced with high-stakes decisions that can profoundly impact the lives of both innocent and guilty suspects. The findings presented here contribute to the broader scientific understanding of a subset of these factors, including those under the direct control of legal professionals—such as lineup-specific procedural factors like lineup size and pre-lineup instructions—as well as those that are often unavoidable—such as pre-lineup investigative factors like crime-lineup delays or the provision of culprit descriptions. Importantly, the results indicate that even when procedures are optimized, there are influences outside the control of legal professionals. These influences may still adversely affect eyewitness cognitive processes, given that they can reduce the probability of diagnostic, detection-based processes (such as culprit-presence detection) and foster non-diagnostic, non-detection-based processes (such as guessing-based selection).

The implications of these findings are considerable. Eyewitness responses to lineups carry substantial evidentiary weight in legal proceedings in many jurisdictions around the world (Fitzgerald et al., 2021), yet the two-high threshold eyewitness identification model shows that an identical eyewitness response—such as a suspect identification—can originate from fundamentally different cognitive processes, only one of which provides evidentiary value about a suspect's guilt. This becomes especially problematic in real-world legal contexts, where the suspect's guilt is never definitively known at the time of the lineup. Therefore, a suspect identified through guessing or biased selection may be incorrectly assumed to be the culprit, increasing the risk of wrongful convictions. Furthermore, while post-hoc archival analyses of real-world lineup data are insightful, in most cases, it cannot be definitively verified whether an eyewitness' suspect identification is correct and the suspect is indeed the culprit. The suspect's guilt may remain uncertain, even after conviction, because conclusive evidence—such as DNA matches or reliable confessions—is not always available. As data from wrongful conviction cases and exonerations show, individuals are sometimes convicted mainly based on eyewitness testimony—and in a non-negligible number of cases, wrongly so (Innocence Project, 2024).

It is thus important that legal professionals are educated about the effects of forensically relevant factors on eyewitness responses to lineups and their underlying cognitive processes. Without such knowledge, there is a heightened risk of misinterpreting suspect identifications as compelling evidence of guilt even in situations where they are very likely to be driven by non-detection-based processes such as guessing (e. g., after long crime-lineup delays as demonstrated in Experiment 2). In the worst-case scenario, such misinterpretations may contribute to the conviction of innocent individuals (Innocence Project, 2024).

Furthermore, not only legal professionals but also researchers who address these questions of legal importance must remain aware that the variables they examine concern human lives. While empirical research in controlled laboratory settings is essential for advancing theoretical understanding and informing practice, the transition from laboratory insights to legal application implies immense ethical responsibility. While psychological research is inevitably needed in order to inform science-based legal systems, most recommendations are largely derived from research carried out in laboratory settings—which is why, for some forensically relevant factors, the applicability and relevance to real-world scenarios remains unresolved (e. g., the verbal overshadowing effect in Experiment 3).

The findings from the present dissertation highlight the value of going beyond eyewitness' observable responses to lineups. Applying a well-validated model specifically designed to directly and distinctively measure detection-based and non-detection-based cognitive processes allows for investigating effects of forensically relevant factors on a deeper level, which even allows for deriving strategies regarding their mitigation (Experiments 1a and 1b).

Results from the experiments of the present dissertation provide a valuable foundation for future research. For instance, results of Experiments 1a and 1b could provide a foundation to explore further strategies that aim at reducing innocent-suspect identification rates even more effectively while still obtaining high culprit-identification rates. Given that two-person lineups have been reported to yield higher culprit-presence detection and lower guessing-based selection than larger lineups (Menne et al., 2023a), it is conceivable that combining two-person lineups with low-culprit-probability instructions could reduce innocent-suspect identification rates to levels even lower than those observed for six-person lineups with neutral instructions while still obtaining higher culprit-identification rates. However, two-person lineups have a very high sampling probability of .50. Thus, it is also conceivable that the reduction in guessing-based selection achieved through low-culprit-probability instructions may not be sufficient to offset this high sampling probability. Moreover, four-person lineups have not yet been examined. Accordingly, whether the relationship between lineup size (when combined with low-culprit-probability instructions) and innocent-suspect identification rates follows a linear trend—whereby innocent-suspect identification rates decline as lineup size decreases—remains an open question for future research. The results of Experiment 2 demonstrate a power-function-shaped decline in culprit-presence detection as crime-lineup delay increases. These findings offer a basis for investigating potential countermeasures against this decline—such as context reinstatement. Given that the probability of culprit-presence detection approaches zero after a one-month delay, it could be examined whether reinstating the context of the crime—for instance, by instructing participants to mentally recreate the crime scene, as in the cognitive interview (Fisher et al., 1990; Fisher et al., 1994; Geiselman et al., 1986)—can support eyewitnesses not only in recalling crime details but also in detecting the presence of the culprit during lineups after extended delays. Finally, the results of Experiment 3 offer insights into the cognitive processes underlying the verbal overshadowing effect by demonstrating that providing a culprit description prior to a lineup leads to reduced culprit-presence detection and guessing-based selection, compared to when no description is given. However, the magnitude of the verbal overshadowing effect varies and has been shown to occur reliably predominantly under specific conditions—namely, when a feature-based face description containing all remembered aspects of the culprit's appearance is provided shortly after the crime (with a crime-description delay of 12 to 14 minutes) and no delay occurs between the culprit description and the lineup. Accordingly, it remains an open question whether the verbal overshadowing effect generalizes to real-world contexts, in which delays of several weeks or even months between the description and the lineup are commonly reported (e. g., Horry et al., 2014; Horry et al., 2012). Taken together, the results from all four experiments in the three projects mark an important step in contributing to existing research, but much remains to be explored regarding how forensically relevant factors affect the cognitive processes underlying eyewitness responses to lineups.

Conclusion

In the present dissertation, the two-high threshold eyewitness identification model was applied to systematically examine the effects of forensically relevant factors on the cognitive processes underlying eyewitness responses to lineups. Four experiments conducted across three projects yielded evidence that both lineup-specific factors (e. g., lineup size or pre-lineup instructions) and pre-lineup investigative factors (e. g., delays between crime and lineup or provision of culprit descriptions) affect detection-based and non-detection-based processes, sometimes in combination.

These results underscore the relevance of assessing the effects of such factors not solely in terms of observable behavior, but also in in light of the cognitive processes that underlie them. Even responses that appear superficially identical—such as a suspect identifications—may originate from different cognitive processes. Among these, only the detection of the culprit's presence provides information that is diagnostic of guilt, whereas responses based on biased selection or guessing do not. In practical legal contexts, where neither the suspect's actual guilt nor the cognitive process leading to a suspect identification are known, group-level insights into how certain factors influence cognitive processes leading up to these responses may support the interpretation of suspect identifications.

While the experiments reported in this dissertation contribute meaningful insights into the global scientific understanding of eyewitness responses to lineups and their underlying cognitive processes, they represent only the tip of the iceberg. The effects of many forensically relevant factors—such as the combination of lineup sizes (other than those investigated in the present dissertation) with low-culprit-probability instructions, the use of context reinstatement to counteract crime-lineup-delay-induced declines in culprit-presence detection, or the investigation of the real-world applicability of the verbal overshadowing effect—remain to be systematically explored, altogether posing an important challenge for future research.

References

Akan, M., Robinson, M. M., Mickes, L., Wixted, J. T., & Benjamin, A. S. (2021). The effect of lineup size on eyewitness identification. *Journal of Applied Psychology, 27*(2), 369–392. https://doi.org/10.1037/xap0000340

Alexander, G. E., Satalich, T. A., Shankle, W. R., & Batchelder, W. H. (2016). A cognitive psychometric model for the psychodiagnostic assessment of memory-related deficits. *Psychological assessment, 28*(3), 279-293. https://doi.org/10.1037/pas0000163

Alogna, V. K., Attaya, M. K., Aucoin, P., Bahník, Š., Birch, S., Birt, A. R., Bornstein, B. H., Bouwmeester, S., Brandimonte, M. A., & Brown, C. (2014). Registered replication report: Schooler and Engstler-Schooler (1990). *Perspectives on Psychological Science, 9*(5), 556-578. https://doi.org/10.1177/174569161454565

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin and Review, 6*(1), 57–86. https://doi.org/10.3758/BF03210812

Bayen, U. J., & Murnane, K. (1996). Aging and the use of perceptual and temporal information in source memory tasks. *Psychology and Aging, 11*(2), 293-303. https://doi.org/10.1037//0882-7974.11.2.293

Bell, R., Menne, N. M., Mayer, C., & Buchner, A. (2024). On the advantages of using AI-generated images of filler faces for creating fair lineups. *Scientific Reports, 14*(1), 12304. https://doi.org/10.1038/s41598-024-63004-z

Berres, S., Erdfelder, E., & Kuhlmann, B. G. (2025). Does sleep benefit source memory? Investigating 12-h retention intervals with a multinomial modeling approach. *Memory & cognition, 53*(2), 467-493. https://doi.org/10.3758/s13421-024-01579-8

Brigham, J. C., & Ready, D. J. (1985). Own-race bias in lineup construction. *Law and Human Behavior, 9*(4), 415-424. https://doi.org/10.1007/BF01044480

Brown, C., Lloyd-Jones, T. J., & Robinson, M. (2008). Eliciting person descriptions from eyewitnesses: A survey of police perceptions of eyewitness performance and reported use of interview techniques. *European Journal of Cognitive Psychology, 20*(3), 529-560. https://doi.org/10.1080/09541440701728474

Buchner, A., Erdfelder, E., & Vaterrodt-Plünnecke, B. (1995). Toward unbiased measurement of conscious and unconscious memory processes within the process dissociation framework. *Journal of Experimental Psychology: General, 124*(2), 137–160. https://doi.org/10.1037/0096-3445.124.2.137

Cheung, O. S., & Gauthier, I. (2010). Selective interference on the holistic processing of faces in working memory. *Journal of Experimental Psychology: Human Perception and Performance, 36*(2), 448-461. https://doi.org/10.1037/a0016471

Chin, J. M., & Schooler, J. W. (2008). Why do words hurt? Content, process, and criterion shift accounts of verbal overshadowing. *European Journal of Cognitive Psychology, 20*(3), 396-413. https://doi.org/10.1080/09541440701728623

Clare, J., & Lewandowsky, S. (2004). Verbalizing facial memory: criterion effects in verbal overshadowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(4), 739-755. https://doi.org/10.1037/0278-7393.30.4.739

Clark, S. E. (2012). Costs and Benefits of Eyewitness Identification Reform: Psychological Science and Public Policy. *Perspect Psychol Sci*, *7*(3), 238-259. https://doi.org/10.1177/1745691612439584

Cutler, B. L., Penrod, S. D., O'Rourke, T. E., & Martens, T. K. (1986). Unconfounding the effects of contextual cues on eyewitness identification accuracy. *Social Behaviour*, *1*, 113–134.

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*(3), 287-307. https://doi.org/10.1007/s10979-006-9008-1

Dodson, C. S., Johnson, M. K., & Schooler, J. W. (1997). The verbal overshadowing effect: Why descriptions impair face recognition. *Memory & cognition*, *25*(2), 129-139. https://doi.org/10.3758/BF03201107

Dodson, C. S., Seale-Carlisle, T. M., Garrett, B. L., Kafadar, K., & Yaffe, J. (2024). Persistence of the verbal overshadowing and weapon-focus effects on lineup identification performance. *Journal of Applied Research in Memory and Cognition*. https://doi.org/10.1037/mac0000195

Doob, A. N., & Kirshenbaum, H. M. (1973). Bias in police lineups-partial remembering. *Journal of Police Science and Administration*, *1*(3), 287-293.

Ebbinghaus, H. (1885). *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie [Memory: A Contribution to Experimental Psychology]*. Duncker & Humblot.

Egan, D., Pittner, M., & Goldstein, A. G. (1977). Eyewitness identification: Photographs vs. live models. *Law and Human Behavior*, *1*(2), 199–206. https://doi.org/10.1007/BF01053439

Eisen, M. L., Ying, R. C., Chui, C., & Swaby, M. A. (2022). Comparing witness performance in the field versus the lab: How real-world conditions affect eyewitness decision-making. *Law and Human Behavior*, *46*(3), 175-188. https://doi.org/10.1037/lhb0000485

Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Zeitschrift für Psychologie/Journal of Psychology*, *217*, 108–124. https://doi.org/10.1027/0044-3409.217.3.108

Erdfelder, E., Cüpper, L., Auer, T.-S., & Undorf, M. (2007). The four-states model of memory retrieval experiences. *Zeitschrift für Psychologie/Journal of Psychology*, *215*(1), 61-71. https://doi.org/10.1027/0044-3409.215.1.61

Erickson, W. B., Lampinen, J. M., & Moore, K. N. (2016). Eyewitness identifications by older and younger adults: A meta-analysis and discussion. *Journal of Police and Criminal Psychology*, *31*, 108–121. https://doi.org/10.1007/s11896-015-9176-3

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*(1), 19-28. https://doi.org/10.1080/10683160802047030

Fisher, R., Quigley, K., Brock, P., Chin, D., & Cutler, B. (1990). The effectiveness of the cognitive interview in description and identification tasks. *American Psychology-Law Society, Williamsburg, VA*.

Fisher, R. P., McCauley, M. R., & Geiselman, R. E. (1994). Improving eyewitness testimony with the Cognitive Interview. In *Adult eyewitness testimony: Current trends and developments* (pp. 245-270).

Fitzgerald, R. J., & Price, H. L. (2015). Eyewitness identification across the life span: A meta-analysis of age differences. *Psychological bulletin*, *141*(6), 1228-1265. https://doi.org/10.1037/bul0000013

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*(2), 151–164. https://doi.org/10.1037/a0030618

Fitzgerald, R. J., Rubínová, E., & Juncu, S. (2021). Eyewitness identification around the world. In A. M. Smith, M. Toglia, & J. M. Lampinen (Eds.), *Methods, measures, and theories in eyewitness identification tasks* (pp. 294–322). Routledge/Taylor & Francis Group. https://doi.org/10.4324/9781003138105-16

Gabbert, F., & Brown, C. (2015). Interviewing for face identification. *Forensic facial identification: Theory and practice of identification from eyewitnesses, composites and CCTV*, 15-41. https://doi.org/10.1002/9781118469538.ch2

Geiselman, R. E., Fisher, R. P., MacKinnon, D. P., & Holland, H. L. (1986). Enhancement of Eyewitness Memory with the Cognitive Interview. *The American journal of psychology*, *99*(3), 385-401. https://doi.org/10.2307/1422492

Gettleman, J. N., Grabman, J. H., Dobolyi, D. G., & Dodson, C. S. (2021). A decision processes account of the differences in the eyewitness confidence-accuracy relationship between strong and weak face recognizers under suboptimal exposure and delay conditions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(3), 402–421. https://doi.org/10.1037/xlm0000922

Giacona, A. M., Lampinen, J. M., & Anastasi, J. S. (2021). Estimator variables can matter even for high-confidence lineup identifications made under pristine conditions. *Law and Human Behavior*, *45*(3), 256–270. https://doi.org/10.1037/lhb0000381

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (Vol. 1). Wiley New York.

Guerra, N., Pinto, R., Mendes, P. S., Rodrigues, P. F., & Albuquerque, P. B. (2022). The impact of COVID-19 on memory: Recognition for masked and unmasked faces. *Frontiers in psychology*, *13*, 960941. https://doi.org/10.3389/fpsyg.2022.960941

Hautus, M. J., Macmillan, N. A., & Creelman, C. D. (2021). *Detection theory: A user's guide*. Routledge. https://doi.org/10.4324/9781003203636

His Majesty's Inspectorate of Constabulary and Fire & Rescue Services. (2015). *PEEL: Police effectiveness 2015. A national overview.* Retrieved 22 Mar 2025 from https://assets-hmicfrs.justiceinspectorates.gov.uk/uploads/police-effectiveness-2015.pdf

Holdstock, J. S., Dalton, P., May, K. A., Boogert, S., & Mickes, L. (2022). Lineup identification in young and older witnesses: does describing the criminal help or hinder? *Cognitive Research: Principles and Implications*, *7*(1), 51. https://doi.org/10.1186/s41235-022-00399-1

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*(1), 94–108. https://doi.org/10.1037/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. https://doi.org/10.1037/h0093959

Houston Police Department Office of Planning & Data Governance. (2023). *Statistics. Monthy Operational Summary: May 2023*. Retrieved 29 Apr 2025 from https://www.houstontx.gov/police/department_reports/operational_summary/NIBRS_MonthlyOperationalSummary_May23.pdf

Innocence Project. (2024, 29 Dec 2024). *Exonerations Data*. Retrieved 6 Mar 2025 from https://innocenceproject.org/exonerations-data/

Juncu, S., & Fitzgerald, R. J. (2021). A meta-analysis of lineup size effects on eyewitness identification. *Psychology, Public Policy, and Law*, *27*(3), 295–315. https://doi.org/10.1037/law0000311

Karageorge, A., & Zajac, R. (2011). Exploring the effects of age and delay on children's person identifications: Verbal descriptions, lineup performance, and the influence of wildcards. *British Journal of Psychology*, *102*(2), 161–183. https://doi.org/10.1348/000712610X507902

Krafka, C., & Penrod, S. (1985). Reinstatement of context in a field experiment on eyewitness identification. *Journal of Personality and Social Psychology*, *49*(1), 58–69. https://doi.org/10.1037/0022-3514.49.1.58

Kroneisen, M., Bott, F. M., & Mayer, M. (2021). Remembering the bad ones: Does the source memory advantage for cheaters influence our later actions positively? *Quarterly journal of experimental psychology*, *74*(10), 1669-1685. https://doi.org/10.1177/17470218211007822

Küppers, V., & Bayen, U. J. (2014). Inconsistency effects in source memory and compensatory schema-consistent guessing. *Quarterly journal of experimental psychology*, *67*(10), 2042–2059. https://doi.org/10.1080/17470218.2014.904914

Lampinen, J. M., Race, B., Wolf, A. P., Phillips, P., Moriarty, N., & Smith, A. M. (2020). Comparing detailed and less detailed pre-lineup instructions. *Applied Cognitive Psychology*, *34*(2), 409–424. https://doi.org/10.1002/acp.3627

Lee, J., K., M. J., & and Penrod, S. D. (2022). Validity of mock-witness measures for assessing lineup fairness. *Psychology, Crime & Law*, *28*(3), 215-245. https://doi.org/10.1080/1068316X.2021.1905811

Lee, M. D., Bock, J. R., Cushman, I., & Shankle, W. R. (2020). An application of multinomial processing tree models and Bayesian methods to understanding memory impairment. *Journal of Mathematical Psychology*, *95*, 102328. https://doi.org/10.1016/j.jmp.2020.102328

Levett, L. M. (2013). Co-witness information influences whether a witness is likely to choose from a lineup. *Legal and Criminological Psychology*, *18*(1), 168-180. https://doi.org/10.1111/j.2044-8333.2011.02033.x

Levi, A. M., & Lindsay, R. C. L. (2001). Lineup and photo spread procedures: Issues concerning policy recommendations. *Psychology, Public Policy, and Law*, *7*(4), 776–790. https://doi.org/10.1037/1076-8971.7.4.776

Li, B., Hu, X., Shanks, D. R., Su, N., Zhao, W., Meng, L., Lei, W., Luo, L., & Yang, C. (2024). Confidence ratings increase response thresholds in decision making. *Psychonomic Bulletin & Review*, *31*(3), 1093-1102. https://doi.org/10.3758/s13423-023-02380-5

Lin, W., Strube, M. J., & Roediger III, H. L. (2019). The effects of repeated lineups and delay on eyewitness identification. *Cognitive Research: Principles and Implications*, *4*(1), 16. https://doi.org/10.1186/s41235-019-0168-1

Lindsay, R., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology*, *70*(3), 556-564. https://doi.org/10.1037/0021-9010.70.3.556

Lockamyeir, R. F., Carlson, C. A., Jones, A. R., Wooten, A. R., Carlson, M. A., & Hemby, J. A. (2021). One perpetrator, two perpetrators: The effect of multiple perpetrators on eyewitness identification. *Applied Cognitive Psychology*, *35*(5), 1206-1223. https://doi.org/10.1002/acp.3853

Lucas, C. A., Brewer, N., & Palmer, M. A. (2021). Eyewitness identification: The complex issue of suspect-filler similarity. *Psychology, Public Policy, and Law*, *27*(2), 151–169. https://doi.org/10.1037/law0000243

Malpass, R. S., & Devine, P. G. (1981). Eyewitness identification: Lineup instructions and the absence of the offender. *Journal of Applied Psychology*, *66*(4), 482–489. https://doi.org/10.1037/0021-9010.66.4.482

Manley, K. D., Chan, J. C. K., & Wells, G. L. (2022). Improving face identification of mask-wearing individuals. *Cognitive Research: Principles and Implications*, *7*(1), 27. https://doi.org/10.1186/s41235-022-00369-7

Mayer, C., Bell, R., Menne, N. M., & Buchner, A. (2024). Lineup position affects guessing-based selection but not culprit-presence detection in simultaneous and sequential lineups. *Scientific Reports*, *14*(1), 27642. https://doi.org/10.1038/s41598-024-78936-9

Mayer, C., Bell, R., Menne, N. M., Therre, A., Lichtenhagen, U., & Buchner, A. (in press). Evidence for age-related differences in culprit-presence detection and guessing-based selection in lineups. *Psychology and Aging*.

McKinley, G. L., & Peterson, D. J. (2022). Using objective measures to examine the effect of suspect-filler similarity on eyewitness identification performance. *Cognitive Research: Principles and Implications*, *7*(1), 1–9. https://doi.org/10.1186/s41235-022-00442-1

Meiser, T., Sattler, C., & Von Hecker, U. (2007). Metacognitive inferences in source memory judgements: The role of perceived differences in item recognition. *Quarterly journal of experimental psychology*, *60*(7), 1015–1040. https://doi.org/10.1080/17470210600875215

Meissner, C. A., & Brigham, J. C. (2001a). A meta-analysis of the verbal overshadowing effect in face identification. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *15*(6), 603-616. https://doi.org/10.1002/acp.728

Meissner, C. A., & Brigham, J. C. (2001b). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law*, *7*(1), 3-35. https://doi.org/10.1037//1076-8971.7.1.3

Memon, A., Havard, C., Clifford, B., Gabbert, F., & Watt, M. (2011). A field evaluation of the VIPER system: A new technique for eliciting eyewitness identification evidence. *Psychology, Crime & Law*, *17*(8), 711–729. https://doi.org/10.1080/10683160903524333

Memon, A., Hope, L., & Bull, R. (2003). Exposure duration: Effects on eyewitness accuracy and confidence. *British Journal of Psychology*, *94*(3), 339-354. https://doi.org/10.1348/000712603767876262

Menne, N., Bell, R., Winter, K., & Buchner, A. (2025). Disclosing the number of simultaneous lineups increases guessing-based selection in cases of multiple-culprit crimes. *Journal of Applied Research in Memory and Cognition*. https://doi.org/10.1037/mac0000217

Menne, N. M., Winter, K., Bell, R., & Buchner, A. (2022). A validation of the two-high threshold eyewitness identification model by reanalyzing published data. *Scientific Reports*, *12*(1), 13379. https://doi.org/10.1038/s41598-022-17400-y

Menne, N. M., Winter, K., Bell, R., & Buchner, A. (2023a). The effects of lineup size on the processes underlying eyewitness decisions. *Scientific Reports*, *13*(1), 17190. https://doi.org/10.1038/s41598-023-44003-y

Menne, N. M., Winter, K., Bell, R., & Buchner, A. (2023b). Measuring lineup fairness from eyewitness identification data using a multinomial processing tree model. *Scientific Reports*, *13*(1), 6290. https://doi.org/10.1038/s41598-023-33101-6

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*(4), 361–376. https://doi.org/10.1037/a0030609

Mickes, L., & Wixted, J. T. (2015). On the Applied Implications of the "Verbal Overshadowing Effect". *Perspectives on Psychological Science*, *10*(3), 400-403. https://doi.org/10.1177/1745691615576762

Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, *42*(1), 42–54. https://doi.org/10.3758/BRM.42.1.42

Nosworthy, G. J., & Lindsay, R. (1990). Does nominal lineup size matter? *Journal of Applied Psychology*, *75*(3), 358–361. https://doi.org/10.1037/0021-9010.75.3.358

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

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*(1), 55–71. https://doi.org/10.1037/a0031602

Penrod, S., & Cutler, B. (1995). Witness confidence and witness accuracy: Assessing their forensic relation. *Psychology, Public Policy, and Law*, *1*(4), 817-845. https://doi.org/10.1037/1076-8971.1.4.817

Pezdek, K. (2012). Fallible eyewitness memory and identification. In B. L. Cutler (Ed.), *Conviction of the innocent: Lessons from psychological research* (pp. 105–124). American Psychological Association. https://doi.org/10.1037/13085-005

Pike, G., Brace, N., & Kynan, S. (2002). *The visual identification of suspects: Procedures and practice. Home Office, A Publication of the Policing and Reducing Crime Unit Home Office Research, Development and Statistics Directorate.* Retrieved June 8 2024 from https://webarchive.nationalarchives.gov.uk/ukgwa/20130128103514/http://www.homeoffice.gov.uk/rds/pdfs2/brf202.pdf

Police Executive Research Forum. (2013). *A National Survey of Eyewitness Identification Procedures in Law Enforcement Agencies*. Retrieved 5 Mar 2024 from https://www.policeforum.org/assets/docs/Free_Online_Documents/Eyewitness_Identification/a%20national%20survey%20of%20eyewitness%20identification%20procedures%20in%20law%20enforcement%20agencies%202013.pdf

Pozzulo, J. D., Crescini, C., & Panton, T. (2008). Does methodology matter in eyewitness identification research?: The effect of live versus video exposure on eyewitness identification accuracy. *International Journal of Law and Psychiatry*, *31*(5), 430-437. https://doi.org/https://doi.org/10.1016/j.ijlp.2008.08.006

Pozzulo, J. D., & Lindsay, R. (1999). Elimination lineups: An improved identification procedure for child eyewitnesses. *Journal of Applied Psychology*, *84*(2), 167–176. https://doi.org/10.1037/0021-9010.84.2.167

Protzko, J., Lundmark, S., Walleczek, J., & Schooler, J. W. (2023). Nonlinear effect amplification: Differential susceptibility of verbal overshadowing as a function of time to interference. *Journal of Applied Research in Memory and Cognition*, *12*(4), 564-573. https://doi.org/10.1037/mac0000084

Radvansky, G. A., Doolen, A. C., Pettijohn, K. A., & Ritchey, M. (2022). A new look at memory retention and forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(11), 1698–1723. https://doi.org/10.1037/xlm0001110

Richler, J. J., Cheung, O. S., & Gauthier, I. (2011). Holistic Processing Predicts Face Recognition. *Psychological science*, *22*(4), 464-471. https://doi.org/10.1177/0956797611401753

Richler, J. J., & Gauthier, I. (2014). A meta-analysis and review of holistic face processing. *Psychological bulletin*, *140*(5), 1281–1302. https://doi.org/10.1037/a0037004

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review, 95*(3), 318-339. https://doi.org/10.1037/0033-295X.95.3.318

Rose, C., & Beck, V. (2016). Eyewitness accounts: false facts, false memories, and false identification. *Journal of Crime and Justice, 39*(2), 243-263. https://doi.org/10.1080/0735648X.2014.940999

Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review*, *103*(4), 734–760. https://doi.org/10.1037/0033-295X.103.4.734

Rubínová, E., Fitzgerald, R. J., Juncu, S., Ribbers, E., Hope, L., & Sauer, J. D. (2021). Live presentation for eyewitness identification is not superior to photo or video presentation. *Journal of Applied Research in Memory and Cognition, 10*(1), 167–176. https://doi.org/10.1016/j.jarmac.2020.08.009

Salt Lake City Police Department. (2022). *SLCPD Releases Response Time Data for January 2022*. Retrieved 29 Apr 2025 from https://police.slc.gov/2022/02/01/slcpd-releases-response-time-data-for-january-2022

Sauer, J., Brewer, N., Zweck, T., & Weber, N. (2010). The effect of retention interval on the confidence–accuracy relationship for eyewitness identification. *Law and Human Behavior*, *34*, 337–347. https://doi.org/10.1007/s10979-009-9192-x

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology*, *22*(1), 36-71. https://doi.org/https://doi.org/10.1016/0010-0285(90)90003-M

Shapiro, P. N., & Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological bulletin*, *100*(2), 139–156. https://doi.org/10.1037/0033-2909.100.2.139

Shepherd, J. (1983). Identification after long delays. *Evaluating witness evidence*, 173–187.

Smith, H. M. J., & Flowe, H. D. (2015). ROC Analysis of the Verbal Overshadowing Effect: Testing the Effect of Verbalisation on Memory Sensitivity. *Applied Cognitive Psychology*, *29*(2), 159-168. https://doi.org/https://doi.org/10.1002/acp.3096

Sporer, S. L., Tredoux, C. G., Vredeveldt, A., Kempen, K., & Nortje, A. (2020). Does exposure to facial composites damage eyewitness memory? A comprehensive review. *Applied Cognitive Psychology*, *34*(5), 1166-1179. https://doi.org/10.1002/acp.3705

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. https://doi.org/10.1023/A:1012888715007

Tredoux, C. G., Sporer, S. L., Vredeveldt, A., Kempen, K., & Nortje, A. (2021). Does constructing a facial composite affect eyewitness memory? A research synthesis and meta-analysis. *Journal of Experimental Criminology*, *17*(4), 713-741. https://doi.org/10.1007/s11292-020-09432-z

Tupper, N., Sauerland, M., Sauer, J. D., Broers, N. J., Charman, S. D., & Hope, L. (2018). Showup identification decisions for multiple perpetrator crimes: Testing for sequential dependencies. *PloS one*, *13*(12), e0208403. https://doi.org/10.1371/journal.pone.0208403

Wells, G. L. (2001). Police lineups: Data, theory, and policy. *Psychology, Public Policy, and Law*, *7*(4), 791–801. https://doi.org/10.1037/1076-8971.7.4.791

Wells, G. L. (2014). Eyewitness identification: Probative value, criterion shifts, and policy regarding the sequential lineup. *Current Directions in Psychological Science*, *23*(1), 11-16. https://doi.org/10.1177/0963721413504781

Wells, G. L., Kovera, M. B., Douglass, A. B., Brewer, N., Meissner, C. A., & Wixted, J. T. (2020). Policy and procedure recommendations for the collection and preservation of eyewitness identification evidence. *Law and Human Behavior*, *44*(1), 3–36. https://doi.org/10.1037/lhb0000359

Wells, G. L., & Lindsay, R. (1980). On estimating the diagnosticity of eyewitness nonidentifications. *Psychological bulletin*, *88*(3), 776-784. https://doi.org/10.1037/0033-2909.88.3.776

Wells, G. L., & Olson, E. A. (2003). Eyewitness testimony. *Annual review of Psychology*, *54*(1), 277-295. https://doi.org/10.1146/annurev.psych.54.101601.145028

Wells, G. L., Rydell, S. M., & Seelau, E. P. (1993). The selection of distractors for eyewitness lineups. *Journal of Applied Psychology*, *78*(5), 835-844. https://doi.org/10.1037/0021-9010.78.5.835

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

Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M., & Brimacombe, E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behavior*, *22*(6), 603-647. https://doi.org/10.1023/A:1025750605807

Wells, G. L., Smith, A. M., & Smalarz, L. (2015b). 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*(4), 324–328. https://doi.org/10.1016/j.jarmac.2015.08.010

Wells, G. L., Yang, Y., & Smalarz, L. (2015c). Eyewitness identification: Bayesian information gain, base-rate effect equivalency curves, and reasonable suspicion. *Law and Human Behavior*, *39*(2), 99–122. https://doi.org/10.1037/lhb0000125

Wetmore, S. A., Neuschatz, J. S., Gronlund, S. D., Wooten, A., Goodsell, C. A., & Carlson, C. A. (2015). Effect of retention interval on showup and lineup performance. *Journal of Applied Research in Memory and Cognition*, *4*(1), 8–14. https://doi.org/10.1016/j.jarmac.2014.07.003

Wickham, L. H. V., & Lander, K. (2008). The effect of verbal description and processing type on face identification. *European Journal of Cognitive Psychology*, *20*(3), 577-586. https://doi.org/10.1080/09541440701728433

Wilcock, R., & Bull, R. (2010). Novel lineup methods for improving the performance of older eyewitnesses. *Applied Cognitive Psychology*, *24*(5), 718–736. https://doi.org/10.1002/acp.1582

Wilson, B. M., Seale-Carlisle, T. M., & Mickes, L. (2018). The effects of verbal descriptions on performance in lineups and showups. *Journal of Experimental Psychology: General*, *147*(1), 113-124. https://doi.org/10.1037/xge0000354

Winter, K., Menne, N. M., Bell, R., & Buchner, A. (2022). Experimental validation of a multinomial processing tree model for analyzing eyewitness identification decisions. *Scientific Reports*, *12*(1), 15571. https://doi.org/10.1038/s41598-022-19513-w

Winter, K., Menne, N. M., Bell, R., & Buchner, A. (2023). Evaluating the impact of first-yes-counts instructions on eyewitness performance using the two-high threshold eyewitness identification model. *Scientific Reports*, *13*(1), 6572. https://doi.org/10.1038/s41598-023-33424-4

Wixted, J. T., & Ebbesen, E. B. (1991). On the form of forgetting. *Psychological science*, *2*(6), 409–415. https://doi.org/10.1111/j.1467-9280.1991.tb00175.x

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

Wooten, A. R., Carlson, C. A., Lockamyeir, R. F., Carlson, M. A., Jones, A. R., Dias, J. L., & Hemby, J. A. (2020). The number of fillers may not matter as long as they all match the description: The effect of simultaneous lineup size on eyewitness identification. *Applied Cognitive Psychology*, *34*(3), 590–604. https://doi.org/10.1002/acp.3644

Zajac, R., & Henderson, N. (2009). Don't it make my brown eyes blue: Co-witness misinformation about a target's appearance can impair target-absent line-up performance. *Memory*, *17*(3), 266-278. https://doi.org/10.1080/09658210802623950

Published Articles

Article 1

This article includes Experiments 1a and 1b.

Therre, A., Bell, R., Menne, N. M., Mayer, C., Lichtenhagen, U., & Buchner, A. (2024). On the possible advantages of combining small lineups with instructions that discourage guessing-based selection. *Scientific Reports*, *14*(1), 14126. https://doi.org/10.1038/s41598-024-64768-0

scientific reports

OPEN

On the possible advantages of combining small lineups with instructions that discourage guessing-based selection

Amelie Therre ✉, Raoul Bell, Nicola Marie Menne, Carolin Mayer, Ulla Lichtenhagen & Axel Buchner

The primary argument for including large numbers of known-to-be innocent fillers in lineups is that guessing-based selections are dispersed among a large number of lineup members, leading to low innocent-suspect identification rates. However, a recent study using the two-high threshold eyewitness identification model has demonstrated advantages of smaller lineups at the level of the processes underlying the observable responses. Participants were more likely to detect the presence of the culprit and less likely to select lineup members based on guessing in smaller than in larger lineups. Nonetheless, at the level of observable responses, the rate of innocent-suspect identifications was higher in smaller compared to larger lineups due to the decreased dispersion of guessing-based selections among the lineup members. To address this issue, we combined smaller lineups with lineup instructions insinuating that the culprit was unlikely to be in the lineup. The goal was to achieve a particularly low rate of guessing-based selections. These lineups were compared to larger lineups with neutral instructions. In two experiments, culprit-presence detection occurred with a higher probability in smaller compared to larger lineups. Furthermore, instructions insinuating that the culprit was unlikely to be in the lineup reduced guessing-based selection compared to neutral instructions. At the level of observable responses, the innocent-suspect identification rate did not differ between smaller lineups with low-culprit-probability instructions and larger lineups with neutral instructions. The rate of culprit identifications was higher in smaller lineups with low-culprit-probability instructions than in larger lineups with neutral instructions.

Keywords Police lineups, Eyewitness identification, Two-high threshold eyewitness identification model, Lineup size, Lineup instructions, Multinomial processing tree model

In criminal prosecution, the testimony of an eyewitness can be a valuable element to complement other evidence but also presents potential risks. False identifications by eyewitnesses have been determined to be a major reason of wrongful convictions, being involved in 63 % of DNA exoneration cases[1]. One potential problem is that the identification of a suspect is not only caused by the detection of the culprit but may also be caused by guessing-based selection. In fact, guessing-based selection occurs surprisingly frequently. For instance, in a field study with 1039 real lineups it has been found that "of all identifications made within this sample, approximately forty per cent were of fillers. These data add to a growing body of research showing that eyewitness guessing is not restricted to experimental situations with disinterested college students who know that their choices carry no consequences"[p. 264[2]]. A possible measure to reduce the rate with which an innocent suspect is identified from a lineup based on guessing is to include a large number of known-to-be-innocent fillers who share relevant characteristics with the suspect.

Provided that the lineup is fair, that is, provided that the suspect does not stand out from the fillers such that it is impossible to distinguish the suspect from the fillers without relying on memory for the culprit's appearance[3], a larger lineup has the advantage over a smaller lineup that guessing-based selection, if it occurs, is dispersed among a larger number of lineup members (including known-to-be-innocent fillers, therefore often described as "filler siphoning"[4,5]). Consequently, a larger lineup reduces the probability that guessing-based selection leads to the identification of the suspect among the fillers compared to a smaller lineup. Legal requirements for

Department of Experimental Psychology, Heinrich Heine University Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany. ✉email: Amelie.Therre@hhu.de

the number of fillers in lineups vary across jurisdictions. In addition to the suspect, lineups typically comprise five fillers in the United States[6], at least seven in Germany[7], eight in Great Britain[8] and nine in Canada[9] (for an overview, see[10]). Consequently, the probability of sampling the suspect among the lineup members in case of a guessing-based selection is 1 ÷ 6 in the United States, at most 1 ÷ 8 in Germany, 1 ÷ 9 in Great Britain and 1 ÷ 10 in Canada. If a lineup member is selected based on guessing, the smaller sampling probability associated with a larger number of fillers can be said to protect a suspect from being falsely identified[5,11,12]. This may be counted as an advantage of larger lineups over smaller lineups. However, it has been pointed out that larger lineups compared to smaller lineups are associated not only with a lower rate of innocent-suspect identifications, but also with a lower rate of culprit identifications—a cost–benefit-tradeoff[13]. Some researchers even found that larger lineups lead to a lower discriminability between culprits and innocent suspects compared to smaller lineups[14] while others found no effect of lineup size on discriminability[15,16] or raw identification rates[17].

These findings and considerations pertain to overall evaluations of eyewitness's observable responses, that is, the rates of innocent-suspect and culprit identifications. However, the effects of lineup size on innocent-suspect and culprit-identification rates may result from various latent detection-based and non-detection-based processes. Here the two-high threshold (2-HT) eyewitness identification model[3,18–21] is used to separately measure the latent processes underlying observable responses to lineups. The 2-HT eyewitness identification model (Fig. 1) has been successfully validated both in a series of dedicated validation experiments[20] and by reanalyzing published data obtained in various laboratories[3]. The 2-HT eyewitness identification model belongs to the class of multinomial processing tree models. Multinomial processing tree models are easy-to-use measurement models that have been applied to many domains in cognitive research[22–25]. Multinomial processing tree models imply that overt responses result from latent processes that occur with certain probabilities[25]. The probabilities with which these processes occur are represented by model parameters that can be compared statistically by means of easy-to-use software such as *multiTree*[26].

The 2-HT eyewitness identification model comprises all response categories that may occur when eyewitnesses try to identify a culprit from a lineup. In culprit-present lineups (see the rounded rectangle on the left side of the upper tree in Fig. 1), these response categories are culprit identifications, filler identifications and lineup rejections (see the rectangles on the right side of the upper tree in Fig. 1). In culprit-absent lineups (see the rounded rectangle on the left side of the lower tree in Fig. 1), these response categories are innocent-suspect identifications, filler identifications and lineup rejections (see the rectangles on the right side of the lower tree in Fig. 1).

In a culprit-present lineup (upper tree in Fig. 1), culprit-presence detection occurs with probability dP, as a result of which the culprit is identified. If the presence of the culprit is not detected, which occurs with probability $1 - dP$, the culprit can still be selected as a consequence of non-detection-based processes. Biased selection of the culprit may occur with probability b if the lineup is unfair, for example, if the culprit stands out from the other members of the lineup based on physical appearance or other distinct characteristics of the suspect's

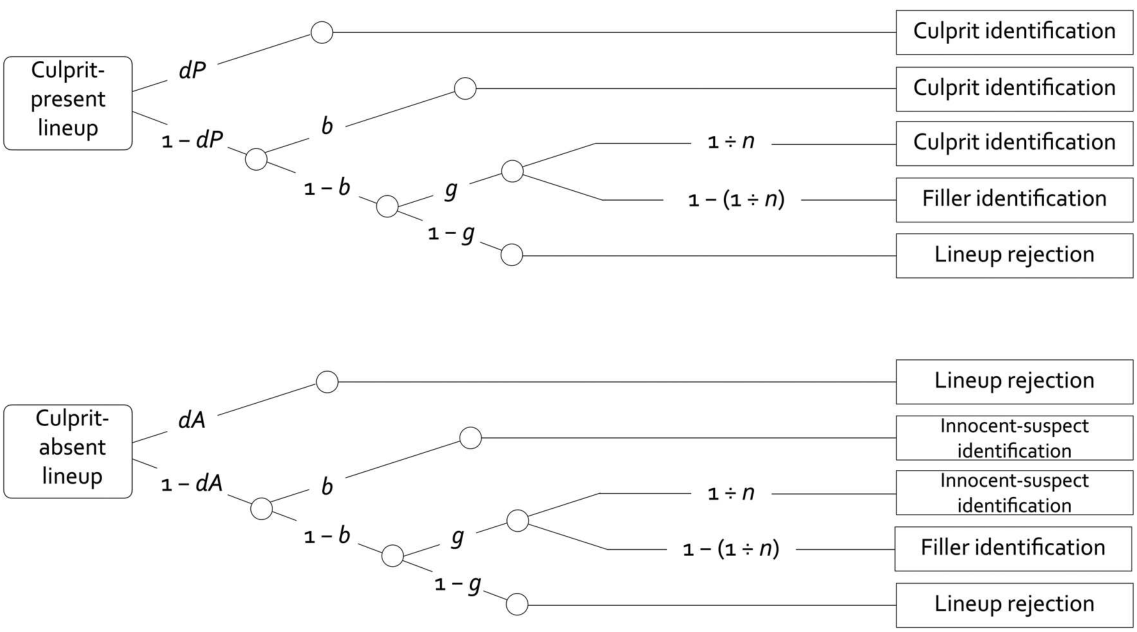

Figure 1. Rounded rectangles on the left represent the two types of possible lineups: culprit-present lineups and culprit-absent lineups. Rectangles to the right represent the observable response categories. Letters along the branches stand for the parameters that represent the latent processes (dP: probability of culprit-presence detection; dA: probability of culprit-absence detection; b: probability of biased suspect selection; g: probability of guessing-based selection). The constant n represents the number of persons in the lineup and $1 \div n$ represents the probability of sampling the culprit (upper tree) or the innocent suspect (lower tree) if guessing-based selection occurs.

photo. With the conditional probability $1-b$, no biased suspect selection occurs. In this case, the eyewitness may still select one of the lineup members based on guessing with the conditional probability g. The probability with which this process leads to the selection of the culprit is given by the sampling probability $1 \div n$, where n is a constant representing the lineup size. For instance, in a lineup with $n=6$ persons, guessing-based selection will lead to the selection of the culprit with a probability of $1 \div 6 \cong 0.17$. With the complementary probability of $1-(1 \div n)=5 \div 6 \cong 0.83$, guessing-based selection will lead to the selection of a filler. A culprit-present lineup is falsely rejected if neither culprit-presence detection $(1-dP)$ nor biased selection $(1-b)$ nor guessing-based selection $(1-g)$ occur.

In a culprit-absent lineup (lower tree in Fig. 1), parameter dA reflects the probability of detecting that the culprit is not in the lineup, resulting in a correct rejection of the lineup. With probability $1-dA$, culprit-absence detection does not occur. In this case, biased suspect selection occurs with the conditional probability b, resulting in a selection of the innocent suspect. In case of no biased suspect selection $(1-b)$, guessing-based selection occurs with the conditional probability g, in which case the probability of selecting the innocent suspect is given by the sampling probability $1 \div n$ while the probability of selecting a filler is given by $1-(1 \div n)$. The culprit-absent lineup is correctly rejected if neither culprit-absence detection $(1-dA)$ nor biased selection $(1-b)$ nor guessing-based selection $(1-g)$ occur.

When the 2-HT eyewitness identification model was used to evaluate the effects of lineup size on the latent processes underlying eyewitnesses' responses to lineups[19], interesting findings emerged on which we build here. Reducing the size of lineups (from six to three or from five to two persons) resulted in two remarkable outcomes: The probability of culprit-presence detection (represented by model parameter dP) was significantly higher and the probability of guessing-based selection (represented by model parameter g) was significantly lower in smaller compared to larger lineups. The model-based analysis thus revealed a favorable aspect of smaller compared to larger lineups at the level of the latent processes in that it seems preferable if observable eyewitness responses to lineups result from a detection process rather than from guessing-based processes.

However, an unfavorable outcome associated with smaller compared to larger lineups was obtained at the level of observable responses: The rate of innocent-suspect identifications was higher in smaller compared to larger lineups[19]. To see why this occurred, it is useful to look at the lower tree in Fig. 1. Here the model structure exposes the interplay of the latent processes behind the changes in observable responses in culprit-absent lineups when the lineup size n is reduced. For simplicity, let us assume that eyewitnesses do not detect the absence of the culprit $(dA=0)$ and that the lineup is perfectly fair $(b=0)$. In this case, the probability of the innocent suspect being identified as the culprit would be given by $g \times (1 \div n)$. Seen in isolation, a reduced probability of guessing-based selection (parameter g) in smaller compared to larger lineups should lead to a reduced rate with which innocent suspects are identified. However, if guessing-based selection occurs, the sampling probability that determines whether innocent suspects are identified is given by the term $1 \div n$, where n represents the lineup size. A decrease in lineup size n implies that the term $1 \div n$ increases. For instance, $1 \div n$ is doubled from 0.17 in six-person lineups to 0.33 in three-person lineups. The smaller probability of guessing-based selection (parameter g) in smaller compared to larger lineups did not compensate for this considerable increase in the probability of randomly sampling the innocent suspect $(1 \div n)$, leading to a net increase in the innocent-suspect identification rate in smaller compared to larger lineups. This is why, despite a reduction in parameter g, there was still a higher rate of innocent-suspect identifications in smaller compared to larger lineups[19].

While the higher rate of innocent-suspect identifications is a clear disadvantage of smaller compared to larger lineups, the structure of the model illustrated in Fig. 1 helps finding a potential remedy: At the level of the latent processes, the innocent-suspect identification rate is strongly affected by the probability of guessing-based selection reflected in parameter g. Therefore, the disadvantage of smaller compared to larger lineups at the level of the innocent-suspect identification rates should become lower if people are discouraged from making guessing-based selections. It seems even possible that the rate of innocent-suspect identifications in smaller lineups reaches the rate of innocent-suspect identifications in larger lineups or perhaps even a lower rate. For this to occur it would be necessary to reduce an eyewitness's tendency to make a guessing-based selection even below the level achieved by reducing the lineup size alone. Here it is useful that the probability of guessing-based selection can be manipulated without affecting the other processes specified in the 2-HT eyewitness identification model simply by applying appropriate lineup instructions[20]. Specifically, instructions insinuating that the culprit is unlikely to be in the lineup have been found to reduce parameter g without affecting the other model parameters[20]. If instructions insinuating that the culprit is unlikely to be in the lineup effectively reduce the probability of guessing-based selection in smaller lineups, then smaller lineups combined with such instructions might have no strong disadvantage in the innocent-suspect-identification rates compared to larger lineups without such instructions. This prediction was tested in the present experiments.

Apart from the unfavorable outcome of a higher rate of innocent-suspect identifications in smaller compared to larger lineups, there was also an unambiguously favorable outcome at the level of observable responses: The rate of culprit identifications was higher in smaller compared to larger lineups[19]. This is to be expected because the dominant determinant of a higher culprit identification rate in smaller compared to larger lineups is the higher probability of culprit-presence detection (parameter dP) in smaller compared to in larger lineups[19]. This is why we expected the rate of culprit identifications to be higher in smaller compared to larger lineups in the present experiments.

Parallel to the previous study[19] which we build on here, three-person lineups were compared to six-person lineups. The lineup size used in the latter condition corresponds to the typical lineup size in the United States[6]. Within each lineup-size condition, about half of the participants received instructions insinuating that the culprit was unlikely to be in the lineup (henceforth low-culprit-probability instructions), which are known to reduce guessing-based selection without affecting culprit-presence detection[20]. The other half of the participants received neutral instructions emphasizing that the culprit may or may not be in the lineup. Such neutral instructions are officially recommended for police lineups[12]. We consider the combination of six-person lineups with neutral instructions to be the standard of comparison. The combination of three-person lineups with low-culprit-probability instructions was compared to this standard in terms of innocent-suspect identification rates and culprit identification rates.

Before testing the novel predictions pertaining to the level of observable responses, it is important first to test whether the previously found effects of lineup size[19] and of low-culprit-probability instructions[20] are robust and can be replicated. In case of a successful replication the probability of culprit-presence detection (parameter dP) should be higher and the probability of guessing-based selection (parameter g) should be lower in three-person lineups than in six-person lineups. Also, low-culprit-probability instructions should lead to a lower probability of guessing-based selection (parameter g) than neutral instructions.

As mentioned above, the two novel predictions that were tested in the present study pertain to the level of observable responses. Both predictions were derived from the analysis of the underlying detection-based and non-detection-based processes as measured by the 2-HT eyewitness identification model. First, the hypothesis was tested that three-person lineups with low-culprit-probability instructions should be associated with a low rate of innocent-suspect identifications, ideally at least as low as the rate observed in six-person lineups with neutral instructions (the standard of comparison). Second, the hypothesis was tested that three-person lineups with low-culprit-probability instructions should be associated with a higher rate of culprit identifications compared to the standard of comparison, that is, the six-person lineups with neutral instructions. To test these hypotheses, we conducted two experiments. Given that both sequential and simultaneous lineups are used in jurisdictions around the world[10], we used both types of lineup formats. Specifically, sequential lineups were used in Experiment 1 and simultaneous lineups were used in Experiment 2, which served as a conceptual replication of Experiment 1.

Experiment 1

Method

Participants

Participants were recruited using the Horizoom research panel (www.horizoom-panel.de). Of the 1063 datasets of participants who had given their informed consent, 25 had to be excluded because participants had not passed the attention check (see below), 3 had to be excluded because of duplicate participation and 29 had to be excluded because participants had not completed the experiment or withdrew their consent. Consequently, datasets of 1006 participants were included in the analysis. Of these participants 528 identified as male, 475 as female and 3 as non-binary. Participants' ages ranged from 18 to 85 years ($M = 51$). Participants were randomly assigned to one of the four experimental groups. A total of 252 participants responded to three-person lineups with low-culprit-probability instructions, 255 participants responded to three-person lineups with neutral instructions, 255 participants responded to six-person lineups with low-culprit-probability instructions and 244 participants responded to six-person lineups with neutral instructions. A sensitivity analysis with G*Power[27] showed that given $N = 1006$ participants and four responses per participant, error probabilities of $\alpha = \beta = 0.05$ and $df = 1$ for tests of parameter equality across two groups, effects as small as $w = 0.06$ could be detected.

Ethics statement

The ethics committee of the Faculty of Mathematics and Natural Sciences at Heinrich Heine University Düsseldorf has granted ethical approval for the experiments reported here. The experiments were conducted in compliance with the Declaration of Helsinki. A requirement for taking part was the participant's declaration of informed consent prior to the experiment. Before the staged-crime video, participants were informed that they would see a video including physical and verbal violence. Participants were asked to continue the study only if they agreed to watch such a video.

Materials and procedure

Materials and procedure were the same as those used in earlier experiments[3,18–21]. The experiment was conducted online and was implemented in *SoSci Survey*[28] (www.socisurvey.de). Participation was possible with a desktop or laptop computer. Participants had to be 18 years old or older (a legal requirement in Germany).

After having given their informed consent, participants provided sociodemographic data. Subsequently, participants saw one of two staged-crime videos (henceforth Video 1 and Video 2). While the actors differed between videos, the events shown as well as their sequence and timing were the same in both videos. In essence, four men dressed in fan clothing of the German soccer club FC Bayern München were the culprits who physically and verbally abused a man dressed in fan clothing of a rival soccer club, Borussia Dortmund, at a bus stop. Actors portraying the same character were selected to be similar in body shape, hair color and hairstyle, that is, the actor portraying Character A in Video 1 was similar to the actor portraying Character A in Video 2, the actor portraying Character B in Video 1 was similar to the actor portraying Character B in Video 2 and so on. The videos were presented at a resolution of 885 × 500 pixels and lasted about 130 s.

The video was followed by an attention-check question requiring participants to indicate the role of the protagonists in the video (with soccer fans being the correct option amidst nine distractor options such as knights, musicians and politicians). Next, participants were informed that they had to identify the FC Bayern München hooligans in a series of photo lineups. All participants received the following instructions (the original was in German):

> "In the film, you saw Bayern München hooligans. Now we want you to identify them. To do this, we will show you several lineups. In each lineup, you will see a series of faces. You will be asked to indicate whether one of the people in the lineup is one of the Bayern München hooligans."

Participants in the groups with low-culprit-probability instructions also received the following instructions:

"It is **unlikely** that one of the culprits is in the lineup. Therefore, you should select the 'Yes, was present' button that belongs to the recognized face only if you are very certain that you have recognized the right person. Otherwise, click on the 'No, this person was not present' button."

In contrast, participants in the groups with neutral instructions received the following instructions:

"It is also possible that no one in the lineup is one of the Bayern München hooligans. If you recognize someone, click on the 'Yes, was present' button that belongs to the recognized face. Otherwise, click on the 'No, this person was not present' button."

Next, four sequential lineups were shown. Depending on the lineup size, the lineups included the facial photos of one suspect and either two or five fillers. The photos were presented one at a time. For each photo, participants had to decide whether or not it depicted one of the culprits by clicking either on a button labeled "Yes, was present" below the person's photo or on a button labeled "No, this person was not present". It was possible to choose more than one person in each lineup. As in prior studies[3,18–21], the last selection was considered to be a revision of any earlier selections and was used in the analyses.

In two lineups, a randomly selected culprit was present. In the other two lineups, an innocent suspect was present. The photos of the innocent suspects were photos of the actors from the video participants had not seen. For instance, if participants had seen Video 1 and the two randomly selected culprits from Video 1 were culprits portraying Characters B and C, then the culprits portraying Characters A and D from Video 2 were selected as innocent suspects in the culprit-absent lineups. This crossed-lineup procedure is identical to the one applied in prior studies[18–21] and ensures that the photos of culprits and innocent suspects (taken right after the videos had been shot) differ to the same degree from the photos of the fillers (taken from a face database[29] with the goal to resemble one of the culprits in body shape, hair color and hairstyle). This is parallel to the real world where the photos of the suspects (whose status of being innocent or guilty is unknown to the police) are often taken from a different source (e.g., from social media) than the photos of the fillers which are typically obtained from face databases. The order of the lineups and the positions of all photos in a lineup, including those of the culprit or innocent suspect, were randomly determined. For three-person lineups, random two-filler subsets were created from the set of five fillers used for the six-person lineups.

Participants in the groups with low-culprit-probability instructions were provided with the following reminder prior to each lineup: "It is unlikely that one of the culprits is in the lineup. Please choose someone only if you are very certain." Participants in the groups with neutral instructions did not receive a reminder.

After having responded to all lineups, participants were asked to reaffirm their consent to the use of their data, debriefed, thanked for their participation and redirected to the panel provider to receive their monetary compensation.

Results

The response frequencies obtained in Experiment 1 are presented in Table 1 (together with those of Experiment 2). The files with the raw frequency data and the equation files needed for the model-based analyses are available at https://osf.io/gcm8x/.

Lineup size	Instructions	Culprit-present lineups			Culprit-absent lineups		
		Culprit identifications	Filler identifications	Lineup rejections	Innocent-suspect identifications	Filler identifications	Lineup rejections
Experiment 1 – sequential lineups							
Three	Low-culprit-probability	212	92	200	73	124	307
	Neutral	260	122	128	131	166	213
Six	Low-culprit-probability	126	166	218	52	205	253
	Neutral	152	214	122	72	252	164
Experiment 2 – simultaneous lineups							
Three	Low-culprit-probability	217	62	227	78	91	337
	Neutral	255	93	170	94	140	284
Six	Low-culprit-probability	149	99	270	39	116	363
	Neutral	186	130	196	62	177	273

Table 1. Response frequencies in culprit-present and culprit-absent lineups for each combination of the lineup-size variable (three-person lineups vs. six-person lineups) and the instruction variable (low-culprit-probability instructions vs. neutral instructions) in Experiments 1 and 2.

Latent processes: Effects on parameters dP and g of the 2-HT eyewitness identification model

All model-based analyses were conducted using *multiTree*[26]. Four instances of the model illustrated in Fig. 1 were needed to analyze the data, one instance for each combination of the lineup-size variable (three-person lineups vs. six-person lineups) and the instruction variable (low-culprit-probability instructions vs. neutral instructions). The term $(1 \div n)$ which represents the probability of randomly sampling the suspect in case of guessing was set to 0.33333 for data obtained with three-person lineups (approximating $1 \div 3$) and 0.16667 for data obtained with six-person lineups (approximating $1 \div 6$).

To arrive at a testable base model, restrictions were applied to the 2-HT eyewitness identification model that were identical to those used in previous studies in which lineup size was manipulated[19]. Specifically, parameter b was set to be equal across all four instances of the model given that the lineups were composed of the same suspects and fillers in all groups such that there was no obvious reason as to why lineup fairness should differ between groups. Parameter dA was also set to be equal across all instances of the model because the manipulations used here were clearly different from those known to affect the probability of culprit-absence detection[3,20]. The base model with these restrictions fit the data, $G^2(6) = 8.04$, $p = 0.235$, providing support for the assumptions implemented in the base model. Parameters b and dA were estimated to be 0.04 ($SE = 0.01$) and 0.03 ($SE = 0.03$), respectively.

The estimates of parameters dP and g are presented in Table 2. One goal of the present research was to test whether the previously found effects of lineup size on culprit-presence detection and guessing-based selection can be replicated[19]. Table 2 shows that the probability of culprit-presence detection (parameter dP) was higher in three-person lineups than in six-person lineups. To test whether this difference was statistically significant, we imposed on the base model the additional restriction that parameter dP did not differ between three-person and six-person lineups, separately for the low-culprit-probability-instruction group and the neutral-instruction group. The decrease in model fit caused by this additional restriction compared to the fit of the base model was statistically significant for both the low-culprit-probability-instruction group, $\Delta G^2(1) = 21.04$, $p < 0.001$, and the neutral-instruction group, $\Delta G^2(1) = 20.12$, $p < 0.001$, implying that the equality restriction is incompatible with the data. This leads to the conclusion that the probability of culprit-presence detection is indeed higher in three-person lineups than in six-person lineups. Table 2 also shows that the probability of guessing-based selection (parameter g) was lower in three-person lineups than in six-person lineups. To test whether this difference was statistically significant, we imposed on the base model the additional restriction that parameter g did not differ between three-person and six-person lineups, separately for the low-culprit-probability-instruction group and the neutral-instruction group. The decrease in model fit caused by this additional restriction compared to the fit of the base model was statistically significant for both the low-culprit-probability-instruction group, $\Delta G^2(1) = 15.42$, $p < 0.001$, and the neutral-instruction group, $\Delta G^2(1) = 12.81$, $p < 0.001$, implying that the equality restriction is incompatible with the data. This leads to the conclusion that the probability of guessing-based selection is indeed lower in three-person lineups than in six-person lineups. It can thus be concluded that the previously found effects of smaller compared to larger lineups on the processes underlying eyewitnesses' responses to lineups[19] are robust and can be replicated.

Next, we tested whether low-culprit-probability instructions reduce the probability of guessing-based selection (parameter g) compared to neutral instructions. Table 2 shows that the estimate of parameter g is lower in the low-culprit-probability-instruction group than in the neutral-instruction group. To test whether this difference is statistically significant, we imposed on the base model the additional restriction that parameter g did not differ as a function of whether low-culprit-probability or neutral instructions were used, separately for the three-person-lineup group and the six-person-lineup group. The decrease in model fit caused by this additional restriction compared to the fit of the base model was statistically significant for both the three-person-lineup group, $\Delta G^2(1) = 52.24$, $p < 0.001$, and the six-person-lineup group, $\Delta G^2(1) = 55.75$, $p < 0.001$, implying that the equality restriction is incompatible with the data. This leads to the conclusion that low-culprit-probability instructions lead to a lower probability of guessing-based selection compared to neutral instructions. As an aside, manipulations aimed at changing the probability with which guessing-based selection occurs should not affect other model parameters such as parameter dP[3,20]. The restriction that parameter dP did not differ as a function of whether low-culprit-probability or neutral instructions were used did not lead to a statistically significant decrease in model fit compared to the fit of the base model for both the three-person-lineup group, $\Delta G^2(1) = 2.27$, $p = 0.132$, and the six-person-lineup group, $\Delta G^2(1) = 1.95$, $p = 0.162$, implying that the equality restriction is compatible with the data. This leads to the conclusion that culprit-presence detection does not differ as a function of the lineup instructions. It can thus be concluded that the previously found effects of low-probability instructions on the latent processes underlying eyewitnesses' responses to lineups[20] are robust and can be replicated.

Lineup Size	Instructions	Estimates of parameter *dP*		Estimates of parameter *g*	
Three	Low-culprit-probability	0.31	(0.03)	0.39	(0.02)
	Neutral	0.37	(0.03)	0.59	(0.02)
Six	Low-culprit-probability	0.15	(0.02)	0.49	(0.02)
	Neutral	0.19	(0.03)	0.68	(0.02)

Table 2. Parameter estimates of parameter dP (representing the probability of culprit-presence detection) and of parameter g (representing the probability of guessing-based selection) in Experiment 1. Values in parentheses are standard errors.

Observable responses: Effects on the rates of innocent-suspect identifications and culprit identifications

Given these successful replications we next tested whether combining three-person lineups with low-culprit-probability instructions would lead to a low rate of innocent-suspect identifications, ideally at least as low as the rate observed in six-person lineups with neutral instructions (the standard of comparison). The rates of innocent-suspect identifications in three-person lineups with low-culprit-probability instructions and six-person lineups with neutral instructions are presented in Fig. 2 (left side). Three-person lineups with low-culprit-probability instructions were associated with a rate of innocent-suspect identifications close to the rate observed in six-person lineups with neutral instructions. A two-proportion z-test showed that there was no statistically significant difference between these groups in the rate of innocent-suspect identifications in Experiment 1, $z = 0.12$, $p = 0.904$. This leads to the conclusion that the two rates are equal.

Finally, we tested whether three-person lineups combined with low-culprit-probability instructions would lead to higher rates of culprit identifications compared to the rates observed in six-person lineups with neutral instructions. The rates of culprit identifications are also presented in Fig. 2 (right side). Three-person lineups with low-culprit-probability instructions were associated with a higher rate of culprit identifications than six-person lineups with neutral instructions. A two-proportion z-test showed that the difference between these groups in the rate of culprit identifications was statistically significant, $z = 3.57$, $p < 0.001$. This leads to the conclusion that the rate of culprit identifications is higher for three-person lineups with low-culprit-probability instructions compared to six-person lineups with neutral instructions.

Discussion

The probability of culprit-presence detection was significantly higher and the probability of guessing-based selection was significantly lower in smaller compared to larger lineups. This replicates earlier findings[19]. In addition, low-culprit-probability instructions led to a significantly smaller probability of guessing-based selection than neutral instructions. In contrast, the type of instructions did not affect the probability of culprit-presence detection. This, too, replicates earlier findings[3,20]. It can be concluded that instructions implying that the culprit is unlikely to be in the lineup effectively discourage guessing-based selection.

Two novel predictions were derived about how small lineups combined with low-culprit-probability instructions should affect observable responses compared to six-person lineups with neutral instructions (the standard of comparison). First, combining three-person lineups with low-culprit-probability instructions should be associated with a low rate of innocent-suspect identifications, ideally at least as low as the rate observed in six-person lineups with neutral instructions. Second, three-person lineups with low-culprit-probability instructions should be associated with a higher rate of culprit identifications compared to six-person lineups with neutral instructions. The fact that both of these predictions were confirmed demonstrates that a deeper understanding of the latent processes underlying eyewitness responses can successfully lead to useful predictions about observable responses.

However, before drawing any firm conclusions it seemed important to test the robustness of these findings in a conceptual replication study, which was the purpose of Experiment 2. Given that not only sequential, but also simultaneous lineups are used in jurisdictions around the world[10], Experiment 2 was parallel to Experiment 1 except that simultaneous lineups were used. We expected to replicate the findings of Experiment 1 in Experiment 2.

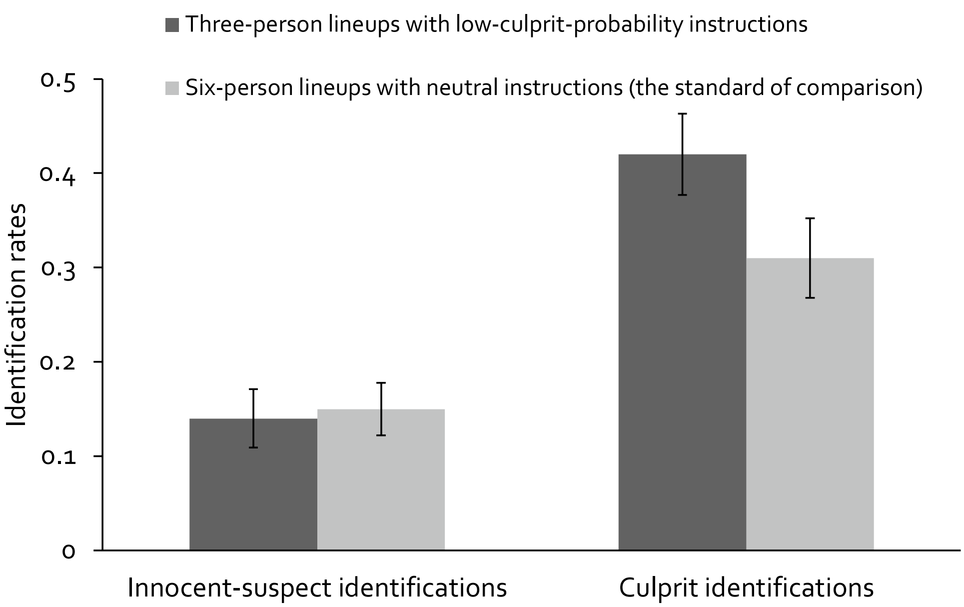

Figure 2. Rates of innocent-suspect identifications (left side) and rates of culprit identifications (right side) for three-person lineups with low-culprit-probability instructions and for six-person lineups with neutral instructions (the standard of comparison) in Experiment 1. Error bars represent the 95 % confidence intervals.

Experiment 2
Method
Participants
Participants were recruited using the Horizoom research panel (www.horizoom-panel.de). Of the 1105 datasets of participants who had given their informed consent, 21 had to be excluded because participants had not passed the attention check, 1 had to be excluded because of duplicate participation and 56 had to be excluded because participants had not completed the experiment or withdrew their consent. Consequently, datasets of 1027 participants, none of whom had participated in Experiment 1, were included in the analyses. Of these participants, 564 identified as male, 458 as female and 5 as non-binary. Participants' ages ranged from 18 to 84 years ($M = 49$). Participants were randomly assigned to one of the four experimental groups. A total of 253 participants responded to three-person lineups with low-culprit-probability instructions, 259 participants responded to three-person lineups with neutral instructions, 259 participants responded to six-person lineups with low-culprit-probability instructions and 256 participants responded to six-person lineups with neutral instructions. A sensitivity analysis with G*Power[27] showed that given $N = 1027$ participants and four responses per participant, error probabilities of $\alpha = \beta = 0.05$ and $df = 1$ for tests of parameter equality across two groups, effects as small as $w = 0.06$ could be detected.

Materials and procedure
Materials and procedure were identical to those of Experiment 1 except that simultaneous rather than sequential lineups were used. For each lineup, all six photos were shown next to each other at the same time. Participants could either select one person by clicking on a button labeled "Yes, was present" below the person's photo or reject the lineup by clicking on a button labeled "No, none of these persons was present".

Results
The response frequencies obtained in Experiment 2 are presented in Table 1 (together with those of Experiment 1). The files with the raw frequency data and the equation files needed for the model-based analyses are available at https://osf.io/gcm8x/.

Latent processes: Effects on parameters dP and g of the 2-HT eyewitness identification model
The term $(1 \div n)$ was set to 0.33333 and 0.16667 for data obtained with three-person lineups and six-person lineups, respectively. The same restrictions as in Experiment 1 were used to arrive at a base model which fit the data, $G^2(6) = 3.40$, $p = 0.757$. Parameters b and dA were estimated to be 0.05 ($SE = 0.01$) and 0.04 ($SE = 0.05$), respectively.

The estimates of parameters dP and g are presented in Table 3. The probability of culprit-presence detection (parameter dP) was higher in three-person lineups than in six-person lineups. Imposing on the base model the additional restriction that parameter dP did not differ between three-person and six-person lineups led to a significant decrease in model fit for both the low-culprit-probability-instruction group, $\Delta G^2(1) = 11.97$, $p < 0.001$, and the neutral-instruction group, $\Delta G^2(1) = 6.89$, $p = 0.009$, implying that the equality restriction is incompatible with the data. This leads to the conclusion that the probability of culprit-presence detection is higher in three-person lineups than in six-person lineups. The probability of guessing-based selection (parameter g) was not lower in three-person lineups than in six-person lineups. Imposing on the base model the additional restriction that parameter g did not differ between the three-person and six-person lineups did not lead to a significant decrease in model fit for both the low-culprit-probability-instruction group, $\Delta G^2(1) = 0.22$, $p = 0.637$, and the neutral-instruction group, $\Delta G^2(1) = 0.06$, $p = 0.815$, implying that the equality restriction is compatible with the data. In sum, then, the effects of smaller compared to larger lineups on the latent processes underlying eyewitnesses' responses to lineups reported in previous research[19] were replicated with respect to culprit-presence detection but not with respect to guessing-based selection. However, in those earlier results the difference in the estimates of parameter g between three-person lineups and six-person lineups was descriptively smaller for simultaneous lineups than for sequential lineups. From this pattern of findings, it seems possible to infer that the effect of lineup size on the probability of guessing-based selection may be relatively small and is therefore not reliably observed in simultaneous lineups. However, this is of course only a post-hoc speculation and the reasons as to why the effect of lineup size on the probability of guessing-based selection might be comparatively small in simultaneous lineups are currently unknown.

Next, we tested whether low-culprit-probability instructions reduce the probability of guessing-based selection (parameter g) compared to neutral instructions. Table 3 shows that the estimate of parameter g is lower

Lineup Size	Instructions	Estimates of parameter *dP*		Estimates of parameter *g*	
Three	Low-culprit-probability	0.33	(0.03)	0.30	(0.02)
	Neutral	0.37	(0.03)	0.45	(0.02)
Six	Low-culprit-probability	0.21	(0.02)	0.29	(0.02)
	Neutral	0.28	(0.03)	0.45	(0.02)

Table 3. Parameter estimates of parameter *dP* (representing the probability of culprit-presence detection) and of parameter *g* (representing the probability of guessing-based selection) in Experiment 2. Values in parentheses are standard errors.

in the low-culprit-probability-instruction group than in the neutral-instruction group. Imposing on the base model the additional restriction that parameter g did not differ as a function of whether low-culprit-probability or neutral instructions were used led to a significant decrease in model fit for both the three-person-lineup group, $\Delta G^2(1) = 28.81$, $p < 0.001$, and the six-person-lineup group, $\Delta G^2(1) = 42.21$, $p < 0.001$, implying that the equality restriction is incompatible with the data. This replicates the results of Experiment 1 and leads to the conclusion that low-culprit-probability instructions lead to a lower probability of guessing-based selection compared to neutral instructions. The restriction that parameter dP did not differ as a function of whether low-culprit-probability or neutral instructions were used did not lead to a statistically significant decrease in model fit compared to the fit of the base model for both the three-person-lineup group, $\Delta G^2(1) = 1.18$, $p = 0.278$, and the six-person-lineup group, $\Delta G^2(1) = 3.72$, $p = 0.054$, implying that the equality restriction is compatible with the data. This replicates the results of Experiment 1 and leads to the conclusion that culprit-presence detection does not differ as a function of the lineup instructions. It can thus be concluded that the previously found effects of low-probability instructions on the latent processes underlying eyewitnesses' responses to lineups[20] are robust and can be replicated.

Observable responses: Effects on the rates of innocent-suspect identifications and culprit identifications
The rates of innocent-suspect identifications in three-person lineups with low-culprit-probability instructions and six-person lineups with neutral instructions are presented in Fig. 3 (left side). Three-person lineups with low-culprit-probability instructions were associated with a rate of innocent-suspect identifications close to the rate observed in six-person lineups with neutral instructions. A two-proportion z-test showed that there was no statistically significant difference between these groups in the rate of innocent-suspect identifications, $z = 1.53$, $p = 0.126$. This replicates the results of Experiment 1 and leads to the conclusion that the two rates are equal.

The rates of culprit identifications are also presented in Fig. 3 (right side). Three-person lineups with low-culprit-probability instructions were associated with a higher rate of culprit identifications than six-person lineups with neutral instructions. A two-proportion z-test showed that the difference between these groups in the rate of culprit identifications was statistically significant, $z = 2.14$, $p = 0.032$. This replicates the results of Experiment 1 and leads to the conclusion that the rate of culprit identifications is higher for three-person lineups with low-culprit-probability instructions compared to six-person lineups with neutral instructions.

Discussion

The results of Experiment 2 largely replicate those of Experiment 1. With regard to the latent processes underlying eyewitness responses, the results confirm that the probability of culprit-presence detection is higher in smaller compared to larger lineups. In addition, low-culprit-probability instructions lead to a lower probability of guessing-based selection compared to neutral instructions.

With regard to observable responses, the results confirm the prediction that combining three-person lineups with low-culprit-probability instructions should be associated with a low rate of innocent-suspect identifications that does not differ from the rate observed in six-person lineups with neutral instructions. The results also confirm the prediction that the rate of culprit identifications should be higher in three-person lineups with low-culprit-probability instructions than in six-person lineups with neutral instructions.

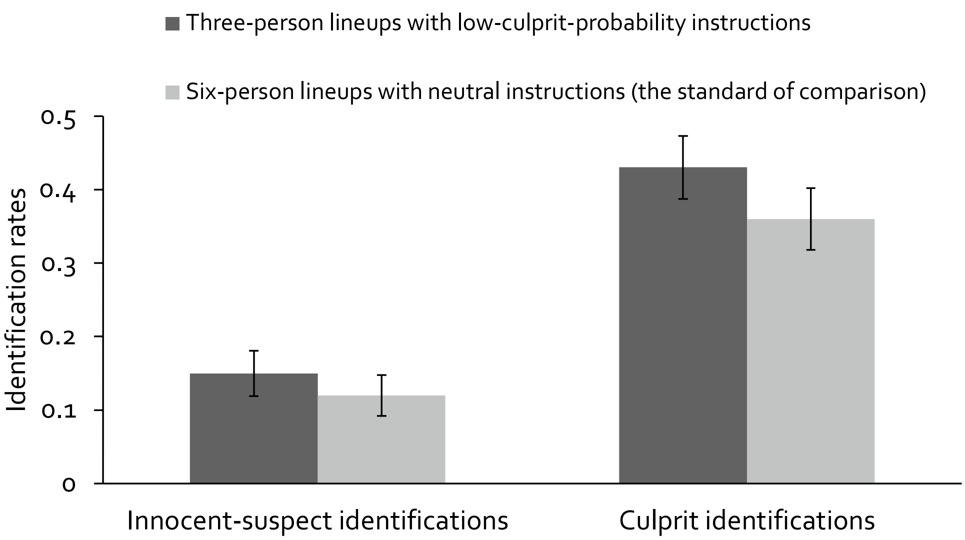

Figure 3. Rates of innocent-suspect identifications (left side) and rates of culprit identifications (right side) for the three-person lineups with low-culprit-probability instructions and for the six-person lineups with neutral instructions (the standard of comparison) in Experiment 2. Error bars represent the 95 % confidence intervals.

General discussion

Here we built on, and largely replicated, earlier findings showing that smaller lineups are associated with a higher probability of culprit-presence detection and a lower probability of guessing-based selection than larger lineups[19]. In the present experiments, the model-based analyses confirmed that the probability of culprit-presence detection is higher in smaller compared to larger lineups. This has to be counted as an advantage of smaller compared to larger lineups in that it seems desirable that lineup procedures support the detection of the culprit. The probability of guessing-based selection was lower in smaller compared to larger lineups in sequential (Experiment 1) but not in simultaneous (Experiment 2) lineups. Guessing-based selection leads to the identification of culprits and innocent suspects with a sampling probability of $1 \div n$ that is inversely related to the lineup size n. A useful aspect of the 2-HT eyewitness identification model (Fig. 1) is that it contains transparent assumptions about how these latent processes concur and bring about overt responses such as innocent-suspect identifications and culprit identifications. The insights gained based on the 2-HT eyewitness identification model and explicated in more detail in the introduction have laid open why a higher probability of culprit-presence detection in smaller compared to larger lineups may, among other factors, cause a higher rate of culprit identifications, and they also helped to clarify why, despite a lower probability of guessing-based selection in smaller compared to larger lineups, the inverse relationship between the sampling probability $1 \div n$ and the lineup size n caused observable rates of innocent-suspect identifications to be higher in smaller compared to larger lineups in previous experiments[19].

However, the 2-HT eyewitness identification model is a helpful tool not only to understand why these phenomena occur but also to identify a possible solution to the problem of the increased rate of innocent-suspect identifications in smaller compared to larger lineups. As explicated in the introduction, if it would be possible to markedly reduce parameter g even below the level achieved by reducing the lineup size, then the rate of innocent-suspect identifications in smaller lineups should be comparatively low, ideally at least as low as the rate of innocent-suspect identifications in larger lineups with neutral instructions. One possible measure to achieve this reduction in parameter g is to provide lineup instructions that discourage guessing-based selection, for instance by insinuating that the culprit is unlikely to be in the lineup[3,20,30]. At the same time, the rate of culprit identifications should remain higher in smaller than in larger lineups. The results of both experiments reported here confirm these predictions. The rate of innocent-suspect identifications was the same in smaller lineups with instructions that discourage guessing-based selection and in larger lineups with neutral instructions (the standard of comparison). At the same time, the rate of culprit identifications was higher in smaller lineups with low-culprit-probability instructions than in larger lineups with neutral instructions.

At a more abstract level, the present findings demonstrate the usefulness of the 2-HT eyewitness identification model not only for measuring the latent processes underlying eyewitnesses' responses to lineups[18,19,21] but also for gaining insights into how these processes may be affected with the goal of potentially improving the outcomes of lineup procedures. The present results show that these insights can be used to generate testable predictions about observable responses. These predictions were confirmed in the experimental tests reported here. Going beyond these specific experimental tests, it may seem tempting to assume that the outcomes of lineup procedures can generally be improved by combining a small lineup size with instructions that discourage guessing-based selection. Here we must sound a note of caution. Whereas it may indeed turn out to be possible to improve the outcomes of lineup procedures in that way, it is at this stage far from clear how well the findings reported here can be generalized. For instance, it is not yet clear what happens if the eyewitnesses' memory for the culprit is extremely poor or extremely good, if lineups are unfair, if lineup sizes differ from the ones investigated here (with showups clearly being out of the question[31]), if the instructions designed to discourage guessing-based selection are different from the ones used here and so on. These questions and many more will all have to be answered in future studies. The study reported here can thus only be the beginning of a larger research effort.

Data availability

The files with the frequency data and the equation files needed for the model-based analyses are available at https://osf.io/gcm8x/.

Received: 15 March 2024; Accepted: 12 June 2024
Published online: 19 June 2024

References

1. Innocence Project. *Exonerations Data*. https://innocenceproject.org/exonerations-data/ (2024). Accessed 12 Mar 2024.
2. Horry, R., Memon, A., Wright, D. B. & Milne, R. Predictors of eyewitness identification decisions from video lineups in England: A field study. *Law Hum. Behav.* **36**, 257–265. https://doi.org/10.1037/h0093959 (2012).
3. Menne, N. M., Winter, K., Bell, R. & Buchner, A. A validation of the two-high threshold eyewitness identification model by reanalyzing published data. *Sci. Rep.* **12**, 13379. https://doi.org/10.1038/s41598-022-17400-y (2022).
4. Wells, G. L., Smalarz, L. & Smith, A. M. ROC analysis of lineups does not measure underlying discriminability and has limited value. *J. Appl. Res. Mem. Cogn.* **4**, 313–317. https://doi.org/10.1016/j.jarmac.2015.08.008 (2015).
5. Wells, G. L. Police lineups: Data, theory, and policy. *Psychol. Public Policy Law* **7**, 791–801. https://doi.org/10.1037/1076-8971.7.4.791 (2001).
6. Police Executive Research Forum. *A National Survey of Eyewitness Identification Procedures in Law Enforcement Agencies*. https://www.policeforum.org/assets/docs/Free_Online_Documents/Eyewitness_Identification/a%20national%20survey%20of%20eyewitness%20identification%20procedures%20in%20law%20enforcement%20agencies%202013.pdf (2013). Accessed 5 Mar 2024.
7. German Federal Ministry of the Interior and Community. *Richtlinien für das Strafverfahren und das Bußgeldverfahren (RiStBV) [Guidelines for criminal proceedings and fine proceedings]*. https://www.verwaltungsvorschriften-im-internet.de/bsvwvbund_28032023_BMJRB3313104000060001.htm (2023). Accessed 2 Mar 2024.
8. Home Office. *Police and Criminal Evidence Act 1984 (PACE) Code D*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/903812/pace-code-d-2017.pdf (2017). Accessed 5 Mar 2024.

9. FPT: Federal/Provincial/Territorial Heads of Prosecutions. Subcommittee on the Prevention of Wrongful Convictions. *Innocence at stake: The need for continued vigilance to prevent wrongful convictions in Canada. Public Prosecution Service of Canada.* https://www.ppsc-sppc.gc.ca/eng/pub/is-ip/index.html (2018). Accessed 30 Dec 2023.
10. Fitzgerald, R. J., Rubínová, E. & Juncu, S. Eyewitness identification around the world. In *Methods, measures, and theories in eyewitness identification tasks* (eds. Smith, A. M., Toglia, M. & Lampinen, J. M.) 294–322. https://doi.org/10.4324/9781003138105-16 (Routledge/Taylor & Francis Group, 2021).
11. Levi, A. M. & Lindsay, R. C. L. Lineup and photo spread procedures: Issues concerning policy recommendations. *Psychol. Public Policy Law* **7**, 776–790. https://doi.org/10.1037/1076-8971.7.4.776 (2001).
12. Wells, G. L. *et al.* Policy and procedure recommendations for the collection and preservation of eyewitness identification evidence. *Law Hum. Behav.* **44**, 3–36. https://doi.org/10.1037/lhb0000359 (2020).
13. Juncu, S. & Fitzgerald, R. J. A meta-analysis of lineup size effects on eyewitness identification. *Psychol. Public Policy Law* **27**, 295–315. https://doi.org/10.1037/law0000311 (2021).
14. Meissner, C. A., Tredoux, C. G., Parker, J. F. & MacLin, O. H. Eyewitness decisions in simultaneous and sequential lineups: A dual-process signal detection theory analysis. *Mem. Cogn.* **33**, 783–792. https://doi.org/10.3758/BF03193074 (2005).
15. Akan, M., Robinson, M. M., Mickes, L., Wixted, J. T. & Benjamin, A. S. The effect of lineup size on eyewitness identification. *J. Appl. Psychol.* **27**, 369–392. https://doi.org/10.1037/xap0000340 (2021).
16. Wooten, A. R. *et al.* The number of fillers may not matter as long as they all match the description: the effect of simultaneous lineup size on eyewitness identification. *Appl. Cogn. Psychol.* **34**, 590–604. https://doi.org/10.1002/acp.3644 (2020).
17. Nosworthy, G. J. & Lindsay, R. Does nominal lineup size matter?. *J. Appl. Psychol.* **75**, 358–361. https://doi.org/10.1037/0021-9010.75.3.358 (1990).
18. Menne, N. M., Winter, K., Bell, R. & Buchner, A. Measuring lineup fairness from eyewitness identification data using a multinomial processing tree model. *Sci. Rep.* **13**, 6290. https://doi.org/10.1038/s41598-023-33101-6 (2023).
19. Menne, N. M., Winter, K., Bell, R. & Buchner, A. The effects of lineup size on the processes underlying eyewitness decisions. *Sci. Rep.* **13**, 17190. https://doi.org/10.1038/s41598-023-44003-y (2023).
20. Winter, K., Menne, N. M., Bell, R. & Buchner, A. Experimental validation of a multinomial processing tree model for analyzing eyewitness identification decisions. *Sci. Rep.* **12**, 15571. https://doi.org/10.1038/s41598-022-19513-w (2022).
21. Winter, K., Menne, N. M., Bell, R. & Buchner, A. Evaluating the impact of first-yes-counts instructions on eyewitness performance using the two-high threshold eyewitness identification model. *Sci. Rep.* **13**, 6572. https://doi.org/10.1038/s41598-023-33424-4 (2023).
22. Batchelder, W. H. & Riefer, D. M. The statistical analysis of a model for storage and retrieval processes in human memory. *Br. J. Math. Stat. Psychol.* **39**, 129–149. https://doi.org/10.1111/j.2044-8317.1986.tb00852.x (1986).
23. Erdfelder, E. *et al.* Multinomial processing tree models: A review of the literature. *Z. Psychol. / J. Psychol.* **217**, 108–124. https://doi.org/10.1027/0044-3409.217.3.108 (2009).
24. Schmidt, O., Erdfelder, E. & Heck, D. W. How to develop, test, and extend multinomial processing tree models: A tutorial. *Psychol. Methods.* https://doi.org/10.1037/met0000561 (2023).
25. Batchelder, W. H. & Riefer, D. M. Theoretical and empirical review of multinomial process tree modeling. *Psychon. Bull. Rev.* **6**, 57–86. https://doi.org/10.3758/BF03210812 (1999).
26. Moshagen, M. multiTree: A computer program for the analysis of multinomial processing tree models. *Behav. Res. Methods.* **42**, 42–54. https://doi.org/10.3758/BRM.42.1.42 (2010).
27. Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods.* **39**, 175–191. https://doi.org/10.3758/BF03193146 (2007).
28. Leiner, D. J. *SoSci Survey [Computer software].* https://www.soscisurvey.de (2022).
29. Minear, M. & Park, D. C. A lifespan database of adult facial stimuli. *Beh. Res. Meth. Instr. Comp.* **36**, 630–633. https://doi.org/10.3758/BF03206543 (2004).
30. Buchner, A., Erdfelder, E. & Vaterrodt-Plünnecke, B. Toward unbiased measurement of conscious and unconscious memory processes within the process dissociation framework. *J. Exp. Psychol. Gen.* **124**, 137–160. https://doi.org/10.1037/0096-3445.124.2.137 (1995).
31. Neuschatz, J. S.*, et al.* A comprehensive evaluation of showups. In *Advances in Psychology and Law: Volume 1* (eds. Miller, M. K. & Bornstein, B. H.) 43–69. https://doi.org/10.1007/978-3-319-29406-3_2 (Springer, 2016).

Author contributions
AT and NMM developed the experiment and implemented suggestions by RB, CM, UL and AB. AT collected the data and analyzed them with contributions of RB, NMM, CM, UL and AB. AT wrote the first draft of the manuscript. RB, NMM, CM, UL and AB revised the manuscript. All authors gave final approval for publication.

Funding
Open Access funding enabled and organized by Projekt DEAL. The work reported herein was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – BU 945/10–2, project number 456214986.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to A.T.

Articles Submitted for Publication

Article 2

This article includes Experiment 2.

Therre, A., Bell, R., Menne, N. M., Mayer, C., Lichtenhagen, U., & Buchner, A. (Under review). Delays reduce culprit-presence detection but do not affect guessing-based selection in response to lineups. Manuscript under review at *Scientific Reports*.

Delays reduce culprit-presence detection but do not affect guessing-based selection in response to lineups

Amelie Therre[1]*, Raoul Bell[1], Nicola Marie Menne[1], Carolin Mayer[1], Ulla Lichtenhagen[1] & Axel Buchner[1]

[1]Department of Experimental Psychology, Heinrich Heine University Düsseldorf, Germany

Author Note

Amelie Therre: https://orcid.org/0009-0002-2345-0218

Nicola Marie Menne: https://orcid.org/0000-0001-9421-8286

Raoul Bell: https://orcid.org/0000-0002-0592-0362

Carolin Mayer: https://orcid.org/0009-0003-5278-4361

Ulla Lichtenhagen: https://orcid.org/0009-0001-8243-4632

Axel Buchner: https://orcid.org/0000-0003-4529-3444

We have no known conflict of interest to disclose.

Correspondence concerning this article should be addressed to Amelie Therre, Department of Experimental Psychology, Heinrich Heine University Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany

E-Mail: Amelie.Therre@hhu.de

Reviewer-Link: https://osf.io/fqus6

Abstract

Police lineups are conducted with varying delays between the crime and the lineup. Crime-lineup delays may adversely affect the cognitive processes underlying eyewitness responses to lineups. In the present study we examined how these processes change across four crime-lineup delays. Participants viewed a staged-crime video and then completed simultaneous photo lineups after no delay or after a delay of one day, one week or one month. The results showed a significant decline in the probability of culprit-presence detection. The form of the decline is best described by a power function with the most rapid decline occurring at short crime-lineup delays. Eyewitnesses did not compensate the declining culprit-presence detection by increasing guessing-based selection, as demonstrated by the fact that the probability of guessing-based selection remained constant across crime-lineup delays. The findings underscore the critical importance of conducting lineups as soon as possible after a crime to maximize the probability of memory-based-culprit detection.

Keywords: police lineups, eyewitness identification, two-high threshold eyewitness identification model, delay, multinomial processing tree model

Responses made by eyewitnesses during police lineups can serve as important evidence in criminal prosecutions [1]. In a lineup, an eyewitness sees a single suspect (who may be guilty or innocent) along with a number of fillers who are known to be innocent. The ability of an eyewitness to detect the presence of the culprit in a given culprit-present lineup or to detect the absence of the culprit in a given culprit-absent lineup depends on the eyewitness' memory. One of the most well-established facts about memory is that it decreases with time [2-4]. Naturally, memory for crime-related details is no exception. For instance, the accuracy of eyewitness responses to crime-related questions has been found to decline with increasing delay [5-8]. This may also affect performance in a lineup given that participants' recall of facial characteristics has been reported to decline significantly after delays ranging from one week [9] to three weeks [10] and one month [11]. Although conflicting results have been reported [12], there is a general trend towards progressively worse accuracy of face memory as a function of an increasing delay [13-18]. Therefore, it may be a concern that lineups can occur with considerable delays due to factors outside of the control of investigators, such as the time needed to identify a suspect and the availability of eyewitnesses. Furthermore, there is good reason for conducting a lineup only after a thorough investigation because this increases the probability that the actual culprit and not an innocent suspect is in the lineup [19]. Delays can also result from resource limitations, as investigators must prioritize different tasks across multiple cases. Archive studies from Great Britain on real-world lineups show that delays between the crime and the lineup range from zero days to nine years [20], with the most frequently reported delays being one to three months [20-23]. Here we test how varying delays affect the cognitive processes underlying eyewitness responses to lineups.

Particularly relevant in this context are studies that address the question of how delay affects eyewitness's responses to lineups [24-35]. Most of these studies have relied on observable response rates, such as rates of correct culprit identifications, filler identifications and lineup rejections [27-32, 34, 35]. However, when the aim is to understand how delay affects the processes underlying eyewitness's responses to lineups, this approach does not yield clear conclusions because changes in the observable response rates may result from different underlying processes. For example, in some studies, the rate of correct culprit identifications from lineups have been found to decline with delays ranging from several minutes to eleven months [27, 28, 31]. In other studies, no differences were found in the culprit identification rates as a function of delay [29, 30, 32, 33]. Here it is unclear whether delay had indeed no effect on the culprit identification accuracy or whether the sensitivity of the statistical tests used in these studies was too low to detect such an effect. A third possibility is that the ability to detect the culprit's presence within the lineup declines, but participants compensate this decline by being more willing to select someone from the lineup based on guessing, resulting in no substantial change in culprit identification rates even after increasing delays. Indeed, in some studies, a descriptive increase in filler identifications in both culprit-present and culprit-absent lineups with increasing delay has been reported [24-26, 33]. For instance, in one study [33] there was no detrimental effect of a 48-hour delay on the rate of culprit identifications, but a descriptive increase in filler identifications in both fair culprit-present lineups (from 10 % to 15 %) and fair culprit-absent lineups (from 24 % to 45 %). This pattern of results raises the question about whether the ability to detect the culprit's presence was truly unaffected by the delay or whether participants perhaps compensated for a deficit in culprit-presence detection by being more willing to select someone based on guessing, the latter of which would be consistent with the observation that the rates of filler identifications were increased at a descriptive level. However, other studies found no increase in filler identification rates as a function of delay [28, 31, 32, 42]. Also, the effects of delay on innocent-suspect identifications are somewhat mixed. Whereas no such effects have been reported in studies with a designated innocent suspect [26, 33], relatively small effects have been reported in studies without a designated innocent suspect in which, therefore, the innocent-suspect-identification rate could not be computed directly but had to be substituted by a value determined by dividing the number of filler identifications in culprit-absent lineups by the lineup size [24, 25].

In addition to studies focusing on observable response rates, a few studies have used ROC analyses to examine the effects of delay on lineup performance [26, 33]. However, ROC analyses were not originally designed for analyzing lineup data. To fit lineup data into the binary format required by ROC analyses, filler identifications and lineup rejections in culprit-present lineups are treated as a single "false rejection" category and filler identifications and lineup rejections in culprit-absent lineups are treated as a single "correct rejection" category. This data reduction discards important distinctions among response types [39-41]. For instance, filler identifications in culprit-absent lineups are false responses, whereas rejections of culprit-absent lineups are correct responses, indicating that different processes underlie these responses. Moreover, ROC analyses yield only a single performance metric—the partial area under the curve—and do not allow for the measurement of multiple cognitive processes. When researchers aim to examine additional processes, such as response bias, they have to leave the measurement model on which ROC analyses are based. In the studies mentioned above, this led researchers to either revert to analyzing raw response rates [33] or use the response bias measure *c* of standard signal detection theory [26]. In each case, this entailed a shift to a different measurement model, grounded in assumptions that differ from those implied by the measurement model on which ROC analyses are based.

For the present study, one comprehensive model is required that is specifically designed to separately measure the processes underlying eyewitness responses to lineups while taking into account the full pattern of lineup-response categories. Ideally, such a model should be supported by validation studies demonstrating that it reliably captures the processes it was designed to measure. In addition, such a model should allow for a formal evaluation of model fit to the data.

Therefore, we used the two-high threshold (2-HT) eyewitness identification model [43-49] to examine the effects of delay on the cognitive processes underlying eyewitness responses to lineups. The 2-HT eyewitness identification model is illustrated in Figure 1. It offers several advantages. First, it allows for the assessment of four distinct cognitive processes within a single, unified framework. These processes, represented by the model's parameters, are culprit-presence detection, biased suspect selection, guessing-based selection and culprit-absence detection. Second, these cognitive processes are determined based on information from all six categories of observable eyewitness responses to both culprit-present and culprit-absent lineups. In culprit-present lineups, these categories are culprit identifications, filler identifications and lineup rejections; in culprit-absent lineups, they are innocent-suspect identifications, filler identifications and lineup rejections (see the rectangles on the right side of Figure 1). Third, the model has been thoroughly validated, showing that its parameters reliably and sensitively reflect the latent cognitive processes they were designed to measure. The model's validity has been demonstrated in a series of experimental studies using the same stimulus materials as those used in the present study [46], and through reanalyses of published data sets [43] from various researchers, laboratories and countries that used diverse lineup procedures, staged-crime videos and photographs of fillers and suspects [32, 33, 50-55].

The 2-HT eyewitness identification model belongs to the class of multinomial processing tree models—a class of straightforward and transparent measurement models for which comprehensive tutorials [56] and easy-to-use software [57] exist. Multinomial processing tree models are widely used in various domains of cognitive research [56, 58-60]. In these models, observable responses are conceived of as being determined by an interplay of latent cognitive processes that occur with certain probabilities [60]. As explicated previously [43-49], the probabilities of these processes occurring are represented by model parameters for which estimates can be determined and which can be statistically compared. Another advantage of the 2-HT eyewitness identification model is that the processes of culprit-presence detection, biased suspect selection, guessing-based selection and culprit-absence detection are precisely and transparently defined by the model's structure, as formalized in the model equations and illustrated in Figure 1. The verbal labels used for the parameters only serve as accessible everyday-language descriptors to simplify communication. The following sections use these labels to explicate the model equations.

In a culprit-present lineup (see the upper tree in Figure 1), the presence of the culprit is detected with probability dP, leading to a correct identification of the culprit. With probability $1 - dP$, the culprit's presence is not detected but the culprit can still be identified through non-detection-based processes referred to as biased selection and guessing-based selection. Biased selection occurs with the conditional probability b. Specifically, if the lineup is unfair and the culprit noticeably stands out based on physical appearance or distinct characteristics of the suspect's photo, biased suspect selection may occur with the conditional probability $b > 0$. In case of no biased suspect selection, which occurs with the conditional probability $1 - b$, a lineup member may still be selected based on guessing, which occurs with the conditional probability g. Conditional upon a guessing-based selection, the probability of the selection of the culprit is determined by the sampling probability $1 \div n$, with n being a constant representing the number of individuals (suspect and fillers) in the lineup. For example, in a lineup consisting of six individuals, the probability of selecting the culprit as a consequence of a guessing-based process is $1 \div 6 = 0.1\overline{6}$. The probability of selecting a filler as a consequence of a guessing-based process is given by the complementary probability $1 - (1 \div n)$. In six-person lineups, this probability is $5 \div 6 = 0.8\overline{3}$. In case of no guessing-based selection, which occurs with the conditional probability $1 - g$, the lineup is falsely rejected.

In a culprit-absent lineup (see the lower tree in Figure 1), the absence of the culprit is detected with probability *dA*, leading to a correct rejection of the lineup. With probability 1 − *dA,* culprit-absence detection does not occur. In this case, biased suspect selection may occur with the conditional probability *b*, resulting in the selection of the innocent suspect. In case of no biased suspect selection, which occurs with the conditional probability 1 − *b*, a lineup member may still be selected based on guessing, which occurs with the conditional probability *g*. The probability of this guessing-based process leading to the selection of the innocent suspect is determined by the sampling probability 1 ÷ *n*, while the probability of selecting a filler is 1 − (1 ÷ *n*). In case of no guessing-based selection, which occurs with the conditional probability 1 − *g*, the lineup is correctly rejected.

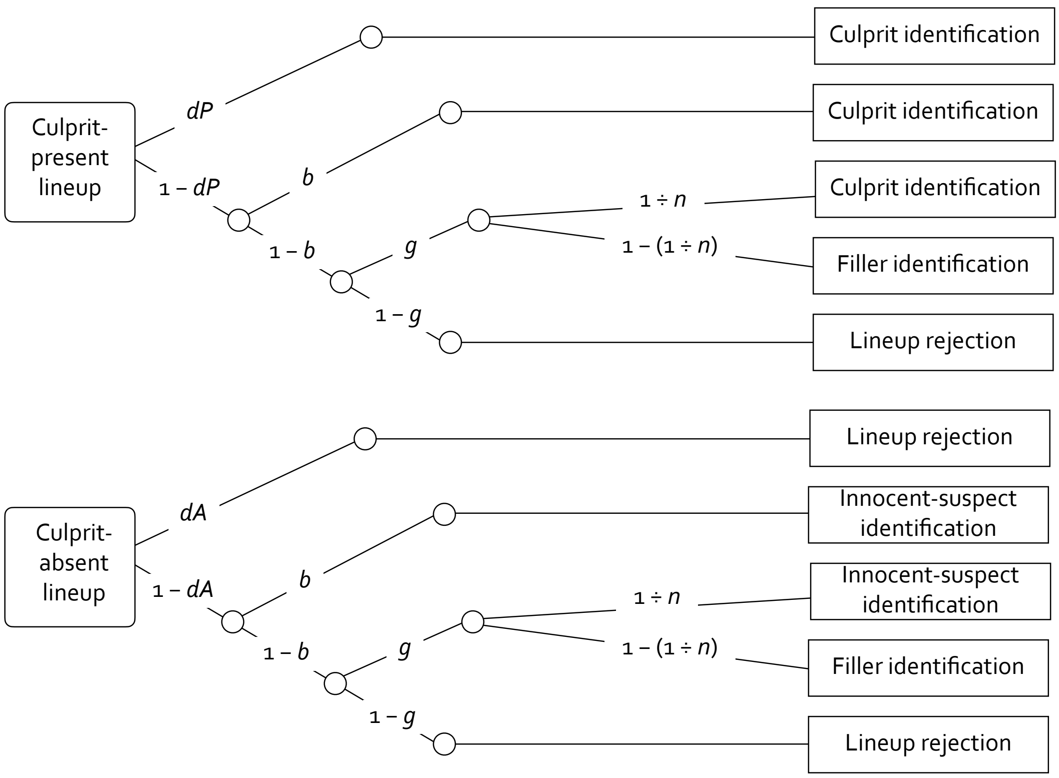

Figure 1. Graphical illustration of the 2-HT eyewitness identification model. Rounded rectangles on the left represent the two types of lineups an eyewitness may be confronted with: culprit-present lineups and culprit-absent lineups. The rectangles on the right represent the categories of observable responses. Letters along the branches denote parameters representing the latent processes specified by the model: *dP* represents the probability of culprit-presence detection, *b* represents the probability of biased suspect selection which occurs in unfair lineups, *g* represents the probability of guessing-based selection and *dA* represents the probability of

culprit-absence detection. The constant 1 ÷ *n* represents the probability of selecting the culprit (upper tree) or the innocent suspect (lower tree) if guessing-based selection occurs, with *n* corresponding to the number of individuals in the lineup.

To date, the effects of delay on culprit-presence detection, guessing-based selection and culprit-absence detection have not been examined directly, which was therefore the aim of the present study. Furthermore, with a few notable exceptions [29, 31, 35], manipulations of the delay variable in previous studies in which lineups were presented were typically limited to only two points in time: one condition with no or a small delay and one condition with a larger delay [24-26, 30, 32-34]. This binary approach, while certainly informative, does not capture the form of the changes occurring over time. We therefore decided to examine the effects of delay across four points in time with the goal to provide a more comprehensive understanding of how the cognitive processes underlying eyewitness responses to lineups change as a function of delay. Specifically, we investigated changes in culprit-presence detection, guessing-based selection and culprit-absence detection as defined within the 2-HT eyewitness identification model with no delay, a delay of one day, a delay of one week and a delay of one month between viewing a staged-crime video and responding to the lineups.

The first prediction about the effects of these delays refers to the memory-based process of culprit-presence detection (parameter *dP*). As memory is susceptible to forgetting [4], we expected parameter *dP* to decline as a function of delay in the form of a typical forgetting function [16]. In contrast, two different predictions as to how guessing-based selection (parameter *g*) changes as a function of delay were possible based on prior research. The inference for deriving one of the predictions begins by noting that guessing-based selection is known to be very sensitive to the probability with which eyewitnesses expect the culprit to be in the lineup based on the instructions they receive [46, 48]. Theses instructions were the same for all delay conditions. Thus, one possible prediction was that guessing-based selection stays constant across delays. Alternatively, the possibility exists that given a decline in culprit-presence detection as a function of delay, participants might engage in compensatory guessing [61-64]. If this were the case, then the probability of guessing-based selection would increase as a function of increasing delays. Finally, the prediction about delay-induced changes in the detection of absent culprits (parameter *dA*) was not as straightforward. This is so because, on the one hand, the memory-based process of culprit-absence detection should become less likely with increasing delay, just like the memory-based process of culprit-presence detection. On the other hand, the typical estimates of parameter *dA* under no-delay conditions are already quite low [44, 47, 48, 65] because culprit-absence detection is an inherently demanding cognitive process. While culprit-presence detection requires just one lineup member (the culprit) to elicit culprit-presence detection, culprit-absence detection requires the eyewitness to rule out each lineup member as the culprit. From this perspective it seemed unrealistic to expect to observe further reductions in the probability with which culprit-absence detection occurs.

Method

Participants

Participants were recruited via the Horizoom research panel (www.horizoom-panel.de), a panel certified under ISO 20252 which ensures rigorous quality control. All participants were first exposed to the staged-crime video and were then randomly assigned to one of four groups defined by the duration of the delay after which the participants were invited to participate in the second phase of the experiment. Of the 3108 data sets of participants who had given their informed consent before being exposed to a staged-crime video, 22 had to be excluded because participants had not passed the attention check (see below), 31 had to be excluded because of duplicate participation and 245 had to be excluded because participants had not completed the experiment or had withdrawn their consent after having completed the first phase of the experiment. Therefore, valid data sets of 2810 participants were available after the first phase of the experiment in which participants had been exposed to the staged-crime video. Of these participants, 550 were assigned to the no-delay condition and were asked to respond to the lineups right after having seen the staged-crime video, 788 participants were assigned to the one-day-delay condition, 706 participants were assigned to the one-week-delay condition and 766 participants were assigned to the one-month-delay condition. More participants were assigned to the with-delay conditions than to the no-delay condition in an attempt to compensate for anticipated dropouts.

Responses to lineups were collected from 550 participants in the no-delay condition, from 532 participants in the one-day-delay condition, from 520 participants in the one-week-delay condition and from 506 participants in the one-month-delay condition, resulting in a total of

2108 data sets that were analyzed. A sensitivity analysis with G*Power [66] showed that given $N = 2108$ participants and four responses per participant, error probabilities of $\alpha = \beta = 0.05$ (and thus a power of 0.95) and $df = 3$ for tests of parameter equality across the four delay conditions, effects of delay as small as $w = 0.05$ could be detected.

The groups were compared with respect to the demographic data that we had collected to assess whether there was any indication that the dropout was selective. Mean age and age range, gender and educational level are reported in Table 1. Neither age, $F(3, 2104) = 0.59$, $p = .619$, nor gender distribution, $\chi^2(6) = 5.23$, $p = .515$, nor educational level, $\chi^2(3) = 4.42$, $p = .220$ differed significantly among groups. Thus, even though the sample size and, hence, the sensitivity of the statistical tests of differences among groups was rather high, there was no evidence that the dropout was selective.

Delay	Mean age (standard deviation)	Gender	Educational level
No delay ($n = 550$)	52 years (15 years)	45 % ♀, 55 % ♂, < 1 % non-binary	55 % A-Levels or higher
One day ($n = 532$)	51 years (14 years)	44 % ♀, 55 % ♂, 1 % non-binary	59 % A-Levels or higher
One week ($n = 520$)	51 years (14 years)	43 % ♀, 57 % ♂	60 % A-Levels or higher
One month ($n = 506$)	51 years (14 years)	41 % ♀, 58 % ♂, 1 % non-binary	55 % A-Levels or higher
All ($n = 2108$)	51 years (14 years)	43 % ♀, 56 % ♂, < 1 % non-binary	57 % A-Levels or higher

Table 1. Mean age (standard deviations in parentheses), gender and educational level by delay. A-Levels include International Baccalaureate (IB) or equivalent qualifications.

Ethics statement

The ethics committee of the Faculty of Mathematics and Natural Sciences at Heinrich Heine University Düsseldorf approved the experiment. The experiment was conducted in accordance with the Declaration of Helsinki. Participants provided informed consent before participating. In the consent form and prior to viewing the staged-crime video, participants were informed that the video would contain physical and verbal violence. They were instructed to proceed with the study only if they were comfortable watching such content.

Materials and procedure

Materials and procedure were the same as those used in a number of previous studies [43-49, 67] except for the manipulation of the delay between viewing the staged-crime video and responding to the lineups. The experiment was conducted online using *SoSci Survey* [68] (www.soscisurvey.de). Participation was possible with a desktop or a laptop computer. Participants were informed that the experiment consisted of two parts, the first of which they would complete directly. In addition, participants were informed that the experiment would include a video portraying physical and verbal violence. They were advised that participation required their consent to view such a video and to the use of their data. Next, participants provided their age, gender and educational level. Participants could participate only if they were least 18 years old (a legal requirement in Germany).

After having been instructed to start the video by clicking on a "Start" button, each participant watched one of two staged-crime videos (referred to as Video 1 and Video 2). The videos were presented at a resolution of 885 × 500 pixels and lasted approximately 130 seconds. The videos depicted the same events in the same sequence and timing but the actors differed between the videos. However, the actors playing the same characters in both videos were chosen to be similar in body shape, hair color and hairstyle. For instance, the actor playing Character A in Video 1 resembled the actor playing Character A in Video 2. The same applied to Characters B, C and D. Both videos featured four men dressed in FC Bayern München soccer club fan clothing who physically and verbally assaulted a man in Borussia Dortmund fan clothing at a bus stop. All culprits were involved in the crime to a similar extent.

By including four culprits, we were able to obtain four data points per participant, thereby increasing the statistical sensitivity of our analyses while also maintaining ecological validity given that more than one third of real-world crimes have been reported to involve multiple culprits [69]. Both in the real world and in study settings, responding to multiple lineups after having witnessed a multiple-culprit crime may be more cognitively demanding than responding to a single lineup after having witnessed a single-culprit crime [70].

Following the video, participants answered an attention-check question in which they had to identify the roles of the protagonists in the video. The correct response was to select "soccer fans" among nine distractor options such as "dancers", "farmers" or "artists". Providing a correct response to the attention-check question was a prerequisite for participation in the second part of the experiment.

Participants assigned to the no-delay condition were then informed that they were about to enter the second part of the experiment. Participants assigned to one of the with-delay conditions were instead informed that they would receive an email inviting them to participate in the second part of the experiment after a delay of 24 hours (one-day-delay condition), seven days (one-week-delay condition) or 30 days (one-month-delay condition). Participants in the with-delay conditions did not always complete the second part of the experiment after the nominal delay, that is, on the same day at which the invitation email had been sent. The average actual delay was therefore somewhat larger than the nominal delay. The average actual delay was 1 day (standard deviation < 1 day) in the one-day-delay condition, 8 days (standard deviation = 1 day) in the one-week-delay condition and 33 days (standard deviation = 3 days) in the one-month-delay condition.

In the second part of the experiment, participants were instructed to identify the FC Bayern München fans—seen in the video during the first part of the experiment—from a series of photo lineups. The following instructions were given (the original instructions were in German):

"In the first part of the experiment, you saw a film with Bayern München fans. Now we want you to identify them. To do this, we will show you several lineups. In each lineup, you will see a series of faces. You will be asked to indicate whether one of the people in the lineup is one of the Bayern München fans. It is also possible that no one in the lineup is one of the Bayern München fans. If you recognize someone, click on the 'Yes, was present' button that belongs to the recognized face. Otherwise, click on the 'No, none of these persons was present' button."

Afterwards, each participant was presented with four separate lineups, one for each of the Bayern München fans from the video. In each lineup, one suspect and five fillers were displayed simultaneously in a single row. This presentation format is a possible method for photographic lineups [71-75] and was chosen for several reasons. First, it resembles the arrangement used in in-person lineups, which remain part of the pertinent guidelines in various jurisdictions [76, 77]. Second, it has been reported that, within these guidelines, "52% described an identification procedure that suggested lineup members would be presented simultaneously (e.g., they would appear in a *line*)", (emphasis added, [76p. 302]). Based on this, we considered the single-row photographic format to be a reasonable choice for the present study.

As in earlier studies [43-49, 67], the crossed-lineup procedure was used. Two of the four lineups were culprit-present lineups and two were culprit-absent lineups. The two culprits for the culprit-present lineups were selected randomly without replacement from the four culprits of the video the participant had seen. The innocent suspects in the culprit-absent lineups were culprits from the parallel video that the participant had not seen. For example, if Characters C and D from Video 1 had randomly been selected as culprits in the culprit-present lineups, then Characters A and B from Video 2 were selected as innocent suspects in the culprit-absent lineups. This crossed-lineup procedure ensures that the photos of the culprits and innocent suspects (taken right after the videos had been recorded) differ to the same degree from the photos of the fillers which were taken from a face database [78] and resembled one of the culprits in body shape, hair color, and hairstyle. This setup is analogous to real-world situations where photos of suspects (whose guilt or innocence is unknown to the police) often come from a different source, such as social media, than the filler photos which are typically taken from a face database. The crossed-lineup procedure is similar, but not identical, to the single-lineup procedure proposed by Oriet and Fitzgerald [79]. In the single-lineup procedure, a single lineup is shown to all participants and the suspect's guilt is determined by which of two videos participants viewed (either featuring the suspect from the lineup or a similar-looking person). By contrast, in the crossed-lineup procedure used here, it is randomly determined for each participant whether a suspect appears in a culprit-present or culprit-absent lineup. This randomization ensures that each suspect has an equal likelihood of being presented as a culprit or as an innocent suspect. This makes the crossed-lineup procedure particularly suitable for cases involving multiple culprits, as it allows for variation across lineups in whether a culprit or an innocent suspect is presented.

All individuals were shown from a frontal view with neutral facial expressions against a black background with no visible clothing. The photographs were adjusted to maintain consistent face sizes and lighting conditions and were displayed at a resolution of 142 × 214 pixels. The positions of the photos in the lineups were determined randomly, as was the sequence of the lineups. Once participants had responded to all lineups, they were asked to reaffirm, or to withdraw, their consent to the use of their data. They were then debriefed, thanked for their participation and redirected to the panel provider to receive their monetary compensation.

Results

The response frequencies obtained in this experiment are presented in Table 2. The files with the raw frequency data and the equations needed for the model-based analyses are available at https://osf.io/fqus6.

Delay	Culprit-present lineups			Culprit-absent lineups		
	Culprit identifications	Filler identifications	Lineup rejections	Innocent-suspect identifications	Filler identifications	Lineup rejections
No delay	401 (37 %)	276 (25 %)	423 (38 %)	133 (12 %)	383 (35 %)	584 (53 %)
One day	265 (25 %)	355 (33 %)	444 (42 %)	137 (13 %)	397 (37 %)	530 (50 %)
One week	188 (18 %)	359 (35 %)	493 (47 %)	117 (11 %)	396 (38 %)	527 (51 %)
One month	134 (13 %)	347 (34 %)	531 (52 %)	109 (11 %)	370 (37 %)	533 (53 %)

Table 2. Response frequencies (and proportions, relative to the condition-specific response frequencies in culprit-present and culprit-absence lineups, respectively, in parentheses) as a function of culprit presence or absence and delay.

All model-based analyses were conducted using *multiTree* [57]. Four instances of the model illustrated in Figure 1 were needed to analyze the data, one instance for each delay condition (no delay, one day, one week, one month). To generate a testable base model, restrictions were applied to the 2-HT eyewitness identification model. As six-person lineups were presented in the current experiment, the term $1 \div n$ which represents the sampling probability of the suspect in case of guessing-based selection was set to 0.16667 for all conditions. Given that the lineups consisted of the same suspects and fillers in all conditions, there was no reason to expect differences in lineup fairness among conditions. Consequently, parameter b was set to be equal for all conditions. The base model incorporating these restrictions fitted the data, $G^2(3) = 2.87$, $p = .412$, supporting the conclusion that lineup fairness, represented by parameter b, did not vary as a function of delay. This implies that the lineups were equally fair across all delays. Parameter b was estimated to be 0.04 (95 % CI [0.03, 0.06]), reflecting a slight inherent unfairness in the lineups across all delays. By taking biased selection due to lineup unfairness into account explicitly in the 2-HT eyewitness identification model, two important goals are achieved. First, the model provides for a direct measure of lineup fairness that is more valid than measures based on the traditional mock-witness task because the cognitive processes of mock witnesses and eyewitnesses differ fundamentally [44]. Second, the model also ensures that the measurements of the other model parameters remain uncontaminated by lineup unfairness, thereby allowing for valid insights into the cognitive processes represented by these model parameters even in unfair lineups [43, 44, 46].

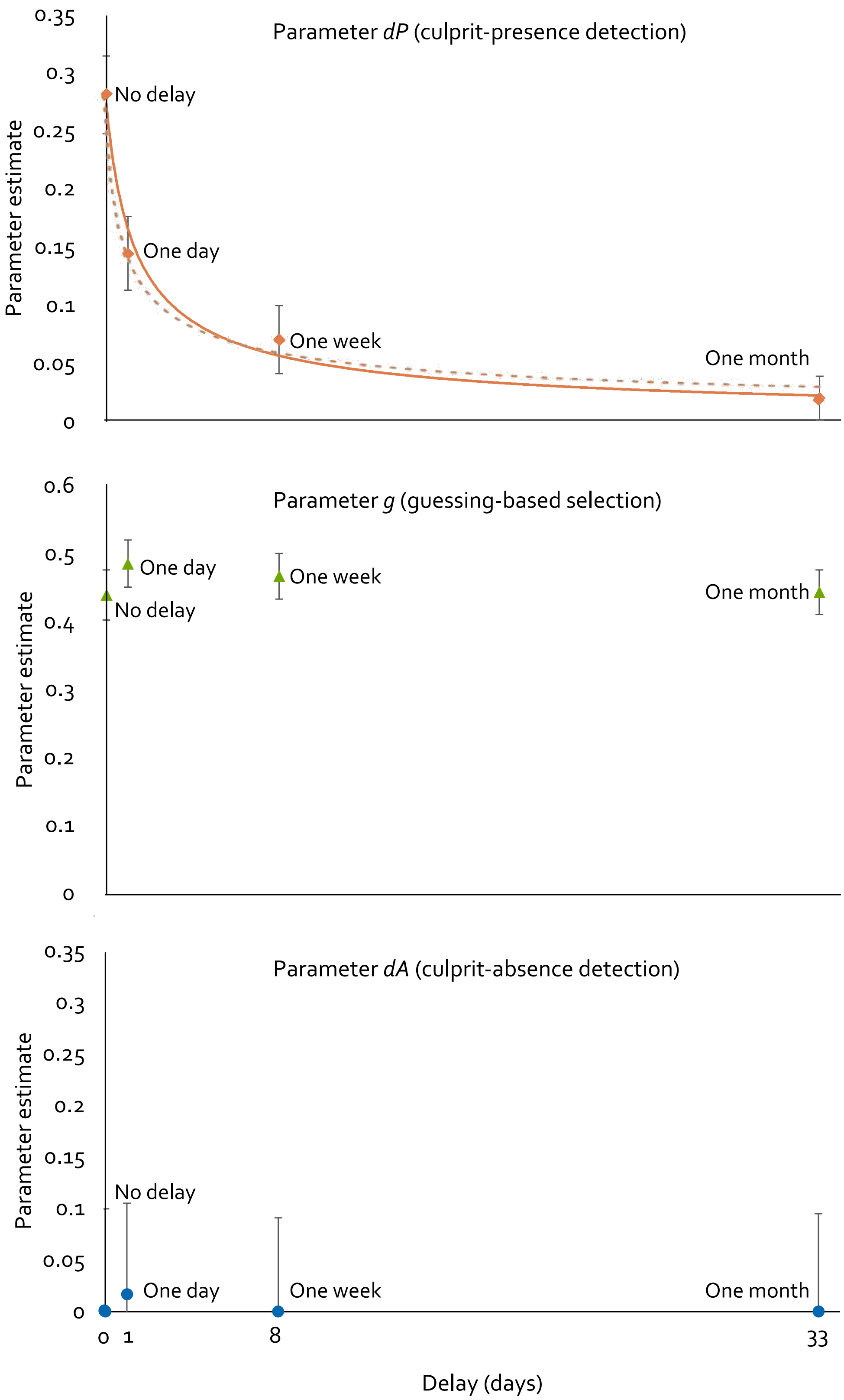

Figure 2. 2-HT eyewitness identification parameter estimates as a function of the average actual delay. For the no-delay condition and the one-day-delay condition, the nominal delay was equal to the average actual delay. For the condition with a nominal one-week delay, the average actual delay was eight days. For the condition with a nominal delay of one month, the average actual delay was 33 days. Estimates of parameter *dP*

(culprit-presence detection) are shown in the upper panel together with graphical illustrations of two functions that have been fitted to the estimates of parameter *dP*. The continuous orange curve shows the standard power function, the dashed orange curve shows the simplified function derived from Wickelgren's [80] power-exponential forgetting theory by Wixted and Carpenter [81] (see text for details). Both curves are plotted so as to align with the empirical data points on the delay axis (0, 1, 8, 33 days). Estimates of parameter *g* (guessing-based selection) are shown in the middle panel. Estimates of parameter *dA* (culprit-absence detection) are shown in the lower panel. Error bars represent 95 % confidence intervals.

The estimates of parameters *dP* (culprit-presence detection), *g* (guessing-based selection) and *dA* (culprit-absence detection) are displayed in Figure 2. The estimates of parameter *dP* (upper panel of Figure 2) clearly decline as a function of delay. To test whether this decline is statistically significant, parameter *dP* was set to be equal across all four delay conditions. The reduction in fit of the model with this equality restriction relative to the base model was statistically significant, $\Delta G^2(3) = 176.43$, $p < .001$. The model implying that parameter *dP* does not differ among delay conditions thus must be rejected, leading to the conclusion that the probability of culprit-presence detection declines as a function of delay. Next, the two most successful of the five 'classic' functions describing forgetting curves [4], a power function and a logarithmic function, were fitted to the estimates of parameter *dP*. Here we took into account that the standard power and logarithmic functions, $dP = \lambda \cdot (delay)^{-\psi}$ and $dP = \lambda \cdot \ln(delay) + \psi$, are undefined at $delay = 0$. In doing so we followed Wixted and Ebbesen [16] and used modified versions of these standard functions of the form $dP = \lambda \cdot (1 + delay)^{-\psi}$ and $dP = \lambda \cdot \ln(1 + delay) + \psi$, which are defined at $delay = 0$ and quickly approximate $dP = \lambda \cdot (delay)^{-\psi}$ and $dP = \lambda \cdot \ln(delay) + \psi$, respectively, as delay increases. Both functions fitted the data well, but the best fitting power function, $\widehat{dP} = 0.2744 \cdot (1 + delay)^{-0.726}$, $R^2 = 0.98$ (shown as the continuous orange curve in Figure 2) fitted the data even better than the best fitting logarithmic function, $\widehat{dP} = -0.0679 \cdot \ln(1 + delay) + 0.2377$, $R^2 = 0.88$ (not shown in Figure 2). Whereas these functions provide an excellent description of the data, they are not theoretically motivated. The latter approach—to fit a theoretically motivated forgetting function—was taken in a meta-analysis by Deffenbacher et al. [18] who fitted the function implied by the power-exponential forgetting theory proposed by Wickelgren [e. g., 80] to eleven data sets obtained in facial memory studies. However, while Deffenbacher et al. [18] noted that this function was of necessity only fitted by eye due to limited data points making formal data fitting impractical [18 p. 145], they also discussed the simplified version of Wickelgren's power-exponential forgetting function [80] proposed by Wixted and Carpenter [81]. This simplification makes this forgetting function more practical for data fitting in empirical memory studies. Specifically, Wixted and Carpenter [81] have shown that, under typical conditions, Wickelgren's power-exponential forgetting function reduces to

$$m = \lambda(1 + \beta t)^{-\psi}, \tag{1}$$

where m is memory strength, λ is the state of long-term memory at $t = 0$, β is a scaling parameter, t is the time delay and ψ is the rate of forgetting. This function can be fitted to the present data given a boundary condition proposed by Wickelgren [80]. This boundary condition is that $m(t = 0) = \lambda$ which was therefore equated with $dP(t = 0)$. An estimate of dP at $t = 0$ is known ($\widehat{dP} = 0.2820$ in the immediate condition, see Figure 2). It is therefore straightforward to set λ = 0. 2820. Given this, the fit of the simplified version of the forgetting function implied by Wickelgren's power-exponential forgetting theory [63, 64] is excellent ($\widehat{dP} = 0.2820 \cdot (1 + 2.6623 \cdot delay)^{-0.502}$, $R^2 = 0.99$, shown as the dashed orange curve in Figure 2). In the light of the similarly excellent fit of the descriptive power function mentioned above, the fact that this theoretically motivated forgetting function fits the present data so well may not be too surprising as it is, after all, also a power function, the only difference to the standard power function being the additional scaling parameter β in the theoretically motivated forgetting function.

The estimates of parameter g, in contrast, seem to be relatively constant across delays (middle panel of Figure 2). To test whether this is indeed the case, parameter g was set to be equal across all four delay conditions. The reduction in fit of the model with this equality restriction relative to the base model was not statistically significant, $\Delta G^2(3) = 6.26$, $p = .100$. The model implying that parameter g does not differ among delay conditions is compatible with the data and need not be rejected, leading to the conclusion that the probability of guessing-based selection does not significantly change as a function of delay.

Finally, the estimates of parameter dA were very low (bottom panel of Figure 2). Given that the parameter estimates were so close to the boundary of the parameter space, we used the parametric bootstrap procedure implemented in multiTree [57] to obtain a p value based on a simulated sampling distribution [56, 82]. The model in which parameter dA was set to be equal across all four delay conditions did not fit significantly worse than the base model, $\Delta G^2(3) = 0.14$, bootstrapped $p = .814$, indicating that the probability of culprit-absence detection does not significantly change as a function of delay.

Discussion

The goal of the present study was to examine how the processes underlying eyewitness responses are affected by the delay between viewing a staged-crime video and responding to lineups. More specifically, we tested hypotheses about how delays affect culprit-presence detection, guessing-based selection and culprit-absence detection. Extending previous studies in which the effects of delay were typically investigated by comparing a condition with no or a small delay to another condition with a larger delay [24-26, 30, 32-34], here the delay variable had four levels: no delay, one day, one week (average actual delay: eight days) and one month (average actual delay: 33 days), allowing us to examine the form of changes in the latent processes underlying eyewitness responses over time. This was done using a large sample of $N = 2108$ participants, each of whom contributed four data points, thus providing for sensitive statistical tests of the effects of delay on the processes underlying eyewitness responses to lineups.

The results are clear-cut for the memory-based process of culprit-presence detection (parameter dP) which declines in a way that is described well by a power function, one of the two 'classic' functions that have been found to best describe the decline in the ability to remember over time [4]. In fact, the present results are strikingly parallel to those from experimental paradigms as diverse as human face memory, matching-to-sample with pigeons and even Ebbinghaus' original savings data for which a power function has been found to fit forgetting curves best and even slightly better than a logarithmic function, just like in the present case [16]. Additionally, the present results align closely with the simplified version of Wickelgren's power-exponential forgetting function [e. g., 80] proposed by Wixted and Carpenter [81]. In sum, then, the changes in the memory-based process of culprit-presence detection (parameter dP) as a function of delay are consistent with what is known about the time-course of forgetting in general. With four levels of the delay variable, the study presented here allows for this conclusion which would not have been possible to draw based on only two levels of the delay variable, the latter of which is characteristic of most studies on the effects of delay on eyewitness memory [24-26, 30, 32-34].

In recommendations of how to perform lineups it has been noted that “eyewitness memory can fade with the passage of time. Hence, a lineup should be conducted as soon as possible after establishing evidence-based suspicion” [1]. The present results demonstrate just how rapid the fading of memory-based culprit-presence detection can be at small delays already. This underscores the critical importance of conducting lineups as soon as possible after a crime to maximize the chances of culprit identifications based on memory. It also highlights the importance of educating those involved in criminal trials such as jurors about the rapid reduction in memory-based culprit-detection processes within the first days following the crime [83], particularly when considering that naïve metacognitive judgements typically do not anticipate the rapid declines reflected in empirical forgetting curves [84, 85].

The results are also clear-cut for guessing-based selection (parameter *g*) which does not change as a function of delay. Specifically, guessing-based selection does not increase parallel to the delay-related reduction in culprit-presence detection; thus, there is no evidence of compensatory guessing [61-64]. Culprit-absence detection (parameter *dA*) did not vary as a function of delay. Notably, the value of *dA* was already close to zero in the no-delay condition and remained at this low level across all delays. This pattern of *dA* is commonly observed in the literature [44, 47, 48, 65] and the explanation for this pattern is straightforward. Whereas culprit-presence detection requires just one lineup member (the culprit) to elicit culprit-presence detection, culprit-absence detection requires the eyewitness to rule out every single lineup member as the culprit which is usually much more difficult.

The present results align with those of previous studies according to which delay primarily affects culprit-present lineups as opposed to culprit-absent lineups [27-29, 42]. Consistent with these findings, the most striking descriptive observation at the level of observable behavior is that culprit identification rates decrease with increasing delay (Table 2). A priori, this pattern could have been attributed to various underlying processes, such as a decline in culprit-presence detection, changes in guessing-based selection, or a combination of these and other processes. The model-based analysis presented here disambiguates this pattern by demonstrating that the decrease in culprit identification rates is driven by a pronounced decline in culprit-presence detection as a function of delay. By contrast, guessing-based selection and all other processes remain constant across delays.

As a limitation, it should be mentioned that the forgetting functions evaluated here describe the decline in culprit-presence detection but do not allow conclusions about mechanisms of forgetting such as decay, interference or consolidation. Future studies could aim at further disentangling the contributions of these mechanisms. Furthermore, the present conclusions rely on a single experiment with a large sample size ($N = 2108$) to assess the effects of four levels of delay (no delay, one day, one week, one month) on the processes underlying eyewitness responses. While our results align well with those of previous research regarding the decline of memory as described by a power function [16], future research could further test the robustness of the conclusions drawn here by investigating even longer delays, sequential lineups versus simultaneous lineups, single-culprit versus multiple-culprit crimes as well as other variations in lineup procedures and stimulus materials [86].

In sum, the results of the present study underscore the critical importance of conducting lineups as soon as possible after a crime to maximize the chances of memory-based culprit-presence detections. Archival studies from Great Britain suggest that the most common delay between a crime and an associated lineup is one to three months [20-23]. Although it is not justified to generalize the exact time course of the decline in culprit-presence detection observed here to real-world cases, the rapid initial reduction in memory-based culprit detections strengthens the argument that lineups should be conducted as soon as possible, ideally within hours or days after the crime, rather than weeks or months later. These findings also highlight the need to educate jurors and others involved in criminal trials about the sharp decline in memory-based culprit-presence-detection within the first days after the crime.

References

1. Wells, G. L., *et al.* Policy and procedure recommendations for the collection and preservation of eyewitness identification evidence. *Law Hum. Behav.* ***44*****,** 3–36. https://doi.org/10.1037/lhb0000359 (2020).

2. Ebbinghaus, H. *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie [Memory: A Contribution to Experimental Psychology]* (Duncker & Humblot, 1885).

3. Radvansky, G. A., Doolen, A. C., Pettijohn, K. A. & Ritchey, M. A new look at memory retention and forgetting. *J. Exp. Psychol. Learn. Mem. Cogn.* ***48*****,** 1698–1723. https://doi.org/10.1037/xlm0001110 (2022).

4. Rubin, D. C. & Wenzel, A. E. One hundred years of forgetting: A quantitative description of retention. *Psychol. Rev.* ***103*****,** 734–760. https://doi.org/10.1037/0033-295X.103.4.734 (1996).

5. Wheatcroft, J. M., Wagstaff, G. F. & Manarin, B. The influence of delay and item difficulty on eyewitness confidence and accuracy. *Int. J. Soc. Sci. Res.* **1,** 1–9. https://doi.org/10.6000/2371-1655.2015.01.01 (2015).

6. Flin, R., Boon, J., Knox, A. & Bull, R. The effect of a five-month delay on children's and adults' eyewitness memory. *Br. J. Psychol.* **83,** 323–336. https://doi.org/10.1111/j.2044-8295.1992.tb02444.x (1992).

7. Odinot, G. & Wolters, G. Repeated recall, retention interval and the accuracy–confidence relation in eyewitness memory. *Appl. Cogn. Psychol.* **20,** 973–985. https://doi.org/10.1002/acp.1263 (2006).

8. Spearing, E. R. & Wade, K. A. Long retention intervals impair the confidence–accuracy relationship for eyewitness recall. *J. Appl. Res. Mem. Cogn.* **11,** 384–391. https://doi.org/10.1037/mac0000014 (2022).

9. Ellis, H., Shepherd, J. & Davies, G. The deterioration of verbal descriptions of faces over different delay intervals. *J. Police Sci. Admin.* **8,** 101–106. (1980).

10. Turtle, J. W. & Yuille, J. C. Lost but not forgotten details: Repeated eyewitness recall leads to reminiscence but not hypermnesia. *J. Appl. Psychol.* **79,** 260–271. https://doi.org/10.1037/0021-9010.79.2.260 (1994).

11. Shepherd, J. W., Gibling, F. & Ellis, H. The effects of distinctiveness, presentation time and delay on face recognition. *Eur. J. Cogn. Psychol.* **3,** 137–145. https://doi.org/10.1080/09541449108406223 (1991).

12. MacLin, O. H., MacLin, M. K. & Malpass, R. S. Race, arousal, attention, exposure and delay: An examination of factors moderating face recognition. *Psychol. Public Policy Law* **7,** 134–152. https://doi.org/10.1037/1076-8971.7.1.134 (2001).

13. Barkowitz, P. & Brigham, J. C. Recognition of faces: Own-race bias, incentive, and time delay. *J. Appl. Soc. Psychol.* **12,** 255-268. https://doi.org/10.1111/j.1559-1816.1982.tb00863.x (1982).

14. Chance, J. E. & Goldstein, A. G. Retention interval and face recognition: Response latency measures. *Bull. Psychon. Soc.* **25,** 415-418. https://doi.org/10.3758/BF03334728 (1987).

15. Courtois, M. R. & Mueller, J. H. Target and distractor typicality in facial recognition? *J. Appl. Psychol.* **66,** 639-645. https://doi.org/10.1037/0021-9010.66.5.639 (1981).

16. Wixted, J. T. & Ebbesen, E. B. On the form of forgetting. *Psychol. Sci.* **2,** 409–415. https://doi.org/10.1111/j.1467-9280.1991.tb00175.x (1991).

17. Shepherd, J. & Ellis, H. The effect of attractiveness on recognition memory for faces. *Am. J. Psychol.* **86,** 627-633. https://doi.org/10.2307/1421948 (1973).

18. Deffenbacher, K. A., Bornstein, B. H., McGorty, E. K. & Penrod, S. D. Forgetting the once-seen face: estimating the strength of an eyewitness's memory

representation. *J. Exp. Psychol. Appl.* ***14*****,** 139–150. https://doi.org/10.1037/1076-898X.14.2.139 (2008).

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

20. Horry, R., Memon, A., Wright, D. B. & Milne, R. Predictors of eyewitness identification decisions from video lineups in England: A field study. *Law Hum. Behav.* ***36*****,** 257–265. https://doi.org/10.1037/h0093959 (2012).

21. Pike, G., Brace, N. & Kynan, S. *The visual identification of suspects: Procedures and practice. Home Office, A Publication of the Policing and Reducing Crime Unit Home Office Research, Development and Statistics Directorate.* https://webarchive.nationalarchives.gov.uk/ukgwa/20130128103514/http://www.homeoffice.gov.uk/rds/pdfs2/brf202.pdf (2002). Accessed June 8 2024.

22. Memon, A., Havard, C., Clifford, B., Gabbert, F. & Watt, M. A field evaluation of the VIPER system: A new technique for eliciting eyewitness identification evidence. *Psychol. Crime Law* **17,** 711–729. https://doi.org/10.1080/10683160903524333 (2011).

23. Horry, R., Halford, P., Brewer, N., Milne, R. & Bull, R. Archival analyses of eyewitness identification test outcomes: What can they tell us about eyewitness memory? *Law Hum. Behav.* ***38*****,** 94–108. https://doi.org/10.1037/lhb0000060 (2014).

24. Sauer, J., Brewer, N., Zweck, T. & Weber, N. The effect of retention interval on the confidence–accuracy relationship for eyewitness identification. *Law Hum. Behav.* ***34*****,** 337–347. https://doi.org/10.1007/s10979-009-9192-x (2010).

25. Palmer, M. A., Brewer, N., Weber, N. & Nagesh, A. The confidence-accuracy relationship for eyewitness identification decisions: Effects of exposure duration, retention interval, and divided attention. *J. Exp. Psychol. Appl.* ***19*****,** 55–71. https://doi.org/10.1037/a0031602 (2013).

26. Lucas, C. A., Brewer, N. & Palmer, M. A. Eyewitness identification: The complex issue of suspect-filler similarity. *Psychol. Public Policy Law* **27,** 151–169. https://doi.org/10.1037/law0000243 (2021).

27. Cutler, B. L., Penrod, S. D., O'Rourke, T. E. & Martens, T. K. Unconfounding the effects of contextual cues on eyewitness identification accuracy. *Soc. Behav.* ***1*****,** 113–134. (1986).

28. Lin, W., Strube, M. J. & Roediger III, H. L. The effects of repeated lineups and delay on eyewitness identification. *Cogn. Res. Princ. Implic.* **4,** 16. https://doi.org/10.1186/s41235-019-0168-1 (2019).

29. Egan, D., Pittner, M. & Goldstein, A. G. Eyewitness identification: Photographs vs. live models. *Law Hum. Behav.* ***1*****,** 199–206. https://doi.org/10.1007/BF01053439 (1977).

30. Krafka, C. & Penrod, S. Reinstatement of context in a field experiment on eyewitness identification. *J. Pers. Soc. Psychol.* ***49*****,** 58–69. https://doi.org/10.1037/0022-3514.49.1.58 (1985).

31. Shepherd, J. Identification after long delays. *Evaluating Witness Evidence***,** 173–187. (1983).

32. Karageorge, A. & Zajac, R. Exploring the effects of age and delay on children's person identifications: Verbal descriptions, lineup performance, and the influence of wildcards. *Br. J. Psychol.* ***102*****,** 161–183. https://doi.org/10.1348/000712610X507902 (2011).

33. Wetmore, S. A., *et al.* Effect of retention interval on showup and lineup performance. *J. Appl. Res. Mem. Cogn.* ***4*****,** 8–14. https://doi.org/10.1016/j.jarmac.2014.07.003 (2015).

34. Giacona, A. M., Lampinen, J. M. & Anastasi, J. S. Estimator variables can matter even for high-confidence lineup identifications made under pristine conditions. *Law Hum. Behav.* ***45*****,** 256–270. https://doi.org/10.1037/lhb0000381 (2021).

35. Gettleman, J. N., Grabman, J. H., Dobolyi, D. G. & Dodson, C. S. A decision processes account of the differences in the eyewitness confidence-accuracy relationship between strong and weak face recognizers under suboptimal exposure and delay conditions. *J. Exp. Psychol. Learn. Mem. Cogn.* ***47*****,** 402–421. https://doi.org/10.1037/xlm0000922 (2021).

36. Wells, G. L. & Lindsay, R. On estimating the diagnosticity of eyewitness nonidentifications. *Psychol. Bull.* ***88*****,** 776-784. https://doi.org/10.1037/0033-2909.88.3.776 (1980).

37. Wixted, J. T. & Mickes, L. The field of eyewitness memory should abandon probative value and embrace receiver operating characteristic analysis. *Perspect. Psychol. Sci.* ***7*****,** 275-278. https://doi.org/10.1177/1745691612442906 (2012).

38. Wixted, J. T. & Mickes, L. A signal-detection-based diagnostic-feature-detection model of eyewitness identification. *Psychol. Rev.* ***121*****,** 262–276. https://doi.org/10.1037/a0035940 (2014).

39. Wells, G. L., Smalarz, L. & Smith, A. M. ROC analysis of lineups does not measure underlying discriminability and has limited value. *J. Appl. Res. Mem. Cogn.* ***4*****,** 313–317. https://doi.org/10.1016/j.jarmac.2015.08.008 (2015).

40. Wells, G. L., Smith, A. M. & Smalarz, L. ROC analysis of lineups obscures information that is critical for both theoretical understanding and applied purposes. *J. Appl. Res. Mem. Cogn.* ***4*****,** 324–328. https://doi.org/10.1016/j.jarmac.2015.08.010 (2015).

41. Smith, A. M., Wells, G. L., Lindsay, R. & Penrod, S. D. Fair lineups are better than biased lineups and showups, but not because they increase underlying discriminability. *Law Hum. Behav.* ***41*****,** 127-145. https://doi.org/10.1037/lhb0000219 (2017).

42. Shapiro, P. N. & Penrod, S. Meta-analysis of facial identification studies. *Psychol. Bull.* ***100*****,** 139–156. https://doi.org/10.1037/0033-2909.100.2.139 (1986).

43. Menne, N. M., Winter, K., Bell, R. & Buchner, A. A validation of the two-high threshold eyewitness identification model by reanalyzing published data. *Sci. Rep.* ***12*****,** 13379. https://doi.org/10.1038/s41598-022-17400-y (2022).

44. Menne, N. M., Winter, K., Bell, R. & Buchner, A. Measuring lineup fairness from eyewitness identification data using a multinomial processing tree model. *Sci. Rep.* ***13*****,** 6290. https://doi.org/10.1038/s41598-023-33101-6 (2023).

45. Menne, N. M., Winter, K., Bell, R. & Buchner, A. The effects of lineup size on the processes underlying eyewitness decisions. *Sci. Rep.* ***13*****,** 17190. https://doi.org/10.1038/s41598-023-44003-y (2023).

46. Winter, K., Menne, N. M., Bell, R. & Buchner, A. Experimental validation of a multinomial processing tree model for analyzing eyewitness identification decisions. *Sci. Rep.* ***12*****,** 15571. https://doi.org/10.1038/s41598-022-19513-w (2022).

47. Winter, K., Menne, N. M., Bell, R. & Buchner, A. Evaluating the impact of first-yes-counts instructions on eyewitness performance using the two-high threshold eyewitness identification model. *Sci. Rep.* ***13*****,** 6572. https://doi.org/10.1038/s41598-023-33424-4 (2023).

48. Therre, A., *et al.* On the possible advantages of combining small lineups with instructions that discourage guessing-based selection. *Sci. Rep.* ***14*****,** 14126. https://doi.org/10.1038/s41598-024-64768-0 (2024).

49. Bell, R., Menne, N. M., Mayer, C. & Buchner, A. On the advantages of using AI-generated images of filler faces for creating fair lineups. *Sci. Rep.* ***14*****,** 12304. https://doi.org/10.1038/s41598-024-63004-z (2024).

50. Colloff, M. F., Wade, K. A. & Strange, D. Unfair lineups make witnesses more likely to confuse innocent and guilty suspects. *Psychol. Sci.* ***27*****,** 1227-1239. https://doi.org/10.1177/0956797616655789 (2016).

51. Smith, A. M. Why do mistaken identification rates increase when either witnessing or testing conditions get worse? *J. Appl. Res. Mem. Cogn.* ***9*****,** 495-507. https://doi.org/10.1016/j.jarmac.2020.08.002 (2020).

52. Memon, A., Hope, L. & Bull, R. Exposure duration: Effects on eyewitness accuracy and confidence. *British Journal of Psychology* ***94*****,** 339-354. https://doi.org/10.1348/000712603767876262 (2003).

53. Wilcock, R. & Bull, R. Novel lineup methods for improving the performance of older eyewitnesses. *Appl. Cogn. Psychol.* ***24*****,** 718–736. https://doi.org/10.1002/acp.1582 (2010).

54. Malpass, R. S. & Devine, P. G. Eyewitness identification: Lineup instructions and the absence of the offender. *J. Appl. Psychol.* ***66*****,** 482–489. https://doi.org/10.1037/0021-9010.66.4.482 (1981).

55. Lampinen, J. M., *et al.* Comparing detailed and less detailed pre-lineup instructions. *Appl. Cogn. Psychol.* ***34*,** 409–424. https://doi.org/10.1002/acp.3627 (2020).

56. Schmidt, O., Erdfelder, E. & Heck, D. W. How to develop, test, and extend multinomial processing tree models: A tutorial. *Psychol. Methods*. https://doi.org/10.1037/met0000561 (2023).

57. Moshagen, M. multiTree: A computer program for the analysis of multinomial processing tree models. *Behav. Res. Methods.* ***42*,** 42–54. https://doi.org/10.3758/BRM.42.1.42 (2010).

58. Batchelder, W. H. & Riefer, D. M. The statistical analysis of a model for storage and retrieval processes in human memory. *Br. J. Math. Stat. Psychol.* ***39*,** 129–149. https://doi.org/10.1111/j.2044-8317.1986.tb00852.x (1986).

59. Erdfelder, E., *et al.* Multinomial processing tree models: A review of the literature. *Z. fur Psychol. / J. Psychol.* ***217*,** 108–124. https://doi.org/10.1027/0044-3409.217.3.108 (2009).

60. Batchelder, W. H. & Riefer, D. M. Theoretical and empirical review of multinomial process tree modeling. *Psychon. Bull. Rev.* ***6*,** 57–86. https://doi.org/10.3758/BF03210812 (1999).

61. Meiser, T., Sattler, C. & Von Hecker, U. Metacognitive inferences in source memory judgements: The role of perceived differences in item recognition. *Q. J. Exp. Psychol.* ***60*,** 1015–1040. https://doi.org/10.1080/17470210600875215 (2007).

62. Küppers, V. & Bayen, U. J. Inconsistency effects in source memory and compensatory schema-consistent guessing. *Q. J. Exp. Psychol.* ***67*,** 2042–2059. https://doi.org/10.1080/17470218.2014.904914 (2014).

63. Hirshman, E. Decision processes in recognition memory: criterion shifts and the list-strength paradigm. *J. Exp. Psychol. Learn. Mem. Cogn.* ***21*,** 302–313. https://doi.org/10.1037/0278-7393.21.2.302 (1995).

64. Riefer, D. M., Hu, X. & Batchelder, W. H. Response strategies in source monitoring. *J. Exp. Psychol. Learn. Mem. Cogn.* ***20*,** 680–693. https://doi.org/10.1037/0278-7393.20.3.680 (1994).

65. Mayer, C., Bell, R., Menne, N. M. & Buchner, A. Lineup position affects guessing-based selection but not culprit-presence detection in simultaneous and sequential lineups. *Scientific Reports* ***14*,** 27642. https://doi.org/10.1038/s41598-024-78936-9 (2024).

66. Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods.* ***39*,** 175–191. https://doi.org/https://doi.org/10.3758/BF03193146 (2007).

67. Menne, N., Bell, R., Winter, K. & Buchner, A. Disclosing the number of simultaneous lineups increases guessing-based selection in cases of multiple-

culprit crimes. *Journal of Applied Research in Memory and Cognition* https://doi.org/10.1037/mac0000217 (2025).

68. Leiner, D. J. *SoSci Survey [Computer software].* https://www.soscisurvey.de (2022).

69. Tupper, N., *et al.* Showup identification decisions for multiple perpetrator crimes: Testing for sequential dependencies. *PloS one* ***13*****,** e0208403. https://doi.org/10.1371/journal.pone.0208403 (2018).

70. Lockamyeir, R. F., *et al.* One perpetrator, two perpetrators: The effect of multiple perpetrators on eyewitness identification. *Appl. Cogn. Psychol.* ***35*****,** 1206-1223. https://doi.org/10.1002/acp.3853 (2021).

71. O'Connell, M. & Synnott, J. A position of influence: Variation in offender identification rates by location in a lineup. *Journal of Investigative Psychology and Offender Profiling* ***6*****,** 139-149. https://doi.org/10.1002/jip.102 (2009).

72. Meisters, J., Birk, D. & and Musch, J. Eyewitness identification in simultaneous and sequential lineups: an investigation of position effects using receiver operating characteristics. *Memory* ***26*****,** 1297-1309. https://doi.org/10.1080/09658211.2018.1464581 (2018).

73. Carlson, C. A., *et al.* A stronger memory for the perpetrator may attenuate effects of the identification procedure on eyewitness accuracy. *Appl. Cogn. Psychol.* ***37*****,** 419-432. https://doi.org/10.1002/acp.4045 (2023).

74. Thompson, W. B. Eyewitness memory: A field study of viewing angle, pose, and eyewitness age. *Applied Psychology in Criminal Justice* ***14*****,** 102-119. (2018).

75. McCormick, K. M. & Semmler, C. Letting go of the Grail: Falsifying the theory of 'true'eyewitness identifications. *PsyArXiv*. https://doi.org/10.31234/osf.io/dszqw (2023).

76. Fitzgerald, R. J., Rubínová, E. & Juncu, S. Eyewitness identification around the world. In *Methods, measures, and theories in eyewitness identification tasks* (Eds. Smith, A. M., Toglia, M. & Lampinen, J. M.) 294–322. https://doi.org/10.4324/9781003138105-16 (Routledge/Taylor & Francis Group, 2021).

77. Fitzgerald, R. J., Price, H. L. & Valentine, T. Eyewitness Identification: Live, Photo, and Video Lineups. *Psychol. Public Policy Law* ***24*****,** 307-325. https://doi.org/10.1037/law0000164 (2018).

78. Minear, M. & Park, D. C. A lifespan database of adult facial stimuli. *Beh. Res. Meth. Instr. Comp.* ***36*****,** 630–633. https://doi.org/10.3758/BF03206543 (2004).

79. Oriet, C. & Fitzgerald, R. J. The single lineup paradigm: A new way to manipulate target presence in eyewitness identification experiments. *Law Hum. Behav.* ***42*****,** 1-12. https://doi.org/10.1037/lhb0000272 (2018).

80. Wickelgren, W. A. Single-trace fragility theory of memory dynamics. *Mem. Cogn.* ***2*****,** 775-780. https://doi.org/10.3758/BF03198154 (1974).

81. Wixted, J. T. & Carpenter, S. K. The Wickelgren power law and the Ebbinghaus savings function. *Psychol. Sci.* ***18*****,** 133-134. https://doi.org/10.1111/j.1467-9280.2007.01862.x (2007).

82. Klauer, K. C. & Oberauer, K. Testing the mental model theory of propositional reasoning. *The Quarterly Journal of Experimental Psychology* ***48*****,** 671-687. https://doi.org/10.1080/14640749508401410 (1995).

83. Cutler, B. L., Penrod, S. D. & Dexter, H. R. Juror sensitivity to eyewitness identification evidence. *Law Hum. Behav.* **14,** 185–191. https://doi.org/10.1007/BF01062972 (1990).

84. Koriat, A., Bjork, R. A., Sheffer, L. & Bar, S. K. Predicting one's own forgetting: the role of experience-based and theory-based processes. *J. Exp. Psychol. Gen.* ***133*****,** 643-656. https://doi.org/10.1037/0096-3445.133.4.643 (2004).

85. Blech, C. & Gaschler, R. Assessing Students' Knowledge about Learning and Forgetting Curves with a Free Production Technique: Measures and Implications for the Development of Learning Aids. *Psychol. Learn. Teach.* **17,** 308-322. https://doi.org/10.1177/1475725718779684 (2018).

86. Fitzgerald, R. J., Rubínová, E., Ribbers, E. & Juncu, S. Eyewitness Lineup Identity (ELI) database: Crime videos and mugshots for eyewitness identification research. *Behav. Res. Methods.* **57,** 63. https://doi.org/10.3758/s13428-024-02585-z (2025).

Data availability

The files with the frequency data and the equations needed for the model-based analyses are available at https://osf.io/fqus6.

Author contributions

AT developed the experiment and implemented suggestions by RB, NMM, CM, UL and AB. AT collected the data and analyzed them with contributions of RB, NMM, CM, UL and AB. AT wrote the first draft of the manuscript. RB, NMM, CM, UL and AB revised the manuscript. All authors gave final approval for publication.

Funding

The work reported herein was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – BU 945/10-1, project number 456214986.

Competing interests

The authors declare no competing interests.

Article 3

This article includes Experiment 3.

Therre, A., Bell, R., Menne, N. M., Mayer, C., Lichtenhagen, U., & Buchner, A. (Under review). On the cognitive processes underlying the verbal overshadowing effect: Culprit descriptions reduce culprit-presence detection and guessing-based selection in eyewitness responses to lineups. Manuscript under review at *Journal of Applied Research in Memory and Cognition*.

On the Cognitive Processes Underlying the Verbal Overshadowing Effect: Culprit Descriptions Reduce Culprit-Presence Detection and Guessing-Based Selection in Eyewitness Responses to Lineups

Amelie Therre*, Raoul Bell, Nicola Marie Menne, Carolin Mayer, Ulla Lichtenhagen & Axel Buchner

Department of Experimental Psychology, Heinrich Heine University Düsseldorf, Germany

Author Note

Amelie Therre: https://orcid.org/0009-0002-2345-0218

Raoul Bell: https://orcid.org/0000-0002-0592-0362

Nicola Marie Menne: https://orcid.org/0000-0001-9421-8286

Carolin Mayer: https://orcid.org/0009-0003-5278-4361

Ulla Lichtenhagen: https://orcid.org/0009-0001-8243-4632

Axel Buchner: https://orcid.org/0000-0003-4529-3444

We have no known conflict of interest to disclose.

This study was not preregistered.

The work reported herein was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – BU 945/10-1, project number 456214986.

Correspondence concerning this article should be addressed to Amelie Therre, Department of Experimental Psychology, Heinrich Heine University Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany

E-Mail: Amelie.Therre@hhu.de

Word Count: 2548 words

Reviewer-Link: https://osf.io/xke8w

Abstract

Culprit descriptions and eyewitness responses to lineups are essential for criminal investigations—the first to locate possible suspects and the latter to provide information relevant to determining guilt or innocence. However, describing a culprit before responding to a lineup can reduce culprit identification rates, a phenomenon known as the verbal overshadowing effect. In the present experiment, the well-validated two-high threshold eyewitness identification model was applied to investigate how culprit descriptions affect the cognitive processes underlying lineup responses. Participants either described or did not describe the culprits before responding to the lineups. If providing culprit descriptions interferes with face recognition, culprit-presence detection should be lower in the culprit-description condition compared to the no-culprit-description condition. Alternatively or additionally, culprit descriptions may discourage guessing-based selection. The results support both hypotheses: Compared to providing no culprit description, providing culprit descriptions reduced culprit-presence detection and guessing-based selection. These findings offer new insights into the processes underlying the verbal overshadowing effect.

Keywords: police lineups, eyewitness identification, two-high threshold eyewitness identification model, multinomial processing tree model, verbal overshadowing effect, culprit descriptions

General Audience Summary

Eyewitnesses play a critical role in criminal investigations. Initially, they often describe a culprit's appearance, helping police to narrow down potential suspects. Later, eyewitnesses might view a police lineup consisting of a suspect—who might be the culprit or innocent—and several known-to-be innocent fillers. The eyewitness's task is either to identify the culprit or to reject the lineup if the culprit is absent. Although both culprit descriptions and lineup responses are essential in criminal investigations, describing a culprit can unintentionally lead to a reduction in correct identifications of the culprit in a lineup. This phenomenon is known as the verbal overshadowing effect. In the present experiment participants either provided or did not provide culprit descriptions prior to viewing lineups. To understand why the verbal overshadowing effect occurs, we tested two possible explanations of the verbal overshadowing effect: First, describing a culprit might interfere with face recognition, making it harder to detect the presence of the culprit in a lineup compared to when no description is given beforehand. Second, describing the culprit may discourage eyewitnesses from selecting a lineup member based on guessing relative to when no description is provided. The results support both explanations: Providing culprit descriptions reduced culprit-presence detection and guessing-based selection relative to not providing culprit descriptions. The results enhance our understanding of how culprit descriptions affect the processes underlying lineup responses.

225 of max. 300 words

On the Cognitive Processes Underlying the Verbal Overshadowing Effect: Culprit Descriptions Reduce Culprit-Presence Detection and Guessing-Based Selection in Eyewitness Responses to Lineups

Obtaining culprit descriptions is often necessary before conducting a lineup (Brown et al., 2008; Dodson et al., 2024; Wells et al., 2020), but providing culprit descriptions can have unintended consequences for eyewitness responses to subsequent lineups. The *verbal overshadowing* effect (Schooler & Engstler-Schooler, 1990) refers to a reduction of culprit identifications from culprit-present lineups after having provided a culprit description (Alogna et al., 2014; Dodson et al., 2024; Holdstock et al., 2022; Wilson et al., 2018). In the present study we applied the well-validated two-high threshold (2-HT) eyewitness identification model (Menne et al., 2022; Winter et al., 2022) to understand how providing culprit descriptions affects the processes underlying eyewitness responses to lineups, using the data from all response categories available from the lineups. This approach enabled us to test whether requiring culprit descriptions affects the probability of the detection of the culprit in the lineup and the probability of guessing-based selections, two distinct processes that have not previously been disentangled within a single unified analytical framework.

One constraint in understanding how providing culprit descriptions affect these processes is that often not all possible data from lineups have been analyzed. In the seminal work by Schooler and Engstler-Schooler (1990) and its large-scale replication (Alogna et al., 2014), only culprit-present lineups were examined. While this research demonstrated that describing a culprit leads to fewer culprit identifications, the lack of culprit-absent lineups limits conclusions on the causes of the verbal overshadowing effect. One possibility is that the cognitive processes involved in describing a face differ fundamentally from those used in efficient face recognition. Whereas faces are processed holistically (Cheung & Gauthier, 2010; Meltzer & Bartlett, 2019; Richler & Gauthier, 2014) and holistic processing predicts subsequent face recognition (Richler et al., 2011), verbalizing the characteristics of a face can result in descriptions of the face as a sum of its parts, creating a mismatch (Chin &

Schooler, 2008; Wickham & Lander, 2008). This mismatch may complicate the detection of the culprit's face in a lineup.

Another possibility is that the difficulty experienced when verbally describing a face leads eyewitnesses to be more reluctant to make guessing-based selections. Guessing is common among eyewitnesses in laboratory and real-world settings (Horry et al., 2012) and a reduction in guessing-based selection would lead to fewer identifications overall. Therefore, reduced guessing-based selection could also cause the observed decrease in culprit identifications. In line with this explanation, when studies include culprit-absent lineups, culprit descriptions have often been found to lead to decreased suspect identification rates in culprit-absent lineups as well (Clare & Lewandowsky, 2004; Dodson et al., 2024; Holdstock et al., 2022; Mickes & Wixted, 2015; Smith & Flowe, 2015; Wilson et al., 2018). Given this, a measurement model is needed to determine whether the verbal overshadowing effect is driven by a reduction in guessing-based selection, culprit-presence detection, or both.

Analyses using receiver operating characteristics (ROC) yielded inconsistent results. Sometimes providing culprit descriptions reduced the discriminability between the culprit and an innocent suspect (Dodson et al., 2024; Smith & Flowe, 2015; Wilson et al., 2018) but not always (Holdstock et al., 2022; Sporer et al., 2016). A limitation of these analyses is that only culprit identification and innocent-suspect identification rates were considered because only these two data categories can be used to determine ROCs. Filler identifications and lineup rejections were not considered separately. Consequently, important information was disregarded (Smith et al., 2017; Wells et al., 2015). For instance, in culprit-absent lineups, identifying a filler is a false response whereas rejecting a culprit-absent lineup is a correct response. The fact that one response is false and the other is correct suggests that they result from different underlying processes. Distinguishing between these response categories could therefore inform the understanding of the underlying processes. A further complication is that the culprit-absent lineups in these studies did not include a designated innocent suspect but only fillers. It was thus impossible to determine proper innocent-suspect identification rates. As a substitute, pseudo-innocent-suspect identification rates were determined by dividing filler identification rates in culprit-absent lineups by lineup size. Given that this method imposes an artificial ceiling on the innocent-suspect identification rate, it is appropriate only if a lineup is perfectly fair, a rather restrictive assumption that may or may not hold (Fitzgerald et al., 2023; Quigley-McBride & Wells, 2021). This is why the culprit-absent lineups used here included designated innocent suspects and the crossed-lineup procedure was used which ensures that culprits and innocent suspects differ to the same degree from the fillers (for details, see the *Materials and procedure* section).

Another limitation of the previous studies is that ROC analyses yield only a single measure—the (partial) area under the curve. In lineup research, this measure reflects how well the culprit can be discriminated from an innocent suspect. Unfortunately, ROC analyses do not provide a measure for testing the hypothesis about the potentially more cautious responding in the culprit-description condition compared to the no-culprit-description condition. To test this important hypothesis, researchers had to turn to measures alien to the ROC measurement model such as signal-detection theory's response-bias measure *c* (Sporer et al., 2016) or raw response rates (Clare & Lewandowsky, 2004; Dodson et al., 2024; Holdstock et al., 2022; Smith & Flowe, 2015; Wilson et al., 2018). However, both approaches require switching from the measurement model underlying ROC analyses to alternative measurement models with different underlying assumptions. Also, both approaches force researchers to rely on only a subset of the available response categories (e. g., culprit identifications and innocent-suspect identifications in case of signal-detection theory's response-bias measure *c*), as a consequence of which substantial and potentially informative data are discarded. To test hypotheses about the processes underlying eyewitness identification responses, it is preferable to use one single empirically validated measurement model that is based on a coherent set of assumptions, incorporates all available response categories and allows for model fit to be evaluated as an indicator of the model's appropriateness, prior to interpreting the model parameters.

The 2-HT eyewitness identification model (Menne et al., 2022; Winter et al., 2022) is such a measurement model. Based on culprit identifications, filler identifications and lineup rejections in culprit-present lineups and innocent-suspect identifications, filler identifications and lineup rejections in culprit-absent lineup, it allows to measure four different types of processes, as will be explicated below. Through four different validation experiments (Winter et al., 2022) and eight different reanalyses (Menne et al., 2022) of published data from different laboratories around the world using diverse stimulus materials and procedures (Colloff et al., 2016; Karageorge & Zajac, 2011; Lampinen et al., 2020; Malpass & Devine, 1981; Memon et al., 2003; Smith, 2020; Wetmore et al., 2015; Wilcock & Bull, 2010), the 2-HT eyewitness identification model has been thoroughly validated, demonstrating that the model parameters accurately measure the latent processes they were designed to measure.

The 2-HT eyewitness identification model (Figure 1) is a multinomial processing tree model (e. g., Batchelder & Riefer, 1986, 1999; Erdfelder et al., 2009). In the model it is transparently specified how latent processes occurring with certain probabilities (Batchelder & Riefer, 1999) lead to observable lineup responses. The probabilities of these processes are represented by model parameters which can be statistically compared across conditions, allowing for statistical hypothesis tests directly at the process level. The model has been described previously (Bell et al., 2024; Mayer et al., 2024; Mayer et al., in press; Menne et al., 2025; Menne et al., 2022, 2023a, 2023b; Therre et al., 2024; Winter et al., 2022; Winter et al., 2023), but its description is repeated here for convenience.

In a culprit-present lineup (upper tree in Figure 1), the presence of the culprit is detected with probability *dP*, leading to a correct culprit identification. With probability 1 − *dP*, culprit-presence detection does not occur in which case biased suspect selection occurs with probability *b*, also leading to a culprit identification. Biased suspect selection refers to the selection of a suspect who stands out from the other lineup members. Parameter *b* thus captures lineup unfairness and allows the other model parameters to be determined without being contaminated by lineup unfairness. No biased suspect selection occurs with probability 1 − *b* in which case a lineup member is selected based on guessing with probability *g*. Within the 2-HT eyewitness identification model, guessing is defined as selecting a lineup member in the absence of detection, without any systematic bias towards selecting the suspect over fillers. If guessing occurs, the constant 1 ÷ *n*, representing the reciprocal of the lineup size *n*, determines the probability of selecting the culprit. Probability 1 − (1 ÷ *n*) determines the probability of selecting a filler. No guessing-based selection occurs with probability 1 − *g*, leading to a false lineup rejection.

In a culprit-absent lineup (lower tree in Figure 1), the absence of the culprit is detected with probability *dA*, leading to a correct lineup rejection. With probability 1 − *dA,* culprit-absence detection does not occur. In this case, the same non-detection-based processes as in culprit-present lineups occur because, in the absence of detection-based processes, eyewitnesses cannot differentiate between lineups with and without culprit.

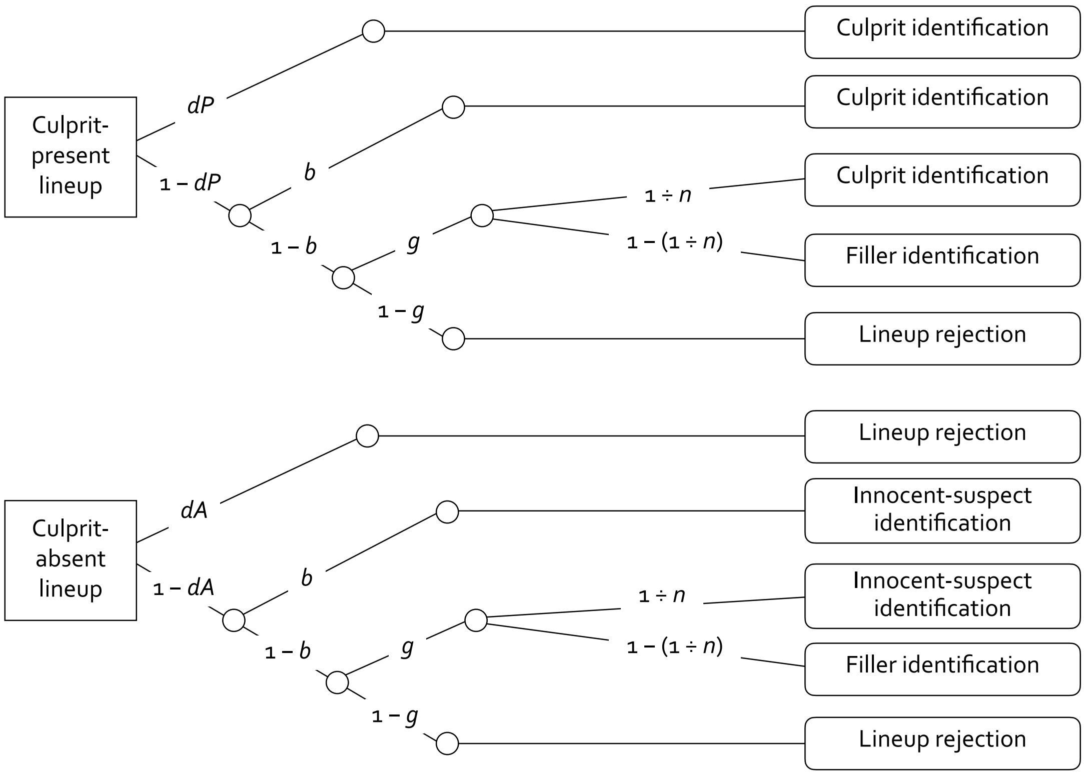

Figure 1. Graphical illustration of the 2-HT eyewitness identification model, as originally proposed and used in previous studies (Bell et al., 2024; Mayer et al., 2024; Mayer et al., in press; Menne et al., 2025; Menne et al., 2022, 2023a, 2023b; Therre et al., 2024; Winter et al., 2022; Winter et al., 2023). Squares on the left represent the two types of lineups an eyewitness may be confronted with: culprit-present lineups and culprit-absent lineups. Letters along the branches denote parameters representing the latent processes specified in the model: *dP* represents the probability of culprit-presence detection, *b* represents the probability of biased suspect selection and is larger than zero in unfair lineups in which the suspect stands out from the fillers, *g* represents the probability of guessing-based selection and *dA* represents the probability of culprit-absence detection. The constant 1 ÷ *n* represents the probability of selecting the culprit (upper tree) or the innocent suspect (lower tree) if guessing-based selection occurs, with *n* corresponding to the lineup size. The rounded rectangles on the right represent the categories of observable responses.

Depending on methodological variations, the size and the robustness of the verbal overshadowing effect vary considerably (Meissner & Brigham, 2001). To facilitate measuring the processes underlying the effect, we implemented conditions known to increase the effect's size: First, we encouraged participants to describe everything they could remember about the culprit's appearance instead of warning them to describe only what they are certain of (Clare & Lewandowsky, 2004; Dodson et al., 1997). Second, we asked for configural face descriptions instead of holistic descriptions (Wickham & Lander, 2008). Third, we implemented a 14-minute delay between witnessing the crime event and providing the culprit descriptions (Protzko et al., 2023).

If the verbal overshadowing effect arises due to an incompatibility between the processes involved in culprit descriptions and those involved in face recognition, then the culprit-detection-parameter *dP* should be lower in the culprit-description condition compared to the no-culprit-description condition. If the difficulty experienced when describing a face renders eyewitnesses more reluctant to make guessing-based selections, then the guessing-based-selection parameter *g* should be lower in the culprit-description condition compared to the no-culprit-description condition. If both components are involved in the verbal overshadowing effect, culprit-presence detection (parameter *dP*) and guessing-based selection (parameter *g*) should be reduced in the culprit-description condition compared to the no-culprit-description condition.

Method

Participants

Participants were recruited through personal contacts, social media and an email list of individuals who had signed up to be notified about psychological experiments in which they could participate. Of the 517 participants who had initially provided their informed consent and sociodemographic data, 106 participants were excluded because they did not complete the experiment or withdrew their consent, one had to be excluded for failing the attention check (see below for details), and two had to be excluded because they reported to be under 18 years old (due to German legal requirements). Therefore, valid data sets of 408 participants were available for analysis. Of those, 280 participants identified as female, 127 as male and three as non-binary. Participants had a mean age of 29 years (*SD* = 14 years). Education level was high, with 391 participants holding at least a university entrance qualification. All participants reported having high German proficiency levels. Psychology students were offered course credit for participation. Participants were randomly assigned to one of two experimental conditions: They responded to lineups either after having provided culprit descriptions (culprit-description condition, *n* = 206) or without having provided culprit descriptions (no-culprit-description condition, *n* = 202). A sensitivity analysis using G*Power (Faul et al., 2007) showed that, given *N* = 408 participants, four responses per participant and error probabilities of $\alpha = \beta = 0.05$ (implying a statistical power of $1 - \beta = .95$), an effect of the size of *w* = 0.09 could be detected when testing whether the culprit-presence-detection and guessing-based-selection parameters can be equated across the culprit-description and no-culprit-description conditions (*df* = 1).

Ethics statement

The ethics committee of the Faculty of Mathematics and Natural Sciences at Heinrich Heine University Düsseldorf has granted approval for the experiment. The experiment was conducted in accordance with the Declaration of Helsinki. Participants provided informed consent before participating. In the consent form and prior to viewing the staged-crime video, participants were informed that a video they would see would contain physical and verbal violence. They were instructed to proceed with the study only if they were comfortable watching such content.

Materials and procedure

Materials and procedure were the same as those used in previous studies (Bell et al., 2024; Mayer et al., in press; Menne et al., 2025; Menne et al., 2022, 2023a, 2023b; Therre et al., 2024; Winter et al., 2022; Winter et al., 2023), except for the delay after the staged-crime video and the culprit-description manipulation described below. The experiment was conducted online using SoSci Survey (Leiner, 2022) and was made accessible only to participants who used a desktop or laptop computer. Participants were first asked to provide their age, gender, German proficiency level and educational background. Afterwards, each participant viewed one of two staged-crime videos, henceforth referred to as Video 1 and Video 2. Both videos portrayed identical events in identical timing and sequence, with only the actors differing between the videos. Actors portraying the identical role were chosen to resemble one another in body shape, hair color and hairstyle. For example, the actor playing Character A in Video 1 was chosen to match the actor portraying Character A in Video 2; the same applies to Characters B, C, and D. In both videos, four men (the culprits) wearing fan clothing such as caps, shirts and scarfs of the German soccer club Bayern München verbally and physically attacked a man (the victim) wearing fan clothing of the soccer club Borussia Dortmund. The videos were displayed at a resolution of 885 × 500 pixels and lasted about 130 seconds. Participants could clearly see the culprits' faces from all angles including the frontal view.

Next, participants completed an attention check question in which they had to correctly identify the roles of the protagonists from the video as "Soccer fans". Participants were only able to continue the experiment if they answered the attention check correctly. Given that the optimal time delay to elicit a verbal overshadowing effect has been reported to be about 14 minutes (Protzko et al., 2023), participants then received a 14-minute distractor task. As in previous studies (Bacharach & Baker, 2024; Baker & Reysen, 2020; Wilson et al., 2018), participants received a brief instruction on how to play Tetris and then played that game. Once the 14-minute delay had elapsed, the experiment proceeded automatically.

Participants assigned to the culprit-description condition were then asked to describe the four Bayern München fans they had previously seen in the video. Parallel to earlier studies (Holdstock et al., 2022; Itoh, 2005; Smith & Flowe, 2015; Wilson et al., 2018), participants were required to spend five minutes completing the culprit descriptions. A timer displayed the time remaining for the task. Similar to the original instructions by Schooler and Engstler-Schooler (1990) that had also been used by others (Alogna et al., 2014; Holdstock et al., 2022; Wilson et al., 2018), participants saw the following instructions (all instructions displayed here were translated from German):

> "Please describe the appearance of the four Bayern München fans in as much detail as possible. Try to describe all facial features as precisely as possible. Write down everything you can think of about the appearance of the four Bayern München fans. It is important that you use the entire five minutes for this. When the time is up, you will automatically be redirected to the next page."

Underneath these instructions, four text fields were provided to describe the four culprits. The first text field was titled "Please describe the first Bayern München fan". The other text fields were titled analogously, referring to the second, third and fourth Bayern München fan. After the five minutes had elapsed, the experiment continued automatically.

Participants assigned to the no-culprit-description condition were asked to name as many countries as possible together with their capitals. Parallel to the culprit-description condition and as in earlier studies (Holdstock et al., 2022; Wilson et al., 2018), participants were required to spend five minutes completing the task. A timer displayed the time remaining for the task. Similar to previous studies (Holdstock et al., 2022; Wilson et al., 2018), participants saw the following instructions:

> "Please list as many countries as possible together with their capitals. Please do not use the internet for assistance. It is important that you use the entire five minutes. When the time is up, you will automatically be redirected to the next page.
>
> Example:
>
> Germany – Berlin
>
> France – Paris"

Underneath this instruction, one text field was provided to list countries together with their capitals. After the five minutes had elapsed, the experiment continued automatically.

Participants in both conditions were then asked to identify the Bayern München fans from the video they had seen in photo lineups. Participants from both groups received the subsequent instructions:

> "In the first part of the experiment, you saw a film with Bayern München fans. Now we want you to identify them. To do this, we will show you several lineups. In each lineup, you will see

> a series of faces. You will be asked to indicate whether one of the people in the lineup is one of the Bayern München fans. It is also possible that no one in the lineup is one of the Bayern München fans. If you recognize someone, click on the 'Yes, was present' button that belongs to the recognized face. Otherwise, click on the 'No, none of these persons was present' button."

Each participant was then successively shown four separate simultaneous lineups, each corresponding to one Bayern München fan from the previously seen video. In each lineup, one suspect was shown alongside five fillers who matched the culprit in body shape, hair color and hairstyle, all displayed in a single row (for an illustration of the lineups, see Winter et al., 2022). For each participant, two lineups were randomly selected to include a culprit (culprit-present lineups) while the other two included an innocent suspect instead of a culprit (culprit-absent lineups). As in prior research (Bell et al., 2024; Mayer et al., 2024; Mayer et al., in press; Menne et al., 2025; Menne et al., 2022, 2023a, 2023b; Therre et al., 2024; Winter et al., 2022; Winter et al., 2023), the crossed-lineup procedure was used. The two culprits in the culprit-present lineups were chosen at random from the four individuals featured in the video the participant had seen. The innocent suspects in the culprit-absent lineups were culprits from the parallel video the participant had not seen. For instance, if Characters B and D from Video 1 were shown as culprits in the culprit-present lineups, then Characters A and C from Video 2 were shown as innocent suspects in the culprit-absent lineups. The crossed-lineup procedure ensures that culprit and innocent-suspect photographs, taken immediately after filming the staged-crime videos, differ to the same degree from the filler photographs taken from a face database (Minear & Park, 2004). This mirrors police practice where the photograph of the suspect, whose innocence or guilt is unknown, may come from a different source, such as social media, than the filler photographs, which are often selected from face databases. As explained elsewhere (Menne et al., 2025), the crossed-lineup procedure is similar to the single-lineup procedure (Oriet & Fitzgerald, 2018) but the two procedures are not identical. For instance, the crossed-lineup procedure can be used to experimentally manipulate culprit presence or absence across lineups, which is not feasible if only a single lineup is used.

All lineup photographs showed faces from a frontal perspective, with neutral expressions, against a black background and without visible clothing. Photographs had been standardized in terms of consistent face sizes and lighting and were shown at a resolution of 142 × 214 pixels. Both the positions of the photographs within each lineup and the order of the lineups were randomized for each individual. After completing all lineups, participants had the option to confirm or to withdraw their consent to the use of their data. Participants were subsequently debriefed, thanked and received their course credit (if required).

Results

Data availability

The files containing the frequency data and the equations required for the model-based analyses have been made publicly available at the Open Science Framework (OSF) and can be accessed at https://osf.io/xke8w.

Model-based analyses

Response frequencies are displayed in Table 1. The model-based analyses were conducted using multiTree (Moshagen, 2010), a free and easy-to-use computer program available for several platforms (see Schmidt et al., , for an excellent introduction to the application of multinomial models). To analyze the present data, two instances of the model depicted in Figure 1 were needed, one for the culprit-description condition and one for the no-culprit-description condition. Our aim was to start with a testable base model that was as simple as possible. To achieve this, the same parameter constraints as in previous studies were imposed onto the 2-HT eyewitness identification model (Mayer et al., 2024; Menne et al., 2025; Menne et al., 2023a; Therre et al., 2024; Winter et al., 2022; Winter et al., 2023). Specifically, given that the lineups included identical photographs of the suspects and fillers in both conditions, no differences in lineup fairness were to be expected. Therefore, the biased-suspect-selection parameter *b* was set to be equal across conditions. In addition, given that the experimental manipulation used here was clearly different from manipulations known to influence culprit-absence detection (Menne et al., 2022; Winter et al., 2022), there was no reason to expect culprit-absence detection to differ between conditions. Therefore, parameter *dA* was set to be equal across conditions as well. Furthermore, given that six-person lineups were shown, the term 1 ÷ *n*, representing the probability of sampling the suspect in case of a guessing-based selection, was set to 0.16667 (approximating 1 ÷ 6) in both conditions. The base model with these constraints fit the data $G^2(2) = 0.31$, $p = .856$ which is important because it shows that the assumptions implemented in the base model are appropriate. Parameters *b* and *dA* were estimated to be .06 (*SE* = .01) and .00 (*SE* = .07), respectively. Parameter estimates for *dP* and *g* are presented in Panels A and B of Figure 2, respectively.

Description Condition	Culprit-present lineups			Culprit-absent lineups		
	Culprit identifications	Filler identifications	Lineup rejections	Innocent-suspect identifications	Filler identifications	Lineup rejections
Culprit Description	137	80	195	49	104	259
No Culprit Description	184	89	131	52	147	205

Table 1. Response frequencies in culprit-present and culprit-absent lineups for the culprit-description and no-culprit-description conditions.

Using multinomial processing tree models, hypotheses can be tested directly at the level of the postulated processes. Panel A of Figure 2 shows that, at a descriptive level, the estimate of parameter *dP* was lower in the culprit-description condition than in the no-culprit-description condition. The hypothesis that this descriptive difference in culprit-presence detection between the culprit-description condition and the no-culprit-description condition is statistically significant can be tested by setting parameter *dP* to be equal across both conditions. This additional constraint generates one degree of freedom. If the model with the additional equality constraint fits the data significantly worse than the base model, it can be concluded that parameter *dP* differs as a function of the culprit description. Indeed, setting parameter *dP* to be equal across conditions significantly decreased the fit of the model to the data relative to the fit of the base model, $\Delta G^2(1) = 10.13$, $p = .001$. This statistical test result leads to the conclusion that that culprit-presence detection is significantly reduced when participants provide a culprit description before responding to the lineup than when no culprit description is provided.

Panel B of Figure 2 shows that, at a descriptive level, the estimate of parameter *g* was lower in the culprit-description condition than in the no-culprit-description condition. Constraining the guessing-based selection parameter *g* to be equal between the culprit-description condition and the no-culprit-description condition significantly decreased the fit of the model to the data relative to the fit of the base model, $\Delta G^2(1) = 20.52$, $p < .001$, leading to the conclusion that describing a culprit before responding to a lineup reduces guessing-based selection relative to not describing a culprit.

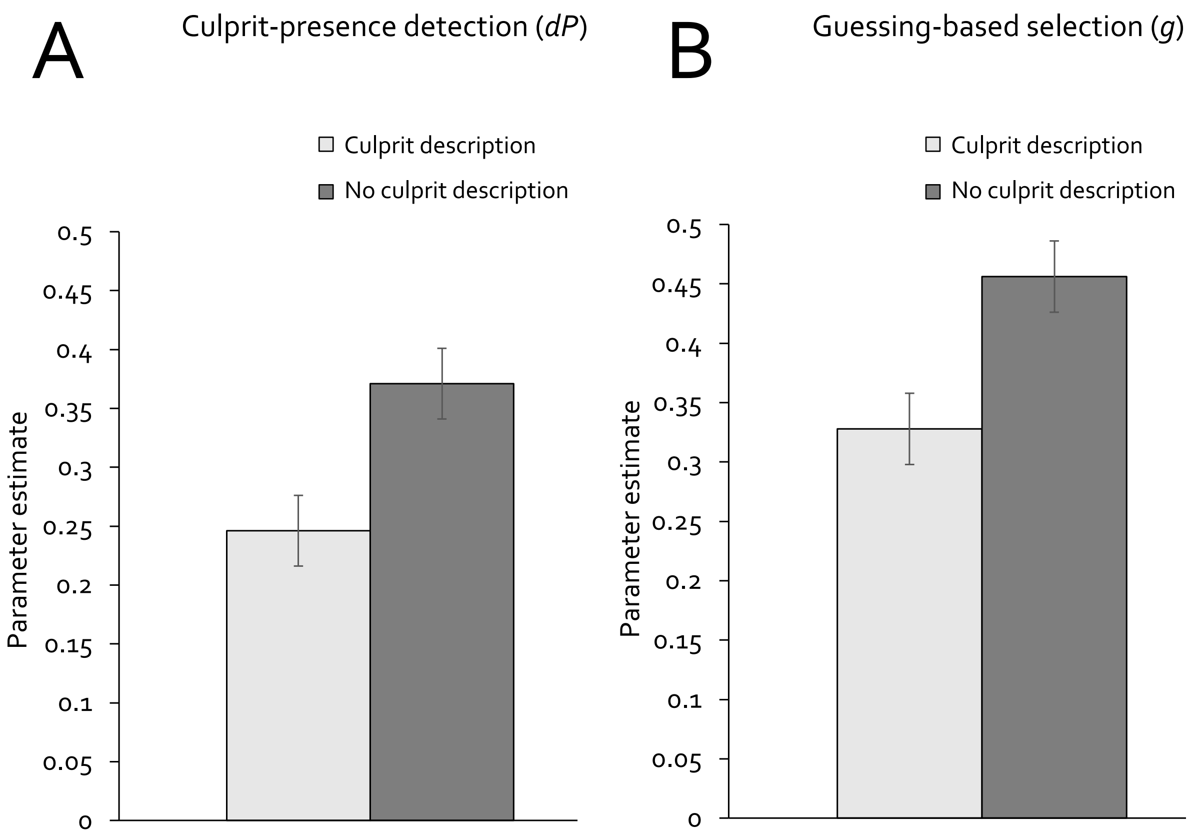

Figure 2. Estimates of the culprit-presence-detection parameter *dP* (Panel A) and the guessing-based-selection parameter *g* (Panel B) as a function of the description condition (culprit description, no culprit description). Error bars represent standard errors.

Discussion

The verbal overshadowing effect primarily refers to the finding that providing a culprit description before responding to a lineup reduces correct culprit identifications (Alogna et al., 2014; Dodson et al., 2024; Holdstock et al., 2022; Schooler & Engstler-Schooler, 1990; Wilson et al., 2018). However, the seminal study by Schooler and Engstler-Schooler (1990) and its large-scale replication (Alogna et al., 2014) included only culprit-present lineups. When culprit-absent lineups were included, identifications in culprit-absent lineups were often reduced as well (Clare & Lewandowsky, 2004; Dodson et al., 2024; Holdstock et al., 2022; Mickes & Wixted, 2015; Smith & Flowe, 2015; Wilson et al., 2018). Given this, the reduction in culprit identifications following culprit descriptions could be due to a decrease in culprit-presence detection, a decrease in guessing-based selection or both. The aim of the present study was to clarify which of these processes underlies the verbal overshadowing effect using the well-validated 2-HT eyewitness identification model (Menne et al., 2022; Winter et al., 2022). With this well-validated model, it is possible to separately measure detection-based and non-detection-based processes from all available lineup responses. The results were clear-cut. Providing culprit descriptions led to decreased culprit-presence detection relative to not providing a description (Panel A of Figure 2). This suggests that the verbal overshadowing effect is at least partly due to culprit descriptions interfering with face recognition, reducing the ability to detect the culprit's presence in the lineup. Furthermore, providing culprit descriptions also decreased guessing-based selection relative to not providing a description (Panel B of Figure 2), supporting the hypothesis that culprit descriptions cause participants to be more reluctant to make guessing-based selections. A key advantage of the 2-HT eyewitness identification model is that it allows both effects to be measured within the same analytical framework, providing a comprehensive understanding of how verbal overshadowing influences lineup performance at the process level.

Law professionals should be aware that obtaining culprit descriptions before a lineup may influence the processes underlying lineup responses in multiple ways. The primary concern with collecting a culprit description is its potential to reduce culprit-presence detection. A decrease in guessing-based selection is not inherently negative because guessing occurs independently of the culprit's presence in the lineup, so that guessing-based selections have no evidentiary value. However, the decrease in culprit-presence detection as a function of providing culprit descriptions creates a fundamental dilemma: Although culprit descriptions are essential for locating possible suspects (Brown et al., 2008; Dodson et al., 2024; Wells et al., 2020), creating favorable conditions for culprit-presence detection is also crucial when eyewitness testimonies serve as key evidence. Given this dilemma, it may be important to increase awareness of the verbal overshadowing effect, which is often overlooked in practice (Brown et al., 2008).

However, as with any cognitive phenomenon studied under controlled conditions, generalization from experimental observations to practice requires caution. In the present study we implemented a 14-minute delay between viewing the crime and providing the description to maximize the likelihood of detecting the verbal overshadowing effect (Protzko et al., 2023). This interval also aligns with the typical delay with which the police arrives at the crime scene and begins interrogating victims and eyewitnesses for crimes comparable in severity to that shown in the staged-crime videos in the present experiment (e. g., Houston Police Department Office of Planning & Data Governance, 2023; Salt Lake City Police Department, 2022). Nevertheless, other aspects of our design such as the short delay between the description and the lineup are not representative of police practice.

In general, both the size and the robustness of the verbal overshadowing effect vary considerably across studies (Alogna et al., 2014; Chin & Schooler, 2008; Dodson et al., 2024; Finger & Pezdek, 1999; Holdstock et al., 2022; Meissner & Brigham, 2001; Schooler & Engstler-Schooler, 1990; Wilson et al., 2018), being influenced by factors such as the length of the crime-description delay, the length of the description-lineup delay, the instructions and the number of repetitions of the verbal description of the culprit. While our findings highlight the risks of obtaining culprit descriptions before a lineup, their implications for police procedures should therefore be considered with caution. Given that both culprit descriptions and lineups are standard components of police investigations, future research should test whether the present results can be replicated under conditions more typical of practical settings and explore interventions to minimize the effect of culprit descriptions on culprit-presence detection—for example, by emphasizing holistic rather than feature-based descriptions.

To summarize, the present study advances our knowledge about the verbal overshadowing effect by disentangling detection-based and non-detection-based processes underlying the effect using one single coherent and well-validated measurement model which takes into account the complete set of lineup response categories (Menne et al., 2022; Winter et al., 2022). The findings lead to clear conclusions about the cognitive processes underlying the verbal overshadowing effect in eyewitness identification: Providing culprit descriptions reduces both culprit-presence detection and guessing-based selection, indicating that verbalization impairs recognition and at the same time leads to more cautious responding.

Acknowledgements

We thank the German Research Foundation for funding this research (BU 945/10-2, project number 456214986).

References

Alogna, V. K., Attaya, M. K., Aucoin, P., Bahník, Š., Birch, S., Birt, A. R., Bornstein, B. H., Bouwmeester, S., Brandimonte, M. A., & Brown, C. (2014). Registered replication report: Schooler and Engstler-Schooler (1990). *Perspectives on Psychological Science*, *9*(5), 556-578. https://doi.org/10.1177/174569161454565

Bacharach, V. R., & Baker, M. A. (2024). Verbal overshadowing and decision criterion effects on recognition memory for faces. *Journal of Cognitive Psychology*, *36*(8), 881-897. https://doi.org/10.1080/20445911.2024.2419888

Baker, M. A., & Reysen, M. B. (2020). The influence of recall instruction type and length on the verbal overshadowing effect. *American Journal of Forensic Psychology*, *38*(3), 3-29.

Batchelder, W. H., & Riefer, D. M. (1986). The statistical analysis of a model for storage and retrieval processes in human memory. *British Journal of Mathematical and Statistical Psychology*, *39*(2), 129–149. https://doi.org/10.1111/j.2044-8317.1986.tb00852.x

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin and Review*, *6*(1), 57–86. https://doi.org/10.3758/BF03210812

Bell, R., Menne, N. M., Mayer, C., & Buchner, A. (2024). On the advantages of using AI-generated images of filler faces for creating fair lineups. *Scientific Reports*, *14*(1), 12304. https://doi.org/10.1038/s41598-024-63004-z

Brown, C., Lloyd-Jones, T. J., & Robinson, M. (2008). Eliciting person descriptions from eyewitnesses: A survey of police perceptions of eyewitness performance and reported use of interview techniques. *European Journal of Cognitive Psychology*, *20*(3), 529-560. https://doi.org/10.1080/09541440701728474

Cheung, O. S., & Gauthier, I. (2010). Selective interference on the holistic processing of faces in working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(2), 448-461. https://doi.org/10.1037/a0016471

Chin, J. M., & Schooler, J. W. (2008). Why do words hurt? Content, process, and criterion shift accounts of verbal overshadowing. *European Journal of Cognitive Psychology*, *20*(3), 396-413. https://doi.org/10.1080/09541440701728623

Clare, J., & Lewandowsky, S. (2004). Verbalizing facial memory: criterion effects in verbal overshadowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(4), 739-755. https://doi.org/10.1037/0278-7393.30.4.739

Colloff, M. F., Wade, K. A., & Strange, D. (2016). Unfair lineups make witnesses more likely to confuse innocent and guilty suspects. *Psychological science*, *27*(9), 1227-1239. https://doi.org/10.1177/0956797616655789

Dodson, C. S., Johnson, M. K., & Schooler, J. W. (1997). The verbal overshadowing effect: Why descriptions impair face recognition. *Memory & cognition*, *25*(2), 129-139. https://doi.org/10.3758/BF03201107

Dodson, C. S., Seale-Carlisle, T. M., Garrett, B. L., Kafadar, K., & Yaffe, J. (2024). Persistence of the verbal overshadowing and weapon-focus effects on lineup identification performance. *Journal of Applied Research in Memory and Cognition*. https://doi.org/10.1037/mac0000195

Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Zeitschrift für Psychologie/Journal of Psychology*, *217*(3), 108–124. https://doi.org/10.1027/0044-3409.217.3.108

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/https://doi.org/10.3758/BF03193146

Finger, K., & Pezdek, K. (1999). The effect of cognitive interview on face identification accuracy: Release from verbal overshadowing. *Journal of Applied Psychology*, *84*(3), 340-248. https://doi.org/10.1037/0021-9010.84.3.340

Fitzgerald, R. J., Tredoux, C. G., & Juncu, S. (2023). Estimation of eyewitness error rates in fair and biased lineups. *Law and Human Behavior*, *47*(4), 463-483. https://doi.org/10.1037/lhb0000538

Holdstock, J. S., Dalton, P., May, K. A., Boogert, S., & Mickes, L. (2022). Lineup identification in young and older witnesses: does describing the criminal help or hinder? *Cognitive Research: Principles and Implications*, *7*(1), 51. https://doi.org/10.1186/s41235-022-00399-1

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. https://doi.org/10.1037/h0093959

Houston Police Department Office of Planning & Data Governance. (2023). *Statistics. Monthy Operational Summary: May 2023*. Retrieved 29 Apr 2025 from https://www.houstontx.gov/police/department_reports/operational_summary/NIBRS_MonthlyOperationalSummary_May23.pdf

Itoh, Y. (2005). The facilitating effect of verbalization on the recognition memory of incidentally learned faces. *Journal of Applied Research in Memory and Cognition*, *19*(4), 421-433. https://doi.org/10.1002/acp.1069

Karageorge, A., & Zajac, R. (2011). Exploring the effects of age and delay on children's person identifications: Verbal descriptions, lineup performance, and the influence of wildcards. *British Journal of Psychology*, *102*(2), 161–183. https://doi.org/10.1348/000712610X507902

Lampinen, J. M., Race, B., Wolf, A. P., Phillips, P., Moriarty, N., & Smith, A. M. (2020). Comparing detailed and less detailed pre-lineup instructions. *Applied Cognitive Psychology*, *34*(2), 409–424. https://doi.org/10.1002/acp.3627

Leiner, D. J. (2022). *SoSci Survey [Computer software]*. https://www.soscisurvey.de

Malpass, R. S., & Devine, P. G. (1981). Eyewitness identification: Lineup instructions and the absence of the offender. *Journal of Applied Psychology*, *66*(4), 482–489. https://doi.org/10.1037/0021-9010.66.4.482

Mayer, C., Bell, R., Menne, N. M., & Buchner, A. (2024). Lineup position affects guessing-based selection but not culprit-presence detection in simultaneous and sequential lineups. *Scientific Reports*, *14*(1), 27642. https://doi.org/10.1038/s41598-024-78936-9

Mayer, C., Bell, R., Menne, N. M., Therre, A., Lichtenhagen, U., & Buchner, A. (in press). Evidence for age-related differences in culprit-presence detection and guessing-based selection in lineups. *Psychology and Aging*.

Meissner, C. A., & Brigham, J. C. (2001). A meta-analysis of the verbal overshadowing effect in face identification. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *15*(6), 603-616. https://doi.org/10.1002/acp.728

Meltzer, M. A., & Bartlett, J. C. (2019). Holistic processing and unitization in face recognition memory. *Journal of Experimental Psychology: General*, *148*(8), 1386-1406. https://doi.org/https://doi.org/10.1037/xge0000640

Memon, A., Hope, L., & Bull, R. (2003). Exposure duration: Effects on eyewitness accuracy and confidence. *British Journal of Psychology*, *94*(3), 339-354. https://doi.org/10.1348/000712603767876262

Menne, N., Bell, R., Winter, K., & Buchner, A. (2025). Disclosing the number of simultaneous lineups increases guessing-based selection in cases of multiple-culprit crimes. *Journal of Applied Research in Memory and Cognition* https://doi.org/10.1037/mac0000217

Menne, N. M., Winter, K., Bell, R., & Buchner, A. (2022). A validation of the two-high threshold eyewitness identification model by reanalyzing published data. *Scientific Reports*, *12*(1), 13379. https://doi.org/10.1038/s41598-022-17400-y

Menne, N. M., Winter, K., Bell, R., & Buchner, A. (2023a). The effects of lineup size on the processes underlying eyewitness decisions. *Scientific Reports*, *13*(1), 17190. https://doi.org/10.1038/s41598-023-44003-y

Menne, N. M., Winter, K., Bell, R., & Buchner, A. (2023b). Measuring lineup fairness from eyewitness identification data using a multinomial processing tree model. *Scientific Reports*, *13*(1), 6290. https://doi.org/10.1038/s41598-023-33101-6

Mickes, L., & Wixted, J. T. (2015). On the Applied Implications of the "Verbal Overshadowing Effect". *Perspectives on Psychological Science*, *10*(3), 400-403. https://doi.org/10.1177/1745691615576762

Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior research methods, instruments, & computers*, *36*, 630–633. https://doi.org/10.3758/BF03206543

Moshagen, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, *42*(1), 42–54. https://doi.org/10.3758/BRM.42.1.42

Oriet, C., & Fitzgerald, R. J. (2018). The single lineup paradigm: A new way to manipulate target presence in eyewitness identification experiments. *Law and Human Behavior*, *42*(1), 1-12. https://doi.org/10.1037/lhb0000272

Protzko, J., Lundmark, S., Walleczek, J., & Schooler, J. W. (2023). Nonlinear effect amplification: Differential susceptibility of verbal overshadowing as a function of time to interference. *Journal of Applied Research in Memory and Cognition*, *12*(4), 564-573. https://doi.org/10.1037/mac0000084

Quigley-McBride, A., & Wells, G. L. (2021). Methodological considerations in eyewitness identification experiments. In *Methods, measures, and theories in eyewitness identification tasks* (pp. 85-112). Routledge. https://doi.org/10.4324/9781003138105-8

Richler, J. J., Cheung, O. S., & Gauthier, I. (2011). Holistic Processing Predicts Face Recognition. *Psychological science*, *22*(4), 464-471. https://doi.org/10.1177/0956797611401753

Richler, J. J., & Gauthier, I. (2014). A meta-analysis and review of holistic face processing. *Psychological bulletin*, *140*(5), 1281–1302. https://doi.org/10.1037/a0037004

Salt Lake City Police Department. (2022). *SLCPD Releases Response Time Data for January 2022*. Retrieved 29 Apr 2025 from https://police.slc.gov/2022/02/01/slcpd-releases-response-time-data-for-january-2022

Schmidt, O., Erdfelder, E., & Heck, D. W. (2023). How to develop, test, and extend multinomial processing tree models: A tutorial. *Psychological Methods*. https://doi.org/10.1037/met0000561

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology*, *22*(1), 36-71. https://doi.org/https://doi.org/10.1016/0010-0285(90)90003-M

Smith, A. M. (2020). Why do mistaken identification rates increase when either witnessing or testing conditions get worse? *Journal of Applied Research in Memory and Cognition*, *9*(4), 495-507. https://doi.org/10.1016/j.jarmac.2020.08.002

Smith, A. M., Wells, G. L., Lindsay, R., & Penrod, S. D. (2017). Fair lineups are better than biased lineups and showups, but not because they increase underlying discriminability. *Law and Human Behavior*, *41*(2), 127-145. https://doi.org/10.1037/lhb0000219

Smith, H. M. J., & Flowe, H. D. (2015). ROC Analysis of the Verbal Overshadowing Effect: Testing the Effect of Verbalisation on Memory Sensitivity. *Applied Cognitive Psychology*, *29*(2), 159-168. https://doi.org/https://doi.org/10.1002/acp.3096

Sporer, S. L., Kaminski, K. S., Davids, M. C., & McQuiston, D. (2016). The verbal facilitation effect: re-reading person descriptions as a system variable to improve identification performance. *Memory*, *24*(10), 1329-1344. https://doi.org/10.1080/09658211.2015.1106561

Therre, A., Bell, R., Menne, N. M., Mayer, C., Lichtenhagen, U., & Buchner, A. (2024). On the possible advantages of combining small lineups with instructions that discourage guessing-based selection. *Scientific Reports*, *14*(1), 14126. https://doi.org/10.1038/s41598-024-64768-0

Wells, G. L., Kovera, M. B., Douglass, A. B., Brewer, N., Meissner, C. A., & Wixted, J. T. (2020). Policy and procedure recommendations for the collection and preservation of eyewitness identification evidence. *Law and Human Behavior*, *44*(1), 3–36. https://doi.org/10.1037/lhb0000359

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. https://doi.org/10.1016/j.jarmac.2015.08.008

Wetmore, S. A., Neuschatz, J. S., Gronlund, S. D., Wooten, A., Goodsell, C. A., & Carlson, C. A. (2015). Effect of retention interval on showup and lineup performance. *Journal of Applied Research in Memory and Cognition*, *4*(1), 8–14. https://doi.org/10.1016/j.jarmac.2014.07.003

Wickham, L. H. V., & Lander, K. (2008). The effect of verbal description and processing type on face identification. *European Journal of Cognitive Psychology*, *20*(3), 577-586. https://doi.org/10.1080/09541440701728433

Wilcock, R., & Bull, R. (2010). Novel lineup methods for improving the performance of older eyewitnesses. *Applied Cognitive Psychology*, *24*(5), 718–736. https://doi.org/10.1002/acp.1582

Wilson, B. M., Seale-Carlisle, T. M., & Mickes, L. (2018). The effects of verbal descriptions on performance in lineups and showups. *Journal of Experimental Psychology: General*, *147*(1), 113-124. https://doi.org/10.1037/xge0000354

Winter, K., Menne, N. M., Bell, R., & Buchner, A. (2022). Experimental validation of a multinomial processing tree model for analyzing eyewitness identification decisions. *Scientific Reports*, *12*(1), 15571. https://doi.org/10.1038/s41598-022-19513-w

Winter, K., Menne, N. M., Bell, R., & Buchner, A. (2023). Evaluating the impact of first-yes-counts instructions on eyewitness performance using the two-high threshold eyewitness identification model. *Scientific Reports*, *13*(1), 6572. https://doi.org/10.1038/s41598-023-33424-4

Declaration of the independent contribution

My dissertation includes three articles, with four experiments in total. Of these, one has been published and two have been submitted for publication in academic journals with an established peer-review process. In the following, I will elaborate how each author individually contributed to the article. The majority of work was always carried out by the first author of the article.

Independent contribution to Article 1

Publication: Therre, A., Bell, R., Menne, N. M., Mayer, C., Lichtenhagen, U., & Buchner, A. (2024). On the possible advantages of combining small lineups with instructions that discourage guessing-based selection. *Scientific Reports*, *14*(1), 14126. https://doi.org/10.1038/s41598-024-64768-0

Study conception: I developed the experimental design of Experiment 1a and 1b with support from Axel Buchner, Raoul Bell, Nicola Marie Menne, Carolin Mayer and Ulla Lichtenhagen.

Implementation: I programmed and conducted Experiment 1a and Experiment 1b with feedback from Axel Buchner, Raoul Bell, Nicola Marie Menne, Carolin Mayer and Ulla Lichtenhagen.

Data analysis: I conducted the statistical analyses independently. Axel Buchner, Raoul Bell, Nicola Marie Menne, Carolin Mayer and Ulla Lichtenhagen reviewed their accuracy.

Manuscript: I prepared the manuscript independently, including an extensive literature review, the design of the figures and the composition and writing of the manuscript. Axel Buchner, Raoul Bell, Nicola Marie Menne, Carolin Mayer and Ulla Lichtenhagen provided feedback which I incorporated after a thorough review and subsequent consultation. I managed the peer-review process through the academic journal. During this process, I made revisions with support from Axel Buchner, Raoul Bell, Nicola Marie Menne, Carolin Mayer and Ulla Lichtenhagen.

Independent contribution to Article 2

Article Submitted for Publication: Therre, A., Bell, R., Menne, N. M., Mayer, C., Lichtenhagen, U., & Buchner, A. (Under review). Delays reduce culprit-presence detection but do not affect guessing-based selection in response to lineups. Manuscript under review at *Scientific Reports*.

Study conception: I developed the experimental design of Experiment 2 with support from Axel Buchner, Raoul Bell, Nicola Marie Menne, Carolin Mayer and Ulla Lichtenhagen.

Implementation: I programmed and conducted Experiment 2 with feedback from Axel Buchner, Raoul Bell, Nicola Marie Menne, Carolin Mayer and Ulla Lichtenhagen.

Data analysis: I conducted the statistical analyses independently. Axel Buchner, Raoul Bell, Nicola Marie Menne, Carolin Mayer and Ulla Lichtenhagen reviewed their accuracy.

Manuscript: I prepared the manuscript independently, including an extensive literature review, the design of the figures and the composition and writing of the manuscript. Axel Buchner, Raoul Bell, Nicola Marie Menne, Carolin Mayer and Ulla Lichtenhagen provided feedback which I incorporated after a thorough review and subsequent consultation. I managed the peer-review process through the academic journal. During this process, I made revisions with support from Axel Buchner, Raoul Bell, Nicola Marie Menne, Carolin Mayer and Ulla Lichtenhagen.

Independent contribution to Article 3

Article Submitted for Publication: Therre, A., Bell, R., Menne, N. M., Mayer, C., Lichtenhagen, U., & Buchner, A. (Under review). On the cognitive processes underlying the verbal overshadowing effect: Culprit descriptions reduce culprit-presence detection and guessing-based selection in eyewitness responses to lineups. Manuscript under review at *Journal of Applied Research in Memory and Cognition*.

Study conception: I developed the experimental design of Experiment 3 with support from Axel Buchner, Raoul Bell, Nicola Marie Menne, Carolin Mayer and Ulla Lichtenhagen.

Implementation: I programmed and conducted Experiment 3 with feedback from Axel Buchner, Raoul Bell, Nicola Marie Menne, Carolin Mayer and Ulla Lichtenhagen.

Data analysis: I conducted the statistical analyses independently. Axel Buchner, Raoul Bell, Nicola Marie Menne, Carolin Mayer and Ulla Lichtenhagen reviewed their accuracy.

Manuscript: I prepared the manuscript independently, including an extensive literature review, the design of the figures and the composition and writing of the manuscript. Axel Buchner, Raoul Bell, Nicola Marie Menne, Carolin Mayer and Ulla Lichtenhagen provided feedback which I incorporated after a thorough review and subsequent consultation. I managed the peer-review process through the academic journal. During this process, I made revisions with support from Axel Buchner, Raoul Bell, Nicola Marie Menne, Carolin Mayer and Ulla Lichtenhagen.

Erklärung an Eides statt

Ich versichere an Eides statt, dass die Dissertation von mir selbständig und ohne unzulässige fremde Hilfe unter Beachtung der „Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf" erstellt worden ist.

Ich versichere insbesondere:

(1) Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt.

(2) Alle wörtlich oder dem Sinn nach aus anderen Texten entnommenen Stellen habe ich als solche kenntlich gemacht; dies gilt für gedruckte Texte ebenso wie für elektronische Ressourcen.

(3) Die Arbeit habe ich in der vorliegenden oder einer modifizierten Form noch nicht als Dissertation vorgelegt—-sei es an der Heinrich-Heine-Universität Düsseldorf oder an einer anderen Universität.

Düsseldorf, 3. Juni 2025

Amelie Frederike Therre