Eyewitnesses' General Metamemory Beliefs Do Not Predict Culprit-Presence Detection

Marie Luisa Schaper, Nicola Marie Menne, Raoul Bell, Carolin Mayer, Axel Buchner

Article - Version of Record

Suggested Citation:
Schaper, M. L., Menne, N. M., Bell, R., Mayer, C., & Buchner, A. (2025). Eyewitnesses' General Metamemory Beliefs Do Not Predict Culprit-Presence Detection. Applied Cognitive Psychology, 39(4), Article e70100. https://doi.org/10.1002/acp.70100

Applied Cognitive Psychology

WILEY

RESEARCH ARTICLE OPEN ACCESS

Eyewitnesses' General Metamemory Beliefs Do Not Predict Culprit-Presence Detection

Marie Luisa Schaper | Nicola Marie Menne | Raoul Bell | Carolin Mayer | Axel Buchner

Department of Experimental Psychology, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

Correspondence: Marie Luisa Schaper (marie.schaper@hhu.de)

Received: 17 January 2025 | **Revised:** 7 July 2025 | **Accepted:** 13 July 2025

Funding: This work was supported by Deutsche Forschungsgemeinschaft, 456214986.

Keywords: culprit-presence detection | eyewitness memory | metacognition | metamemory | multinomial modeling

ABSTRACT

If eyewitnesses' general beliefs about their memory predicted whether they detect the culprit in a lineup, it could be useful in legal investigations to systematically assess general metamemory beliefs. Using a process-oriented approach, we tested this hypothesis in two experiments. In Experiment 1, 1627 participants responded to either several metamemory-belief questionnaires or, in the control condition, to a personality questionnaire, then watched a video of a staged crime and finally were presented with four lineups. In Experiment 2, 1467 participants first watched the staged-crime video, then responded to the questionnaires and finally were presented with the lineups. Using hierarchical multinomial modeling, we tested whether general metamemory beliefs were associated with culprit-presence detection. The results of both experiments provide evidence against the hypothesis that general metamemory beliefs predict culprit-presence detection. Accordingly, we caution against using general metamemory beliefs as indicators of how well a culprit can be detected.

General beliefs about how good or bad one's memory is represent a critical aspect of metamemory (for an overview, see Tauber and Dunlosky 2016). The purpose of the present research was to test whether systematically assessing these general metamemory beliefs could be useful in the context of police lineups. In a lineup, a suspect—who may be guilty or innocent—is presented alongside several known-to-be-innocent fillers. The eyewitnesses' task is to make an identification if they detect the culprit in the lineup or to reject the lineup if they do not. If people's general metamemory beliefs about how well their memory functions predicted whether they are likely to detect the culprit in the lineup, the systematic assessment of general metamemory beliefs could be valuable in police investigations and legal trials. For instance, one may consider not performing lineups with eyewitnesses who believe they are poor eyewitnesses if this actually implies that they are indeed unlikely to detect the culprit in the lineup. Therefore, the present research set out to test whether general metamemory beliefs are related to the detection of the presence of a culprit in a lineup.

Before going into more detail, it seems necessary to clarify that general metamemory beliefs are not to be confused with confidence judgments. Just like general metamemory beliefs, confidence judgments are an aspect of metamemory; however, other than general metamemory beliefs, confidence judgments pertain to specific evaluations of past memory processes. In the context of eyewitness responses to lineups, confidence judgments reflect how certain an eyewitness is that a preceding response to a lineup is accurate. When lineups are administered under ideal conditions, confidence judgments may well be useful in the context of police lineups (e.g., Saraiva et al. 2020; Wixted and Mickes 2022; Wixted and Wells 2017), even though this is still critically discussed (e.g., Berkowitz et al. 2022; Sauer et al. 2019; Smith et al. 2021). In contrast, whether general metamemory beliefs predict the detection of the culprit in a lineup is less well researched and is thus the focus of the research reported here.

There are indeed instances in which people are able to predict how well their memory functions. For example, people are reasonably good at predicting that repeatedly studied word pairs will be remembered better than word pairs studied only once (Koriat 1997). When such predictions about memory are made on an item-by-item basis (Rhodes 2016), as in the instance just mentioned, the standard of comparison for judging how well a particular item can be remembered is how well other items can be remembered by the same person. The standard of comparison is thus directly available to the person making the predictions, which should facilitate making metamemory-based predictions. However, even under these favorable conditions, metamemory-based predictions are not always accurate (Bindemann et al. 2014; Frank and Kuhlmann 2017; Luna and Albuquerque 2022; Mieth et al. 2021; Schaper et al. 2019b; Undorf and Zimdahl 2019). For instance, people believe that their memory is better for items paired with an expected source (kitchen-oven) than for items paired with an unexpected source (kitchen-toothpaste) even though memory is actually enhanced for the unexpected source (Schaper et al. 2019b). Metamemory-based predictions may be even less likely to be accurate when individuals are asked to assess their general memory abilities in comparison to the memory abilities of others—for example, when judging whether they have better or worse eyewitness memory than the average person. In such cases, the standard of comparison must be inferred indirectly from one's own memory of other people's behavior in tasks involving memory, the latter of which may or may not be related to the task faced when an eyewitness is asked to identify a culprit in a lineup (for a general discussion of the relationship between metamemory questionnaires and memory, see Gopi and Madan 2024), which may be particularly challenging. In line with these considerations, it has been found that responses to metamemory questionnaires in which one's face recognition ability is to be judged relative to other people's face recognition ability are unrelated to actual face recognition performance (Bindemann et al. 2014). Overall, these findings raise the question of whether people's general metamemory beliefs, particularly those reflecting a comparison of one's own memory to that of others, can meaningfully predict differences in the detection of the culprit in a lineup.

In the context of the present research question, one specific aspect of a study by Saraiva et al. (2020) is of particular relevance. In their study, participants completed three different metamemory questionnaires, including the Eyewitness Metamemory Scale (Saraiva et al. 2019), before watching a staged-crime video and responding to either fair or unfair lineups. The Eyewitness Metamemory Scale was specifically designed to assess eyewitnesses' general metamemory beliefs, such as whether one considers oneself a good eyewitness or whether one focuses on specific facial features such as nose and eyes when paying attention to a face. The authors examined whether the scores on seven metamemory (sub)scales predicted the accuracy of responses to lineups, analyzing data separately for fair and unfair lineups and for participants who did or did not select a face. Significant predictive relationships emerged in two out of 28 instances: The score of the Discontentment subscale of the Eyewitness Metamemory Scale (Saraiva et al. 2019)—which includes items such as "My ability to remember faces is much worse than other people's ability to remember faces" (Item 12, see Saraiva et al. 2019)—negatively predicted response accuracy of participants who selected a face in fair lineups and the same was true for the response accuracy of participants who selected a face in unfair lineups (Saraiva et al. 2020, tables 5 and 6). In other words, participants who indicated to be more discontent with their memory tended to be less likely to select the correct face.

Saraiva et al. (2020) noted that their study provided initial evidence for a potential relationship between general-metamemory-belief questionnaires and the accuracy of lineup responses, but they also acknowledged the need for further research (p. 127). Although their hypotheses were tested in a rigorous design, a general statistical consideration reinforces the call for additional research: A total of 28 statistical tests were reported, of which two—approximately 7%—were statistically significant. Although the issue of alpha-error inflation due to multiple testing was carefully addressed, keeping the actual alpha-error level at the nominal level of 5%, the percentage of significant results of statistical tests being close to this alpha-error level raises the possibility that the observed effects could have occurred by chance. Further, while the work by Saraiva et al. (2020) provides an important starting point for examining the role of general metamemory beliefs in the context of lineups, they focused on the relationship between metamemory and response accuracy. Response accuracy was calculated by coding each lineup response as correct or incorrect, depending on whether participants accurately identified the culprit in culprit-present lineups or correctly rejected the lineup in culprit-absent lineups. The current study builds on this work by taking a more process-oriented approach: Instead of asking whether general metamemory beliefs predict response accuracy, we specifically investigate whether these beliefs are related to the detection of the culprit in a lineup.

When observable responses from memory tasks are used as the unit of analyses, one puts up with the fact that these observable responses typically represent the result of a conglomerate of several different types of underlying cognitive processes. For instance, a correct culprit identification may be caused by a memory-based detection of the presence of the culprit, by a biased selection of the culprit due to the fact that the culprit stands out from the fillers, or by a selection of the culprit based on guessing. In the current study, our goal was to directly examine the relationship between general metamemory beliefs and the process of detecting the presence of the culprit in a lineup. Testing this relationship at the process level is interesting because culprit-presence detection is diagnostically relevant in the determination of a suspect's guilt. In contrast, guessing-based selection, which may also lead to an identification of the culprit, occurs irrespective of the culprit's actual presence in the lineup and therefore lacks diagnostic value. For instance, if general metamemory beliefs were related to culprit identifications because individuals who believe they have good memory are actually better at detecting the culprit in the lineup, it could be useful to consider general metamemory beliefs in lineup contexts. By contrast, if general metamemory beliefs were associated with culprit identifications only because individuals with overly optimistic beliefs in the quality of their memory are more likely to make a guessing-based selection which occasionally

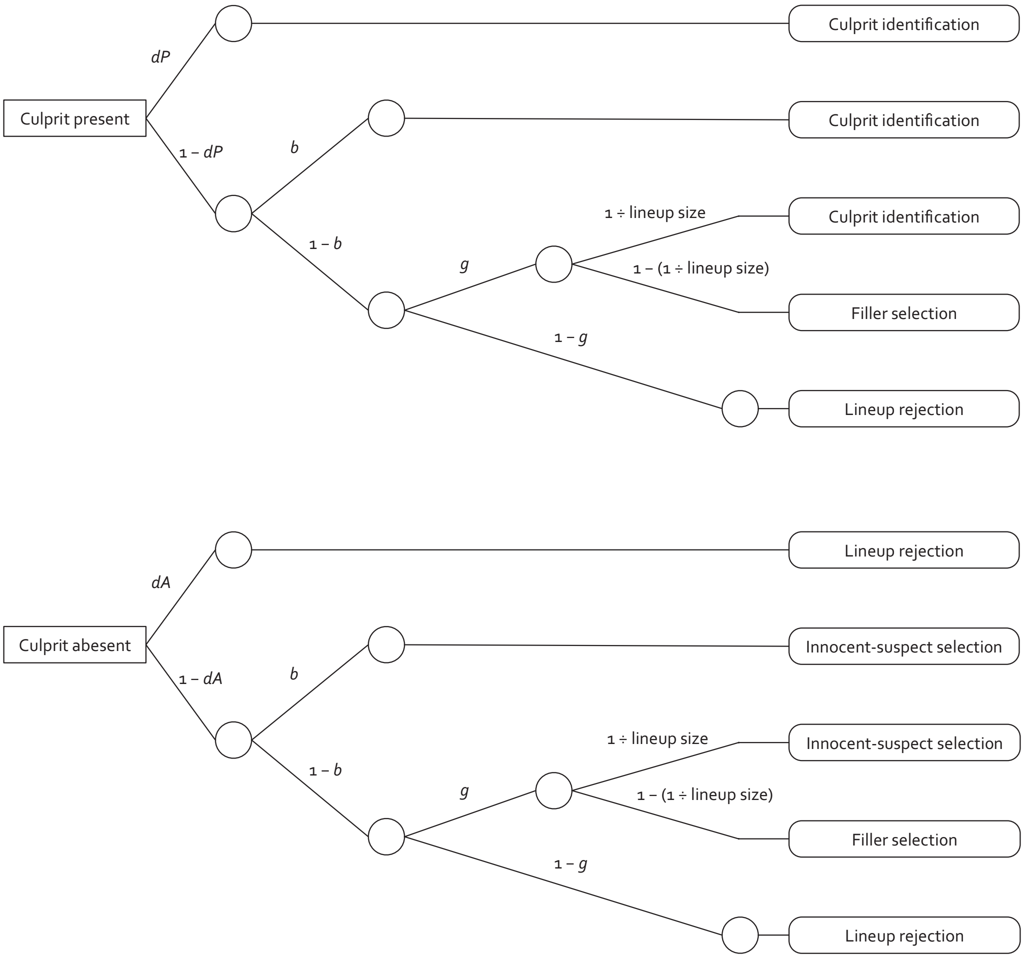

FIGURE 1 | Processing-tree illustration of the 2-HT eyewitness identification model. The rectangles on the left represent the stimuli that are presented: Culprit-present lineups (upper tree) and culprit-absent lineups (lower tree). The rounded rectangles on the right represent the response categories observed in these lineups. The letters along the branches represent the probabilities of the postulated latent processes (*dP*: Culprit-presence detection; *b*: Biased suspect selection; *g*: Guessing-based selection among the lineup members; *dA*: Culprit-absence detection). The sampling probability of randomly selecting the suspect among the lineup members in case of a guessing-based selection is given by 1 ÷ lineup size.

leads to the identification of the culprit, then general metamemory beliefs would not be useful. From this perspective, general metamemory-belief questionnaires would only be of diagnostic value if they specifically predicted culprit-presence detection. To disentangle the processes involved in lineup responses, we used the well-validated two-high threshold (2-HT) eyewitness identification model (Menne et al. 2022; Winter et al. 2022). With this multinomial measurement model, it is possible to estimate the probabilities of four latent cognitive processes contributing to observable lineup responses: two memory-based processes—culprit-presence detection and culprit-absence detection—and two non-memory-based processes—biased suspect selection and guessing-based selection. The estimates are based on all response categories observed in lineups: correct culprit identifications, false selections of innocent suspects, false and correct rejections of lineups, as well as filler selections in culprit-present and culprit-absent lineups. Thus, the 2-HT eyewitness identification model does not require collapsing across different response categories with distinct underlying processes, unlike, for example, Receiver Operating Characteristic analysis, which has been criticized for relying on such an aggregation in the context of lineup data (Smith et al. 2017; Wells, Smalarz, and Smith 2015; Wells, Smith, and Smalarz 2015). For example, in culprit-absent lineups, filler selections represent false responses while lineup rejections are correct responses, reflecting fundamentally different underlying processes. Another important aspect is that the 2-HT eyewitness identification model has been extensively validated both in a series of validation experiments specifically designed for this purpose (Winter et al. 2022) and in a series of eight 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 and Zajac 2011; Lampinen et al. 2020; Malpass and Devine 1981; Memon et al. 2003; Smith 2020; Wetmore et al. 2015; Wilcock and Bull 2010). Notably, the culprit-presence detection parameter *dP* has been shown to respond sensitively to manipulations that can be expected to affect this process, such as exposure time and viewing conditions. For those unfamiliar with multinomial measurement models, easy-to-read introductory texts and comprehensive reviews exist (Erdfelder et al. 2009; Schmidt et al. 2023).

The 2-HT eyewitness identification model is illustrated in Figure 1. The processes of culprit-presence detection, culprit-absence detection, biased suspect selection and guessing-based selection are precisely and transparently defined by the model's structure. The verbal labels serve as accessible descriptors to simplify communication. The upper tree in Figure 1 illustrates the processes assumed to occur when eyewitnesses are presented with culprit-present lineups. The presence of the culprit is detected with probability dP, resulting in a correct culprit identification. If the presence of the culprit is not detected which occurs with the complementary probability $1-dP$, the culprit can still be selected based on two non-memory-based processes. Conditional on the culprit's presence not being detected ($1-dP$), biased suspect selection occurs with probability b. The unconditional probability that a lineup member is selected due to biased suspect selection in culprit-present lineups is thus given by $(1-dP) \cdot b$. Parameter b is larger than zero if the photograph of the culprit stands out because it differs in some characteristic (e.g., a distinctive facial feature, color, contrast or lighting) from the photographs of the fillers and if the difference is picked up by the eyewitness. Parameter b thus captures lineup unfairness and allows for the other model parameters to be determined without being contaminated by lineup unfairness (Menne et al. 2022, 2023b; Winter et al. 2022). With the complementary probability $1-b$, no biased suspect selection occurs. Conditional on the culprit's presence not being detected ($1-dP$) and on the suspect not being selected due to lineup unfairness ($1-b$), eyewitnesses may select one of the lineup members based on guessing with probability g. The unconditional probability that a lineup member is selected via guessing-based selection in culprit-present lineups is thus given by $(1-dP) \cdot (1-b) \cdot g$. Guessing-based selection is defined as a selection process that is neither driven by detection nor by bias, implying that there is no systematic preference for selecting the suspect over the fillers. Following guessing-based selection, the sampling probability of selecting the suspect in a six-person lineup is thus $0.1\overline{6}$ while the complementary probability of selecting one of the fillers is $0.8\overline{3}$, illustrating that, as Wixted and Wells (2017) have pointed out, "a witness who chooses randomly is far more likely to land on a filler than the suspect" (p. 11). With probability $1-g$, the lineup is rejected. The lower tree depicted in Figure 1 illustrates the processes assumed to occur when eyewitnesses are presented with culprit-absent lineups. The absence of the culprit is detected with probability dA, resulting in a correct lineup rejection. If the absence of the culprit is not detected, which occurs with the complementary probability $1-dA$, then biased suspect selection (which is conditional on the absence of the culprit not being detected) and guessing-based selection (which is conditional on the absence of the culprit not being detected and conditional on the suspect not being selected due to lineup unfairness) occur with the same probabilities as in culprit-present lineups. However, in this case, these processes may lead to the selection of the innocent suspect in lieu of the culprit, or to filler selections and lineup rejections. Thus, in culprit-absent lineups, the unconditional probability that a lineup member is selected due to biased suspect selection is given by $(1-dA) \cdot b$ and the unconditional probability that a lineup member is selected via guessing-based selection is given by $(1-dA) \cdot (1-b) \cdot g$.

One aim of the present experiments was to test whether general metamemory beliefs predict culprit-presence detection, represented by parameter dP. To this end, we used the same metamemory questionnaires as Saraiva et al. (2020). However, instead of focusing on lineup responses as the unit of analysis, we examined the underlying process of culprit-presence detection, as measured via the 2-HT eyewitness identification model. We conducted two experiments: one in which participants first responded to the metamemory questionnaires and then watched the staged-crime video, as in Saraiva et al. (2020), and another one in which this order was reversed. Additionally, we tested whether the results depended on the lineup presentation format: Participants were either presented with simultaneous lineups (i.e., all faces in the lineup presented at the same time) or sequential lineups (i.e., all faces in the lineup presented one after the other; both methods are well-established, see Fitzgerald et al. 2021).

It is also important to consider how the assessment of general metamemory beliefs might reactively affect eyewitness memory. For example, Saraiva et al. (2020) acknowledged the possibility that, in their study, responding to metamemory questionnaires prior to watching a staged-crime video may have reactively enhanced encoding of the culprit's face. This concern is plausible insofar as reactivity of metamemory assessment has been observed in various memory paradigms (e.g., Besken 2016; Besken and Mulligan 2013, 2014; Mieth et al. 2021; Mitchum et al. 2016; Myers et al. 2020; Schaper et al. 2019a; Senkova and Otani 2021; Soderstrom et al. 2015). Such reactivity may occur, for instance, when responding to metamemory questionnaires draws attention to certain aspects of the task that otherwise would be less well attended. Consider, for instance, Item 22 of the Eyewitness Metamemory Scale (Saraiva et al. 2019): "I often focus on specific facial features such as nose and eyes when I am paying attention to a face". Responding to this item may influence how participants attend to faces in a subsequent staged-crime video and, as a consequence, how well they later detect those faces in a lineup. To test for such reactive effects, both experiments included a control condition: one group of participants responded to the metamemory questionnaires, whereas another group of participants responded to an equal number of items taken from the HEXACO personality questionnaire (Ashton and Lee 2007; Ashton et al. 2007; Lee and Ashton 2018). In Experiment 1, participants first responded to the metamemory questionnaires, then watched the staged-crime video and were subsequently presented with lineups (directly replicating the procedure used by Saraiva et al. 2020), allowing to examine whether responding to metamemory questionnaires may affect encoding. In Experiment 2, in contrast, participants first watched the staged-crime video and only then responded to the metamemory questionnaire, prior to lineup presentation. This reversed order allowed us to test whether responding to metamemory questionnaires might reactively affect retrieval, for example by encouraging retrieval practice or increasing retrieval effort at test.

1 | Experiment 1

1.1 | Method

1.1.1 | Participants

The experiment was conducted online on the SoSci Survey platform (https://www.soscisurvey.de). Participants were recruited using the research panel of Bilendi (https://www.bilendi.de), a panel provider certified under ISO 20252. We aimed at a sample size of at least 1500 valid data sets and ended data collection at the end of the day this criterion was reached. Of the 1897 data sets of the participants who had responded to the socio-demographic questionnaire at the start of the experiment, 221 had to be excluded from the analyses because the participants had not completed the experiment or had withdrawn their consent to the use of their data. Data of 19 participants had to be excluded because they had participated and seen the video of the culprits twice (only the data from the second participation were excluded). Data of 30 participants had to be excluded because they had failed the attention check (see *Procedure* section). The final data set contained data of 1627 participants (706 female, 918 male, 3 non-binary) with a mean age of 50 (SD = 16) years. Participants were randomly assigned to one of the four groups: the simultaneous-lineups metamemory group (414 participants), the simultaneous-lineups control group (417 participants), the sequential-lineups metamemory group (387 participants) and the sequential-lineups control group (409 participants).

1.1.2 | Ethics Statement

Approval was received from the local ethics committee for a series of experiments to which the present experiments belong. Informed consent was obtained from all participants prior to participation. They were informed that they would be asked to answer several questions and would then see a short video in which several persons would physically and verbally abuse another person. Participants were asked not to proceed if they felt uncomfortable expecting to watch such a video. At the end of the experiment, participants were thanked for their participation and the purpose of the experiment was disclosed, as was the fact that the crime depicted in the video had been staged.

1.1.3 | Materials

We used the same staged-crime videos and facial photographs of the suspects and fillers as in previous studies (Bell et al. 2024; Mayer et al. 2024; Menne et al. 2025, 2023a, b; Therre et al. 2024; Winter et al. 2022, 2023). We repeat the description of these materials here for convenience. In each of two staged-crime videos, four different culprits, dressed in fan clothing of the German soccer team FC Bayern München, insulted and beat a victim, dressed in fan clothing of the German soccer team Borussia Dortmund. The sequence and timing of events were the same in both videos, but the culprits and the victims were played by different actors who were chosen to match in basic features such as hair color, hair style, and body type. Specifically, Actor A in Video 1 matched Actor A in Video 2, Actor B in Video 1 matched Actor B in Video 2, and so on. Including four culprits allowed us to obtain four data points per participant, thereby enhancing parameter-estimation precision in our analyses. This procedure is ecologically valid as a notable share of real-world crimes involves multiple culprits (Hobson and Wilcock 2011; Hobson et al. 2012; Tupper et al. 2019). Responding to multiple lineups after having witnessed a multiple-culprit crime may be more demanding than responding to a single lineup after having witnessed a single-culprit crime (Lockamyeir et al. 2021). The videos had a resolution of 885 × 500 pixels and a duration of 130 s. In both videos, the culprits' faces were clearly visible from multiple angles.

Color portrait photographs taken from each of the culprits of each video served as suspect photographs. The facial filler photographs were taken from the Center for Vital Longevity face database (Minear and Park 2004). Each set of five filler photographs roughly matched both suspects of a suspect pair (e.g., the suspects played by Actor A in Video 1 and by Actor A in Video 2) in basic physical features such as hair color, hair style, and body type. The suspect and filler photographs showed the faces from a frontal view with a neutral facial expression. Using Affinity Photo, the faces were cropped from the original images, placed on a black background, and adjusted in color and contrast. All photographs were shown with a resolution of 142 × 214 pixels to facilitate the online presentation of the faces.

We used the same three metamemory questionnaires as Saraiva et al. (2020) to measure general metamemory beliefs: First, the Eyewitness Metamemory Scale (Saraiva et al. 2019) is designed to assess general metamemory beliefs specifically for face memory and eyewitness identification settings with three subscales. The Contentment subscale is related to positive self-perceptions of memory (10 items, e.g., "My ability to remember faces is much better than other people's ability to remember faces"). The Discontentment subscale is related to negative self-perceptions of memory (8 items, e.g., "Compared to other people, I think I would be a much worse eyewitness"). The Strategy subscale is related to the use of mnemonic strategies (5 items, e.g., "Compared to other people, I more often use a strategy [e.g., focus on specific facial features such as eyes] to remember a person's face"). Participants responded to all items of the Eyewitness Memory Scale by indicating their level of agreement or disagreement using a 7-point scale ranging from "strongly disagree" to "strongly agree". Second, the Multifactorial Memory Questionnaire (Troyer and Rich 2002) is designed to assess general metamemory beliefs on three subscales. The Contentment subscale is related to positive and negative self-perceptions of memory (18 items, e.g., "I am generally pleased with my memory ability"). Participants responded to the Contentment subscale by indicating their level of agreement or disagreement using a 5-point scale ranging from "strongly disagree" to "strongly agree". The Ability subscale is related to self-appraisal of one's memory capabilities (20 items, e.g., "How often do you forget to pay a bill on time?"). Participants responded to the Ability subscale by indicating how often the event described in an item occurred during the 2 weeks preceding the assessment using a 5-point scale ranging from "always" to "never". The Strategy subscale is related to the use of strategies (14 items, e.g., "How often do you use a timer or alarm to remind you when to do something?"; the final five items on this subscale were omitted due to an implementation error). Participants responded to the Strategy subscale by indicating how often the event described in an item occurred during the 2 weeks preceding the assessment using a 5-point scale ranging from "never" to "always". Third, the Squire Subjective Memory Questionnaire (van Bergen et al. 2010) is designed to measure distrust in one's memory (18 items, e.g., "My ability to search through my mind and recall names or memories I know are there is…"). Participants responded to the Squire Subjective Memory Questionnaire by indicating their self-assessment using a 9-point scale ranging from "worse than ever before" to "better than ever before". The metamemory questionnaires were translated into German by the current authors. In the control group, we used an equal number of items (i.e., 93 and thus all but the final seven items) taken from the 100-item HEXACO personality questionnaire (Ashton and Lee 2007; Ashton et al. 2007; Lee and Ashton 2018) which measures personality on six subscales. The German version of the HEXACO was retrieved in April 2023 from https://hexaco.org/hexaco-inventory. The results of the HEXACO are irrelevant to the current research question and are therefore not reported. The questionnaires were implemented in the online platform used for the present experiments.

1.1.4 | Procedure

Apart from administering the questionnaires, we used essentially the same procedure as in previous studies (Bell et al. 2024; Mayer et al. 2024; Menne et al. 2025, 2023a, b; Therre et al. 2024; Winter et al. 2022, 2023). We repeat the description of the procedure here for convenience.

After having provided informed consent and having indicated their age, gender, and education level, participants were randomly assigned to one of the four groups. Participants in the metamemory groups responded first to the Eyewitness Metamemory Scale (Saraiva et al. 2019), then to the Multifactorial Memory Questionnaire (Troyer and Rich 2002) and then to the Squire Subjective Memory Questionnaire (van Bergen et al. 2010). All participants in the control groups responded to the same sequence of items of the HEXACO personality questionnaire.

Next, participants watched one of the two staged-crime videos described above. After the video, participants were asked to answer an attention-check question. They passed if they selected, from a randomized list of 10 alternatives, the answer that the video had shown soccer fans.

Participants were then shown four separate lineups in a random order without being informed about the number of lineups to expect. Participants were asked to identify the culprit if present or to reject the lineup if the culprit was absent. Each lineup consisted of six photographs, one photograph of the suspect's face and five filler photographs. The positions of the photographs in the lineups were randomly determined. Two randomly selected lineups contained a photograph of a culprit and two contained a photograph of an innocent suspect. The innocent suspect was one of the actors from the video participants had not seen. For instance, if participants had seen Video 1, then two randomly selected actors from Video 1 (e.g., Actor A and Actor D) served as culprits while two actors from Video 2 (in this example, Actor B and Actor C) served as innocent suspects. Each actor was equally likely to be presented as the culprit or innocent suspect. This so-called *crossed-lineup procedure* ensures that, on average, the culprits and innocent suspects differ to the same degree from the fillers. The crossed-lineup procedure also has the advantage that culprit-absent lineups contain a designated innocent suspect to whom the fillers are matched, which represents a more ecologically valid lineup procedure than using only fillers in culprit-absent lineups. In practice, the photographs of the suspects (whose guilt or innocence is unknown) are typically taken from other sources (e.g., social media) than the photographs of the fillers, which are usually taken from face databases. Another advantage of the crossed-lineup procedure is that culprit-present and culprit-absent lineups include the same filler faces; only the identity of the suspect changes. Which of the two suspects served as the culprit and which as the innocent suspect depended on the random assignment to one of the two videos (see above). In that way, it was ensured that, on average, the degree of fairness was the same in culprit-present and culprit-absent lineups.

Before completing the lineup task, participants received the following instructions: "In the video, you saw aggressive Bayern München fans. You are now asked to identify them. For this purpose, 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 aggressive Bayern München fans. It is also possible that no one in the lineup is one of the Bayern München fans." Participants were then instructed on how make a response depending on the lineup condition.

In simultaneous lineups, the six photographs were shown in a single row with a button labeled "Yes, was present" below each photograph. Participants clicked on one of these buttons to select a photograph or they clicked on a "No, none of these persons was present" button to the right of the photographs to reject the lineup. Participants also indicated their confidence in their response to make the procedure similar to that of a real police lineup. Participants confirmed their response by clicking on a "Continue" button (all self-paced). After a click on this button, the next lineup was shown unless participants had already seen all four lineups in which case they were debriefed and thanked for participating.

In sequential lineups, the six photographs were presented one at a time. Participants decided for each face whether it belonged to one of the culprits by clicking either on a "Yes, was present" button below the photograph or on a "No, this person was not present" button presented to the right of the photograph. Participants also indicated their confidence in their response to make the procedure similar to that of a real police lineup. Participants confirmed their response by clicking on a "Continue" button (all self-paced). Clicking on this button initiated the display of the next photograph until all six photographs of a lineup had been presented. Then, the next lineup started unless participants had already seen the fourth lineup in which case they were debriefed and thanked for participating. In many experiments with sequential lineups participants know that only their first selection counts and that the lineup is terminated as soon as a face is selected. However, just like Steblay et al. (2011), we are not aware of any jurisdiction that implements sequential lineups in this way (Bill Blackwood Law Enforcement Management Institute of Texas 2022; Home Office 2017; Horry et al. 2012; U.S. Department of Justice: Office of the Deputy Attorney General 2017; Wells et al. 2011; Wells, Steblay, and Dysart 2015). Thus, in the interest of ecological validity (Horry et al. 2021; Winter et al. 2023), the lineups continued even after participants had selected a face. In case of more than one selection in a single lineup, later selections were considered as revisions of earlier selections. If no photograph in the lineup was selected, the lineup was categorized as rejected.

1.1.5 | Open Practices

The data, analysis code, and supplementary analyses of both experiments reported here are available at the Open Science Framework at https://osf.io/mspzk/. The experiments were not pre-registered.

1.2 | Results and Discussion

The main purpose of the current study was to test whether general metamemory beliefs predict culprit-presence detection. To this end, we analyzed participants' response frequencies (see Table 1) in the lineups with the 2-HT eyewitness identification model (Menne et al. 2022; Winter et al. 2022) to obtain parameter estimates for culprit-presence detection. The constant 1 ÷ lineup size was set to 0.16667.

Parameters were estimated within the Bayesian-hierarchical latent trait approach (Klauer 2010) using the *R* package *TreeBUGS* (Heck et al. 2018). To obtain baseline parameter estimates, we calculated one model per group. The latent-trait approach extracts participant-specific parameters from a multivariate normal distribution of probit-transformed parameters. Via Bayes' theorem, the a priori distribution of parameters is updated toward a posterior distribution, given the data. Samples are drawn from the posterior distribution with the Markov chain Monte Carlo algorithm from which a parameter estimate and a corresponding 95% Bayesian credibility interval are calculated. The true parameter value is found within the Bayesian credibility interval with 95% confidence.

For each group, we used 16 chains in which we obtained 1000000 samples with a burn-in period of 500000 samples each. We retained every 100th sample. Parameter convergence is indicated by the Gelman-Rubin statistic (Gelman and Rubin 1992) being $\hat{R} < 1.01$. Parameter convergence was good for all parameters and all parameter correlations. An assessment of model fit is provided by posterior predictive tests (Meng 1994). For Bayesian-hierarchical multinomial processing tree models, Klauer (2010) introduced two model-fit statistics which are based on the posterior samples. Statistic T_1 represents the distance between the observed and predicted mean of the individual frequencies and statistic T_2 represents the summed differences between the observed and predicted covariances of the individual frequencies (Klauer 2010). For both T_1 and T_2, posterior predictive p values are calculated, which indicate good model fit if $p < 0.05$ (cf. Heck et al. 2018). Parameter estimates and the posterior predictive p values as indicators of model fit for each group are presented in Table 2. As can be seen from the table, model fit was good in all groups. Here we were primarily interested in parameter dP (culprit-presence detection).[1] We first tested the effect of lineup type (simultaneous or sequential) on parameter dP by calculating

TABLE 1 | Observed response frequencies for culprit identifications, filler selections and false lineup rejections in culprit-present lineups and innocent-suspect selections, filler selections and correct lineup rejections in culprit-absent lineups as a function of lineup type (simultaneous, sequential) and questionnaire (metamemory, control) in Experiments 1 and 2. Response frequencies are aggregated across lineups and participants.

			Culprit-present lineups			Culprit-absent lineups		
Experiment	Lineup type	Questionnaire group	Culprit identifications	Filler selections	False lineup rejections	Innocent-suspect selections	Filler selections	Correct lineup rejections
1	Simultaneous	Metamemory	336	192	300	103	276	449
		Control	285	243	306	109	293	432
	Sequential	Metamemory	304	323	147	119	353	302
		Control	260	394	164	111	468	239
2	Simultaneous	Metamemory	234	208	290	103	297	332
		Control	238	229	273	89	267	384
	Sequential	Metamemory	218	393	117	115	453	302
		Control	225	376	133	93	422	160

TABLE 2 | Parameter estimates and posterior predictive *p* values as an indicator of model fit as a function of lineup type (simultaneous, sequential) and questionnaire (metamemory, control) for Experiments 1 and 2.

Experiment	Lineup type	Questionnaire	Parameters				Model fit (Posterior predictive, *p*)	
			dP	*dA*	*b*	*g*	T_1	T_2
1	Simultaneous	Metamemory	0.29 (0.23, 0.35)	0.04 (<0.01, 0.11)	0.05 (0.01, 0.09)	0.41 (0.35, 0.48)	0.448	0.524
		Control	0.17 (0.08, 0.25)	0.05 (<0.01, 0.14)	0.06 (0.03, 0.10)	0.48 (0.41, 0.56)	0.492	0.545
	Sequential	Metamemory	0.21 (0.12, 0.28)	0.11 (0.02, 0.19)	0.07 (0.03, 0.11)	0.80 (0.72, 0.87)	0.521	0.575
		Control	0.18 (0.12, 0.23)	0.04 (<0.01, 0.08)	0.02 (<0.01, 0.05)	0.86 (0.80, 0.92)	0.499	0.152
2	Simultaneous	Metamemory	0.17 (0.09, 0.24)	0.02 (<0.01, 0.07)	0.06 (0.03, 0.09)	0.51 (0.45, 0.57)	0.142	0.295
		Control	0.19 (0.11, 0.25)	0.07 (<0.01, 0.17)	0.05 (0.01, 0.09)	0.51 (0.45, 0.58)	0.511	0.326
	Sequential	Metamemory	0.08 (0.02, 0.16)	0.02 (<0.01, 0.05)	0.03 (<0.01, 0.06)	0.87 (0.82, 0.91)	0.501	0.555
		Control	0.15 (0.07, 0.22)	0.06 (0.01, 0.12)	0.02 (<0.01, 0.05)	0.87 (0.81, 0.92)	0.516	0.462

Note: dP: Culprit-presence detection; *dA*: Culprit-absence detection; *b*: Biased suspect selection (conditional on the presence or absence of the culprit not being detected); *g*: Guessing-based selection among the lineup members (conditional on the presence or absence of the culprit not being detected and conditional on the suspect not being selected due to biased suspect selection). 95% Bayesian credibility intervals are shown in parentheses.

the absolute parameter differences between groups. A parameter difference is deemed statistically significant if the corresponding 95% Bayesian credibility interval does not contain zero. Lineup type did not affect parameter *dP*, neither in the control condition, difference = 0.01 [−0.09, 0.12], nor in the metamemory condition, difference = 0.09 [−0.01, 0.19]. For this reason, the following analyses were calculated using the combined data from both lineup types. For the most part, the results were the same when the analyses were conducted for the lineup types separately (results split by lineup type are reported in Tables S2 and S3 at the Open Science Framework). We will note the few instances where this was not the case.

1.2.1 | The Relation of General Metamemory Beliefs and Culprit-Presence Detection

The mean summary scores, their standard deviations, and the correlations among the metamemory-belief scales are presented in Table S4 at the Open Science Framework (see Open Practices section). These data agree well with the corresponding data published by Saraiva et al. (2020, Table 4).

To test whether general metamemory beliefs predict culprit-presence detection, we regressed the participant-specific parameter estimates for parameter *dP* (culprit-presence detection) to the summary scores calculated for each of the subscales of the metamemory questionnaires[2]. For these regression analyses, we calculated Bayes factors indicating how much more likely the observed data are under the null hypothesis than under the alternative hypothesis. Based on the conventions suggested by Raftery (1995), Bayes factors larger than 3, 20, and 150 indicate positive, strong, and very strong evidence, respectively. The regression coefficients (together with their 95% Bayesian credibility intervals) and Bayes factors are presented in Table 3. A positive regression weight indicates that culprit-presence detection is more likely as a predictor increases. Conversely, a negative regression weight indicates that culprit-presence detection is less likely as the predictor increases. All regression weights were close to zero, and all 95% Bayesian credibility intervals included zero. Critically, all Bayes factors indicated positive to strong evidence in favor of the null hypothesis.

Regression analyses use the full information in the data, but in the current case, the participant-specific parameter estimates were based on only four observations, which is somewhat unique in multinomial modeling. In particular, the small number of observations per participant may impact the reliability of the participant-specific parameter estimates, for example due to shrinkage (i.e., the individual parameters being pulled towards the mean parameter estimate due to the hierarchical model structure, e.g., Gelman et al. 2013), which could impact the regression results. We therefore tested whether the regression results could be replicated using group-level parameter estimates, which are based on the response frequencies both at the individual and the group level. To do this, we performed median splits for the summary scores of all scales of the metamemory questionnaires and fitted one model for the lower half and one for the upper half (including the median itself) of each metamemory scale.[3] We compared the group-level estimates of parameter

TABLE 3 | Regression weights and Bayes factors for the null hypothesis for the metamemory subscales predicting culprit-presence detection (*dP*).

Experiment	Metamemory questionnaire	Subscale	Regression weight	Bayes factor
1	Eyewitness metamemory scale	Contentment	< 0.01 (> −0.01, 0.01)	15.85
		Discontentment	> −0.01 (−0.02, < 0.01)	12.89
		Strategy	> −0.01 (−0.01, 0.01)	22.84
	Multifactorial metamemory questionnaire	Contentment	< 0.01 (−0.01, 0.01)	21.21
		Ability	< 0.01 (−0.01, 0.01)	17.18
		Strategy	< 0.01 (−0.01, 0.01)	22.10
	Squire subjective memory questionnaire		< 0.01 (−0.01, 0.01)	24.01
2	Eyewitness metamemory scale	Contentment	0.02 (< 0.01, 0.03)	0.68[a]
		Discontentment	−0.01 (−0.03, < 0.01)	5.96
		Strategy	0.02 (> −0.01, 0.04)	4.97
	Multifactorial metamemory questionnaire	Contentment	0.01 (> −0.01, 0.02)	5.17
		Ability	−0.01 (−0.02, 0.01)	11.95
		Strategy	0.01 (−0.01, 0.02)	10.99
	Squire subjective memory questionnaire		0.01 (> −0.01, 0.01)	5.42

Note: 95% Bayesian credibility intervals are shown in parentheses.
[a]The corresponding Bayes factor for the alternative hypothesis is 1.47, which is considered inconclusive according to the conventions by Raftery (1995).

TABLE 4 | Results of the median-split analyses for culprit-presence detection (*dP*).

			Median split		
Experiment	Metamemory questionnaire	Subscale	Lower half	Upper half	Difference
1	Eyewitness metamemory scale	Contentment	0.25 (0.17, 0.31)	0.25 (0.17, 0.32)	< 0.01 (−0.11, 0.10)
		Discontentment	0.25 (0.16, 0.32)	0.24 (0.17, 0.30)	0.01 (−0.10, 0.11)
		Strategy	0.27 (0.20, 0.33)	0.22 (0.13, 0.30)	0.05 (−0.06, 0.16)
	Multifactorial metamemory questionnaire	Contentment	0.25 (0.17, 0.31)	0.25 (0.16, 0.31)	< 0.01 (−0.10, 0.11)
		Ability	0.24 (0.15, 0.30)	0.26 (0.18, 0.32)	0.02 (−0.08, 0.13)
		Strategy	0.24 (0.14, 0.31)	0.26 (0.19, 0.32)	0.03 (−0.08, 0.14)
	Squire subjective memory questionnaire		0.28 (0.21, 0.34)	0.21 (0.12, 0.28)	0.07 (−0.03, 0.18)
2	Eyewitness metamemory scale	Contentment	0.11 (0.03, 0.18)	0.14 (0.05, 0.22)	0.03 (−0.09, 0.14)
		Discontentment	0.15 (0.06, 0.23)	0.11 (0.02, 0.18)	0.05 (−0.07, 0.17)
		Strategy	0.14 (0.05, 0.20)	0.11 (0.03, 0.20)	0.03 (−0.09, 0.14)
	Multifactorial metamemory questionnaire	Contentment	0.10 (0.02, 0.17)	0.16 (0.07, 0.23)	0.06 (−0.05, 0.17)
		Ability	0.17 (0.08, 0.24)	0.07 (0.01, 0.16)	0.10 (−0.02, 0.20)
		Strategy	0.12 (0.03, 0.19)	0.13 (0.04, 0.21)	0.02 (−0.10, 0.13)
	Squire subjective memory questionnaire		0.14 (0.06, 0.20)	0.10 (0.02, 0.19)	0.04 (−0.08, 0.14)

Note: 95% Bayesian credibility intervals are in parenthesis.

dP (culprit-presence detection) between the upper and lower halves. The results are presented in Table 4. The descriptive differences of the parameter estimates between the upper and lower halves were in the predicted direction in only three out of seven comparisons. In line with the regression analyses presented in the previous paragraph, parameter differences were all non-significant, indicated by all 95% Bayesian credibility intervals of the difference including zero. There is thus converging evidence that general metamemory beliefs do not predict culprit-presence detection.

1.2.2 | Reactive Effects of Metamemory Assessment

Assessing general metamemory beliefs prior to participants' watching the staged-crime video may reactively enhance culprit-presence detection (*dP*) compared to the control group. Descriptively, the estimate of *dP* was higher for the metamemory group compared to the corresponding control group (see Table 2).[4] Averaged across both lineup types, however, this difference narrowly missed statistical significance, difference = 0.07 [> −0.01, 0.15]. When analyzed separately by lineup type, the reactive effect was significant in the simultaneous-lineups condition, difference = 0.12 [0.02, 0.23], but not in the sequential-lineups condition, difference = 0.03 [−0.08, 0.12]. However, the lineup type × questionnaire group interaction was not significant, interaction effect = 0.10 [−0.04, 0.25]. Evidence for reactivity of a priori metamemory-belief assessment was thus limited.

2 | Experiment 2

General metamemory beliefs can be held prior to any experience with a memory task (e.g., Frank and Kuhlmann 2017; Schaper et al. 2019b). Such a priori metamemory beliefs were assessed in Experiment 1 in which participants responded to the metamemory questionnaires before watching the staged-crime video—just like the participants in the study of Saraiva et al. (2020). Thus, when responding to the questionnaires, these participants had little information about the task they were about to complete. They only knew that they would be asked to answer several questions and would then see a short video in which several persons would physically and verbally abuse another person. However, general metamemory beliefs may also change over the course of a task as a function of the information that becomes available during performing the task (e.g., Schaper et al. 2022). Such a situation was implemented in Experiment 2 in which participants first watched the staged-crime video and then responded to the questionnaires. Thus, participants in Experiment 2 knew more about the task at hand (e.g., they knew how long they had seen the culprits' faces) than participants in Experiment 1 when they responded to the metamemory questionnaires. Such knowledge may be integrated with general metamemory beliefs, potentially improving the relationship between participants' general metamemory beliefs and culprit-presence detection.

2.1 | Method

2.1.1 | Participants

Participants were recruited as in Experiment 1. Of the 1801 data sets of the participants who had responded to the sociodemographic questionnaire at the start of the experiment, 279 had to be excluded from the analysis because the participants had not completed the experiment or had withdrawn their consent to the use of their data. Data of 28 participants had to be excluded because they had participated and seen the video with the culprits twice (only the data from the second participation were excluded). Data of 27 participants had to be excluded because they had failed the attention check. The final data set contained data of 1467 participants (625 female, 839 male, 3 non-binary) with a mean age of 48 (SD = 16) years. Participants were randomly assigned to one of the four groups: the simultaneous-lineups metamemory group (366 participants), the simultaneous-lineups control group (370 participants), the sequential-lineups metamemory group (364 participants) and the sequential-lineups control group (367 participants).

2.1.2 | Materials and Procedure

Materials and procedure were identical to those of Experiment 1, except that participants first watched a staged-crime video, then responded to the questionnaires and were subsequently presented with the lineups.

2.2 | Results and Discussion

As in Experiment 1, we analyzed eyewitness memory with the 2-HT eyewitness identification model. All analyses were conducted in the same manner as in Experiment 1. The observed response frequencies are presented in Table 1. Convergence was good for all parameters in all groups as indicated by the Gelman-Rubin statistic (Gelman and Rubin 1992) being $\hat{R} < 1.01$. Parameter estimates and the posterior predictive *p* values as indicators of model fit for each group are presented in Table 2. As can be seen from the table, model fit was good in all groups. As in Experiment 1, lineup type (simultaneous or sequential) did not affect parameter *dP* (culprit-presence detection), neither in the control condition, difference = 0.04 [−0.06, 0.14], nor in the metamemory condition, difference = 0.09 [−0.02, 0.19]. We calculated the following analyses using the combined data from both lineup types.

2.2.1 | The Relation of General Metamemory Beliefs and Culprit-Presence Detection

The mean summary scores, their standard deviations, and the correlations among the metamemory-belief scales are presented in Table S4 at the Open Science Framework (see Open Practices section). These data agree well with the corresponding data published by Saraiva et al. (2020, Table 4) and with the results of Experiment 1.

We again regressed the participant-specific parameter estimates for parameter *dP* (culprit-presence detection) to the summary scores calculated for each of the subscales of the metamemory questionnaires. The regression weights (together with their 95% Bayesian credibility intervals) and Bayes factors are presented in Table 3. All regression weights were very close to zero. With one exception, the 95% Bayesian credibility intervals for the regression weights included zero, and the Bayes factors indicated positive evidence in favor of the null hypothesis. Only the Contentment subscale of the Eyewitness Metamemory Scale showed a slightly positive regression weight, but the corresponding Bayes factor was inconclusive. This effect disappeared when simultaneous and sequential lineups were analyzed separately.

We thus consider this effect to be sporadic and refrain from interpretation.

As in Experiment 1, we also performed median splits for all scales of the metamemory questionnaires (see Table 4) to test whether the results reported for participant-specific parameter estimates can be replicated using group-level parameter estimates. The descriptive differences of the parameter estimates between the upper and lower halves were in the predicted direction in only four out of seven comparisons. In line with the results of the regression analyses, parameter differences between the upper and lower halves were all non-significant.[5] There is thus again converging evidence that general metamemory beliefs do not predict culprit-presence detection.

2.2.2 | Reactive Effects of Metamemory Assessment

Metamemory was assessed after participants had watched the staged-crime video but before lineup presentation. Thus, responding to questions such as "Compared to other people, I more often use a strategy [e.g., focus on specific facial features such as eyes] to remember a person's face", could not have affected encoding, but might have influenced retrieval processes. However, there was no evidence for such a reactive effect on culprit-presence detection, difference = 0.05 [−0.03, 0.13].

3 | General Discussion

General metamemory beliefs could prove highly useful in applied settings if they would reliably predict how well eyewitnesses detect the presence of a culprit in a lineup. For instance, if an eyewitness's general metamemory beliefs clearly indicated that the eyewitness is unlikely to detect the culprit in a lineup, investigators might choose not to conduct the lineup. The present research built upon a previous study on this topic by Saraiva et al. (2020) who presented preliminary findings which they interpreted as suggesting that general metamemory beliefs may be related to response accuracy in lineups, although they acknowledged that more research is necessary in order to arrive at more solid conclusions. Building on this foundation, a process-oriented approach was adopted in the present study to further test whether general metamemory beliefs are meaningfully related to eyewitness memory. Specifically, we tested whether general metamemory beliefs predict culprit-presence detection, a process that is diagnostically relevant in determining a suspect's guilt. In contrast, guessing-based selection, which may also lead to an identification of the culprit, occurs irrespective of the suspect's actual presence in the lineup and therefore lacks diagnostic value. From this perspective, metamemory-belief questionnaires would only be of practical relevance in legal contexts if they specifically predicted culprit-presence detection. To address this question, we conducted two experiments in which participants responded to a series of metamemory questionnaires (or, in the control groups, a personality questionnaire) before watching (Experiment 1) or after having watched (Experiment 2) a staged-crime video. Then, participants were asked to identify culprits from four lineups. In the following, we will discuss the results pertaining to the relationship between general metamemory beliefs and culprit-presence detection, as well as whether responding to the metamemory questionnaires reactively affected culprit-presence detection.

3.1 | General Metamemory Beliefs Do Not Predict Culprit-Presence Detection

Using the same three metamemory questionnaires with seven subscales as in the study by Saraiva et al. (2020), we found no evidence that general metamemory beliefs predict culprit-presence detection. In Experiment 1, participants' general metamemory beliefs were assessed before they watched the staged-crime video. There was no evidence that general metamemory beliefs predict the detection of the presence of the culprit. In Experiment 2, participants witnessed the crime first and had more information than in Experiment 1 about the culprits they would have to identify from lineups before they were asked to respond to the metamemory questionnaires. Nevertheless, we again found no evidence indicating that general metamemory beliefs predict the detection of the presence of the culprit. All analyses but one yielded positive or even strong evidence in favor of the null hypothesis. Consistent with these results, median-split analyses of all seven metamemory scales revealed no differences in culprit-presence detection between participants with higher and lower metamemory scores. Specifically, the descriptive differences of the estimates of parameter dP between the upper and lower halves were in the predicted direction in only half of the comparisons (three out of seven in Experiment 1, four out of seven in Experiment 2). Thus, both methods of data analysis and both experiments provided converging evidence against a predictive relationship between general metamemory belief and culprit-presence detection (see Tables 3 and 4).

From a practical point of view, the current results call into question the usefulness of general metamemory beliefs for predicting eyewitness memory. Accordingly, we caution against using eyewitnesses' general metamemory beliefs as an indicator of the detection of the culprit in a lineup. This conclusion is consistent with prior research showing that general metamemory beliefs are often inaccurate (Bindemann et al. 2014; Frank and Kuhlmann 2017; Luna and Albuquerque 2022; Mieth et al. 2021; Schaper et al. 2019b; Undorf and Zimdahl 2019). On an item-by-item basis, people are somewhat more accurate in distinguishing between information that they will remember well and information they will remember less well (Rhodes 2016), because the standard of comparison is directly available to the person making the predictions. However, judging one's overall memory ability in comparison to that of others may be particularly challenging (Carroll and Nelson 1993). A reason for this may be that the appropriate standard of comparison is not directly available but must be inferred indirectly from one's own memory of other people's behavior in tasks involving memory, which may or may not be related to the task faced when an eyewitness is asked to identify a culprit in a lineup.

3.2 | Reactive Effects of Metamemory Assessment

Another aim of the current study was to test whether responding to metamemory-belief questionnaires may reactively enhance culprit-presence detection. Whereas self-assessments of metamemory have sometimes been shown to guide attention to specific aspects of tasks, leading to improved memory for those aspects (e.g., Besken 2016; Besken and Mulligan 2013, 2014; Mieth et al. 2021; Mitchum et al. 2016; Myers et al. 2020; Schaper et al. 2019a; Senkova and Otani 2021; Soderstrom et al. 2015), the current data showed only limited support for the assumption that measuring general metamemory beliefs reactively affects culprit-presence detection. Compared to the control condition, metamemory-belief assessment enhanced culprit-presence detection when metamemory was assessed before witnessing the crime in Experiment 1, but the effect was statistically significant only for simultaneous lineups and not for sequential lineups or when lineup types were combined. We thus conclude that reactive effects of metamemory-belief assessment on culprit-presence detection appear to be negligible. This is in line with the increasing awareness that metamemory reactivity may be particularly strong for relatively simple materials and may not hold the same relevance in applied settings (Schäfer and Undorf 2024).

3.3 | Limitations and Directions for Future Research

Some limitations of the current research should be noted. First, our results and interpretations rest on the use of the 2-HT eyewitness identification model (Menne et al. 2022; Winter et al. 2022). As any measurement model of eyewitness memory, this model makes critical assumptions about the cognitive processes that underlie eyewitness identifications. For example, it is assumed that different cognitive processes (e.g., memory and guessing) can be separately measured and that, following guessing-based selections, there is no systematic preference for selecting the suspect over the fillers. Such assumptions may be challenged. Winter et al. (2022, p.14) have argued in this regard that “all models are necessarily simplifications of the complexities of reality. [...] The question is not which set of simplifications is correct because all simplifications of the complexities of reality must necessarily be false. Instead, the question is which set of simplifications is sufficiently useful”. A model may in this sense be considered useful if its parameters can be shown to measure what they were intended to measure. For the parameters of the 2-HT eyewitness identification model, this has been shown extensively in a series of validation experiments (Winter et al. 2022) and in a series of eight reanalyses (Menne et al. 2022) of published data (Colloff et al. 2016; Karageorge and Zajac 2011; Lampinen et al. 2020; Malpass and Devine 1981; Memon et al. 2003; Smith 2020; Wetmore et al. 2015; Wilcock and Bull 2010). Combined with the fact that, in the current study, the 2-HT eyewitness identification model produced a good model fit in all groups of both experiments, the available data suggest that the 2-HT eyewitness identification model provides a useful tool for separating the detection-based and guessing-based processes underlying eyewitness responses.

Second, each experiment included only one attention check, raising the possibility that some participants may have responded carelessly to the metamemory questionnaires. However, the fact that the pattern of correlations among the seven subscales of the three metamemory questionnaires was highly consistent both across Experiments 1 and 2 and with the corresponding pattern of correlations reported by Saraiva et al. (2020) suggests that participants engaged with the task in a meaningful way both here and in the study of Saraiva et al. (2020).

Third, we intentionally used the same materials and procedure in both experiments to systematically examine potential reactivity effects of metamemory assessments. In this setting, the hypothesis that general metamemory beliefs predict culprit-presence detection was clearly disconfirmed in two experiments. In future studies aimed at this hypothesis, other general metamemory questionnaires, mock-crime scenarios, and lineup procedures could be used to test whether these measures of general metamemory beliefs might predict culprit-presence detection under different circumstances.

Fourth, our focus was limited to whether general metamemory beliefs predict culprit-presence detection. As such, the present results leave open the possibility that other types of metamemory measures—such as item-level predictions or confidence ratings at test—may provide better indicators of culprit-presence detection. Exploring how different metamemory measures relate to culprit-presence detection would be a valuable direction for future research.

4 | Conclusion

Taken together, the current results suggest that general metamemory beliefs do not predict the detection of the presence of the culprit in a lineup. Accordingly, at this stage we caution against using eyewitnesses' general metamemory beliefs as indicators of the detection of the culprit in a lineup.

Author Contributions

Marie Luisa Schaper: data curation, methodology, project administration, resources, validation, visualization, writing – original draft, formal analysis. **Nicola Marie Menne:** conceptualization, methodology, software, data curation, investigation, formal analysis, writing – review and editing, resources. **Raoul Bell:** conceptualization, methodology, resources, writing – review and editing. **Carolin Mayer:** conceptualization, writing – review and editing. **Axel Buchner:** conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, supervision, validation, visualization, writing – original draft.

Acknowledgements

The research reported in this article was supported by the German Research Foundation: BU 945/10-2, project number 456214986. This research was presented at the 53rd DGPs (German Psychological Society) congress and the 65th annual meeting of the Psychonomic Society. Open Access funding enabled and organized by Projekt DEAL.

Ethics Statement

Approval was received from the ethics committee of the Faculty of Mathematical and Natural Sciences of the Heinrich-Heine-Universität Düsseldorf for a series of experiments to which the present experiments belong.

Consent

Informed consent was obtained from all participants prior to participation.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data and analysis code of the experiments and supplementary analyses are available at the project page of the Open Science Framework at https://osf.io/mspzk/.

Endnotes

[1] Supplementary analyses pertaining to parameter *dA* for both experiments can be found in Table S1 at the Open Science Framework (see Open Practices section).

[2] In the regression analyses, we calculated one model for each metamemory scale and included only that scale as a predictor of *dP*. This was necessary in order to be able to calculate the Bayes factors without making further assumptions about the regression weights (Heck 2019). Further, for these regression analyses, we retained every 500th sample to keep the output files at a manageable size.

[3] For the median-split analyses, we implemented lineup type as predictor of all four parameters in the model (lineup type did not predict parameter *dP*).

[4] For the reactivity analyses, we implemented lineup type as predictor of all four parameters in the model (lineup type did not predict parameter *dP*). Further, we retained every 500th sample to keep the output files at a manageable size.

[5] When the analyses were carried out separately for simultaneous and sequential lineups, there was one significant effect for simultaneous lineups and the Contentment subscale of the Multifactorial Metamemory Questionnaire only. However, this effect was not replicated for sequential lineups, for the two lineup types combined, in the regression analyses or in Experiment 1. We consider this effect to be sporadic and refrain from interpretation.

References

Ashton, M. C., and K. Lee. 2007. "Empirical, Theoretical, and Practical Advantages of the HEXACO Model of Personality Structure." *Personality and Social Psychology Review* 11, no. 2: 150–166. https://doi.org/10.1177/1088868306294907.

Ashton, M. C., K. Lee, B. Marcus, and R. E. De Vries. 2007. "German Lexical Personality Factors: Relations With the HEXACO Model." *European Journal of Personality* 21, no. 1: 23–43. https://doi.org/10.1002/per.597.

Bell, R., N. M. Menne, C. Mayer, and A. Buchner. 2024. "On the Advantages of Using AI-Generated Images of Filler Faces for Creating Fair Lineups." *Scientific Reports* 14: 12304. https://doi.org/10.1038/s41598-024-63004-z.

Berkowitz, S. R., B. L. Garrett, K. M. Fenn, and E. F. Loftus. 2022. "Convicting With Confidence? Why We Should Not Over-Rely on Eyewitness Confidence." *Memory* 30, no. 1: 10–15. https://doi.org/10.1080/09658211.2020.1849308.

Besken, M. 2016. "Picture-Perfect Is Not Perfect for Metamemory: Testing the Perceptual Fluency Hypothesis With Degraded Images." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 42, no. 9: 1417–1433. https://doi.org/10.1037/xlm0000246.

Besken, M., and N. W. Mulligan. 2013. "Easily Perceived, Easily Remembered? Perceptual Interference Produces a Double Dissociation Between Metamemory and Memory Performance." *Memory & Cognition* 41, no. 6: 897–903. https://doi.org/10.3758/s13421-013-0307-8.

Besken, M., and N. W. Mulligan. 2014. "Perceptual Fluency, Auditory Generation, and Metamemory: Analyzing the Perceptual Fluency Hypothesis in the Auditory Modality." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 40, no. 2: 429–440. https://doi.org/10.1037/a0034407.

Bill Blackwood Law Enforcement Management Institute of Texas. 2022. "Model Policy on Eyewitness Identification." http://www.lemitonline.org/resources/documents/ewid_final.pdf.

Bindemann, M., J. Attard, and R. A. Johnston. 2014. "Perceived Ability and Actual Recognition Accuracy for Unfamiliar and Famous Faces." *Cogent Psychology* 1, no. 1: 986903. https://doi.org/10.1080/23311908.2014.986903.

Carroll, M., and T. O. Nelson. 1993. "Effect of Overlearning on the Feeling of Knowing Is More Detectable in Within-Subject Than in Between-Subject Designs." *American Journal of Psychology* 106, no. 2: 227–235. https://doi.org/10.2307/1423169.

Colloff, M. F., K. A. Wade, and D. Strange. 2016. "Unfair Lineups Make Witnesses More Likely to Confuse Innocent and Guilty Suspects." *Psychological Science* 27, no. 9: 1227–1239. https://doi.org/10.1177/0956797616655789.

Erdfelder, E., T.-S. Auer, B. E. Hilbig, A. Assfalg, M. Moshagen, and L. Nadarevic. 2009. "Multinomial Processing Tree Models: A Review of the Literature." *Zeitschrift Für Psychologie/Journal of Psychology* 217, no. 3: 108–124. https://doi.org/10.1027/0044-3409.217.3.108.

Fitzgerald, R. J., E. Rubínová, and S. Juncu. 2021. "Eyewitness Identification Around the World." In *Methods, Measures, and Theories in Eyewitness Identification Tasks*, edited by A. M. Smith, M. P. Toglia, and J. M. Lampinen, 294–322. Routledge. https://doi.org/10.4324/9781003138105.

Frank, D. J., and B. G. Kuhlmann. 2017. "More Than Just Beliefs: Experience and Beliefs Jointly Contribute to Volume Effects on Metacognitive Judgments." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 43, no. 5: 680–693. https://doi.org/10.1037/xlm0000332.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2013. *Bayesian Data Analysis*. 3rd ed. Chapman and Hall/CRC. https://doi.org/10.1201/b16018.

Gelman, A., and D. B. Rubin. 1992. "Inference From Iterative Simulation Using Multiple Sequences." *Statistical Science* 74, no. 4: 457–472. https://doi.org/10.1214/ss/1177011136.

Gopi, Y., and C. R. Madan. 2024. "Subjective Memory Measures: Metamemory Questionnaires Currently in Use." *Quarterly Journal of Experimental Psychology* 77, no. 5: 924–942. https://doi.org/10.1177/17470218231183855.

Heck, D. W. 2019. "A Caveat on the Savage-Dickey Density Ratio: The Case of Computing Bayes Factors for Regression Parameters." *British Journal of Mathematical and Statistical Psychology* 72, no. 2: 316–333. https://doi.org/10.1111/bmsp.12150.

Heck, D. W., N. R. Arnold, and D. Arnold. 2018. "TreeBUGS: An R Package for Hierarchical Multinomial-Processing-Tree Modeling." *Behavior Research Methods* 50, no. 1: 264–284. https://doi.org/10.3758/s13428-017-0869-7.

Hobson, Z. J., and R. Wilcock. 2011. "Eyewitness Identification of Multiple Perpetrators." *International Journal of Police Science and Management* 13, no. 4: 286–296. https://doi.org/10.1350/ijps.2011.13.4.253.

Hobson, Z. J., R. Wilcock, and T. Valentine. 2012. "Multiple Suspect Showing: A Survey of Police Identification Officers." *Policing* 7, no. 1: 79–87. https://doi.org/10.1093/police/pas021.

Home Office. 2017. *Police and Criminal Evidence Act 1984 (PACE) Code D Revised: Code of Practice for the Identification of Persons by Police Officers*. TSO. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/903812/pace-code-d-2017.pdf.

Horry, R., R. J. Fitzgerald, and J. K. Mansour. 2021. "“Only Your First Yes Will Count”: The Impact of Prelineup Instructions on Sequential Lineup Decisions." *Journal of Experimental Psychology: Applied* 27, no. 1: 170–186. https://doi.org/10.1037/xap0000337.

Horry, R., A. Memon, D. B. Wright, and R. Milne. 2012. "Predictors of Eyewitness Identification Decisions From Video Lineups in England: A Field Study." *Law and Human Behavior* 36, no. 4: 257–265. https://doi.org/10.1037/h0093959.

Karageorge, A., and R. Zajac. 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, no. 2: 161–163. https://doi.org/10.1348/000712610X507902.

Klauer, K. C. 2010. "Hierarchical Multinomial Processing Tree Models: A Latent-Trait Approach." *Psychometrika* 75, no. 1: 70–98. https://doi.org/10.1007/s11336-009-9141-0.

Koriat, A. 1997. "Monitoring One's Own Knowledge During Study: A Cue-Utilization Approach to Judgments of Learning." *Journal of Experimental Psychology: General* 126, no. 4: 349–370. https://doi.org/10.1037/0096-3445.126.4.349.

Lampinen, J. M., B. Race, A. P. Wolf, P. Phillips, N. Moriarty, and A. M. Smith. 2020. "Comparing Detailed and Less Detailed Pre-Lineup Instructions." *Applied Cognitive Psychology* 34, no. 2: 409–424. https://doi.org/10.1002/acp.3627.

Lee, K., and M. C. Ashton. 2018. "Psychometric Properties of the HEXACO-100." *Assessment* 25, no. 5: 543–556. https://doi.org/10.1177/1073191116659134.

Lockamyeir, R. F., C. A. Carlson, A. R. Jones, A. R. Wooten, M. A. Carlson, and J. A. Hemby. 2021. "One Perpetrator, Two Perpetrators: The Effect of Multiple Perpetrators on Eyewitness Identification." *Applied Cognitive Psychology* 35, no. 5: 1206–1223. https://doi.org/10.1002/acp.3853.

Luna, K., and P. B. Albuquerque. 2022. "Do Beliefs About Font Size Affect Retrospective Metamemory Judgments in Addition to Prospective Judgments? A Registered Report." *Experimental Psychology* 69, no. 3: 172–184. https://doi.org/10.1027/1618-3169/a000549.

Malpass, R. S., and P. G. Devine. 1981. "Eyewitness Identification: Lineup Instructions and the Absence of the Offender." *Journal of Applied Psychology* 66, no. 4: 482–489. https://doi.org/10.1037/0021-9010.66.4.482.

Mayer, C., R. Bell, N. M. Menne, and A. Buchner. 2024. "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.

Memon, A., L. Hope, and R. Bull. 2003. "Exposure Duration: Effects on Eyewitness Accuracy and Confidence." *British Journal of Psychology* 94, no. 3: 339–354. https://doi.org/10.1348/000712603767876262.

Meng, X.-L. 1994. "Posterior Predictive *p*-Values." *Annals of Statistics* 22, no. 3: 1142–1160. http://www.jstor.org/stable/2242219.

Menne, N. M., R. Bell, K. Winter, and A. Buchner. 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., K. Winter, R. Bell, and A. Buchner. 2022. "A Validation of the Two-High Threshold Eyewitness Identification Model by Reanalyzing Published Data." *Scientific Reports* 12: 13379. https://doi.org/10.1038/s41598-022-17400-y.

Menne, N. M., K. Winter, R. Bell, and A. Buchner. 2023a. "The Effects of Lineup Size on the Processes Underlying Eyewitness Decisions." *Scientific Reports* 13: 17190. https://doi.org/10.1038/s41598-023-44003-y.

Menne, N. M., K. Winter, R. Bell, and A. Buchner. 2023b. "Measuring Lineup Fairness From Eyewitness Identification Data Using a Multinomial Processing Tree Model." *Scientific Reports* 13: 6290. https://doi.org/10.1038/s41598-023-33101-6.

Mieth, L., M. L. Schaper, B. G. Kuhlmann, and R. Bell. 2021. "Memory and Metamemory for Social Interactions: Evidence for a Metamemory Expectancy Illusion." *Memory & Cognition* 49, no. 1: 14–31. https://doi.org/10.3758/s13421-020-01071-z.

Minear, M., and D. C. Park. 2004. "A Lifespan Database of Adult Facial Stimuli." *Behavior Research Methods, Instruments, & Computers* 36, no. 4: 630–633. https://doi.org/10.3758/BF03206543.

Mitchum, A. L., C. M. Kelley, and M. C. Fox. 2016. "When Asking the Question Changes the Ultimate Answer: Metamemory Judgments Change Memory." *Journal of Experimental Psychology: General* 145, no. 2: 200–219. https://doi.org/10.1037/a0039923.

Myers, S. J., M. G. Rhodes, and H. E. Hausman. 2020. "Judgments of Learning (JOLs) Selectively Improve Memory Depending on the Type of Test." *Memory & Cognition* 48, no. 5: 745–758. https://doi.org/10.3758/s13421-020-01025-5.

Raftery, A. E. 1995. "Model Selection in Social Research." *Sociological Methodology* 25: 111–163. https://doi.org/10.2307/271063.

Rhodes, M. G. 2016. "Judgments of Learning: Methods, Data, and Theory." In *The Oxford Handbook of Metamemory*, edited by J. Dunlosky and S. K. Tauber, 65–80. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199336746.013.4.

Saraiva, R. B., I. van Boeijen, L. Hope, R. Horselenberg, M. Sauerland, and P. J. Koppen. 2020. "Eyewitness Metamemory Predicts Identification Performance in Biased and Unbiased Line-Ups." *Legal and Criminological Psychology* 25, no. 2: 111–132. https://doi.org/10.1111/lcrp.12166.

Saraiva, R. B., I. M. van Boeijen, L. Hope, R. Horselenberg, M. Sauerland, and P. J. Koppen. 2019. "Development and Validation of the Eyewitness Metamemory Scale." *Applied Cognitive Psychology* 33, no. 5: 964–973. https://doi.org/10.1002/acp.3588.

Sauer, J. D., M. A. Palmer, and N. Brewer. 2019. "Pitfalls in Using Eyewitness Confidence to Diagnose the Accuracy of an Individual Identification Decision." *Psychology, Public Policy, and Law* 25, no. 3: 147–165. https://doi.org/10.1037/law0000203.

Schäfer, F., and M. Undorf. 2024. "On the Educational Relevance of Immediate Judgment of Learning Reactivity: No Effects of Predicting One's Memory for General Knowledge Facts." *Journal of Applied Research in Memory and Cognition* 13, no. 1: 113–123. https://doi.org/10.1037/mac0000113.

Schaper, M. L., U. J. Bayen, and C. V. Hey. 2022. "Delaying Metamemory Judgments Corrects the Expectancy Illusion in Source Monitoring: The Role of Fluency and Belief." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 48, no. 7: 975–1000. https://doi.org/10.1037/xlm0001088.

Schaper, M. L., B. G. Kuhlmann, and U. J. Bayen. 2019a. "Metacognitive Expectancy Effects in Source Monitoring: Beliefs, In-The-Moment Experiences, or Both?" *Journal of Memory and Language* 107: 95–110. https://doi.org/10.1016/j.jml.2019.03.009.

Schaper, M. L., B. G. Kuhlmann, and U. J. Bayen. 2019b. "Metamemory Expectancy Illusion and Schema-Consistent Guessing in Source Monitoring." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 45, no. 3: 470–496. https://doi.org/10.1037/xlm0000602.

Schmidt, O., E. Erdfelder, and D. W. Heck. 2023. "How to Develop, Test, and Extend Multinomial Processing Tree Models: A Tutorial." *Psychological Methods*. https://doi.org/10.1037/met0000561.

Senkova, O., and H. Otani. 2021. "Making Judgments of Learning Enhances Memory by Inducing Item-Specific Processing." *Memory & Cognition* 49, no. 5: 955–967. https://doi.org/10.3758/s13421-020-01133-2.

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, no. 4: 495–507. https://doi.org/10.1016/j.jarmac.2020.08.002.

Smith, A. M., L. Smalarz, R. Ditchfield, and N. T. Ayala. 2021. "Evaluating the Claim That High Confidence Implies High Accuracy in Eyewitness Identification." *Psychology, Public Policy, and Law* 27, no. 4: 479–491. https://doi.org/10.1037/law0000324.

Smith, A. M., G. L. Wells, R. C. L. Lindsay, and S. D. Penrod. 2017. "Fair Lineups Are Better Than Biased Lineups and Showups, but Not Because They Increase Underlying Discriminability." *Law and Human Behavior* 41, no. 2: 127–145. https://doi.org/10.1037/lhb0000219.

Soderstrom, N. C., C. T. Clark, V. Halamish, and E. L. Bjork. 2015. "Judgments of Learning as Memory Modifiers." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 41, no. 2: 553–558. https://doi.org/10.1037/a0038388.

Steblay, N. K., J. E. Dysart, and G. L. Wells. 2011. "Seventy-Two Tests of the Sequential Lineup Superiority Effect: A Meta-Analysis and Policy Discussion." *Psychology, Public Policy, and Law* 17, no. 1: 99–139. https://doi.org/10.1037/a0021650.

Tauber, S. K., and J. Dunlosky. 2016. "A Brief History of Metamemory Research and Handbook Overview." In *The Oxford Handbook of Metamemory*, edited by J. Dunlosky and S. K. Tauber, 7–21. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199336746.013.13.

Therre, A., R. Bell, N. M. Menne, C. Mayer, U. Lichtenhagen, and A. Buchner. 2024. "On the Possible Advantages of Combining Small Lineups With Instructions That Discourage Guessing-Based Selection." *Scientific Reports* 14: 14126. https://doi.org/10.1038/s41598-024-64768-0.

Troyer, A. K., and J. B. Rich. 2002. "Psychometric Properties of a New Metamemory Questionnaire for Older Adults." *Journals of Gerontology: Series B: Psychological Sciences and Social Sciences* 57, no. 1: P19–P27. https://doi.org/10.1093/geronb/57.1.P19.

Tupper, N., M. Sauerland, J. D. Sauer, and L. Hope. 2019. "Eyewitness Identification Procedures for Multiple Perpetrator Crimes: A Survey of Police in Sweden, Belgium, and The Netherlands." *Psychology, Crime & Law* 25, no. 10: 992–1007. https://doi.org/10.1080/1068316X.2019.1611828.

Undorf, M., and M. F. Zimdahl. 2019. "Metamemory and Memory for a Wide Range of Font Sizes: What Is the Contribution of Perceptual Fluency?" *Journal of Experimental Psychology: Learning, Memory, and Cognition* 45, no. 1: 97–109. https://doi.org/10.1037/xlm0000571.

U.S. Department of Justice: Office of the Deputy Attorney General. 2017. "Eyewitness Identification: Procedures for Conducting Photo Arrays." https://www.justice.gov/archives/opa/press-release/file/923201/dl.

van Bergen, S., I. Brands, M. Jelicic, and H. Merckelbach. 2010. "Assessing Trait Memory Distrust: Psychometric Properties of the Squire Subjective Memory Questionnaire." *Legal and Criminological Psychology* 15, no. 2: 373–384. https://doi.org/10.1348/135532509X471960.

Wells, G. L., L. Smalarz, and A. M. Smith. 2015. "ROC Analysis of Lineups Does Not Measure Underlying Discriminability and has Limited Value." *Journal of Applied Research in Memory and Cognition* 4, no. 4: 313–317. https://doi.org/10.1016/j.jarmac.2015.08.008.

Wells, G. L., A. M. Smith, and L. Smalarz. 2015. "ROC Analysis of Lineups Obscures Information That Is Critical for Both Theoretical Understanding and Applied Purposes." *Journal of Applied Research in Memory and Cognition* 4, no. 4: 324–328. https://doi.org/10.1016/j.jarmac.2015.08.010.

Wells, G. L., N. Steblay, and J. E. Dysart. 2011. "A Test of the Simultaneous vs. Sequential Lineup Methods." *American Judicature Society.*

Wells, G. L., N. K. Steblay, and J. E. Dysart. 2015. "Double-Blind Photo Lineups Using Actual Eyewitnesses: An Experimental Test of a Sequential Versus Simultaneous Lineup Procedure." *Law and Human Behavior* 39, no. 1: 1–14. https://doi.org/10.1037/lhb0000096.

Wetmore, S. A., J. S. Neuschatz, S. D. Gronlund, A. Wooten, C. A. Goodsell, and C. A. Carlson. 2015. "Effect of Retention Interval on Showup and Lineup Performance." *Journal of Applied Research in Memory and Cognition* 4, no. 1: 8–14. https://doi.org/10.1016/j.jarmac.2014.07.003.

Wilcock, R., and R. Bull. 2010. "Novel Lineup Methods for Improving the Performance of Older Eyewitnesses." *Applied Cognitive Psychology* 24, no. 5: 718–736. https://doi.org/10.1002/acp.1582.

Winter, K., N. M. Menne, R. Bell, and A. Buchner. 2022. "Experimental Validation of a Multinomial Processing Tree Model for Analyzing Eyewitness Identification Decisions." *Scientific Reports* 12: 15571. https://doi.org/10.1038/s41598-022-19513-w.

Winter, K., N. M. Menne, R. Bell, and A. Buchner. 2023. "Evaluating the Impact of First-Yes-Counts Instructions on Eyewitness Performance Using the Two-High Threshold Eyewitness Identification Model." *Scientific Reports* 13: 6572. https://doi.org/10.1038/s41598-023-33424-4.

Wixted, J. T., and L. Mickes. 2022. "Eyewitness Memory Is Reliable, but the Criminal Justice System Is Not." *Memory* 30, no. 1: 67–72. https://doi.org/10.1080/09658211.2021.1974485.

Wixted, J. T., and G. L. Wells. 2017. "The Relationship Between Eyewitness Confidence and Identification Accuracy: A New Synthesis." *Psychological Science in the Public Interest* 18, no. 1: 10–65. https://doi.org/10.1177/1529100616686966.