 doi:10.20944/preprints201812.0012.v1

Metamemory and lineup instructions:

Asking eyewitnesses to recollect reduces false identification

Nicola Guerin[a1], Nathan Weber[a], and Ruth Horry[b]

[a]Flinders University, School of Psychology,

GPO Box 2100, Adelaide, South Australia, 5001

[b]Swansea University, Department of Psychology,

College of Human and Health Sciences, Vivian Tower, Swansea, SA2 8PP, UK

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

Ethics Approval: The four experiments were conducted in accordance with APA, Australia's National Health and Medical Research Council guidelines, and The Code of Ethics of the World Medical Association (Declaration of Helsinki) for ethical research.

Correspondence concerning this article should be addressed to Nicola Guerin, Cancer Council Victoria,615 St Kilda Road, Melbourne, Victoria, 3004, Australia. Email: nicola.guerin@cancervic.org.au

[1] Present address: Cancer Council Victoria, 615 St Kilda Road, Melbourne, Victoria, 3004, Australia

 doi:10.20944/preprints201812.0012.v1

Metamemory and lineup instructions: Asking eyewitnesses to recollect reduces false identification

Abstract

Little theoretically-informed research investigates how non-standard eyewitness identification tasks or metacognitive instructions might improve identification accuracy. We used a continuous dual-process model of recognition to explain familiarity-based identification errors and design modified lineup tasks and metacognitive instructions that increased eyewitness recollection and discriminability. In four studies we examined identification performance across lineups (standard simultaneous, elimination, delayed-choice) and instructions (task-related, phenomenological, standard). Participants viewed photos of targets and made identification decisions about a lineup for each target. Instructions about memory phenomenology improved discriminability in delayed-choice lineups, while task-related instructions were ineffective. Metacognitive instructions about how to better evaluate memory quality in modified lineup tasks could improve recollection for greater identification accuracy even when memory is poor. While immediate post-decision confidence is a good predictor of identification accuracy, lineup modifications that improve eyewitness memory use would provide better evidence of suspect guilt or innocence. We discuss implications for lineup theory and design.

Keywords: eyewitness identification; feedback; recollection; metacognition; metamemory; recognition

 doi:10.20944/preprints201812.0012.v1

Eyewitness identifications play a key role in many criminal cases. In this role, lineup tasks are tests to diagnose the guilt or innocence of a suspect (Wells & Luus, 1990). Missed identifications are missed opportunities to bring an offender to justice, but potentially the worst diagnostic errors are false positives, carrying serious consequences for innocent suspects. Ultimately, the value of identification procedures rests on their diagnostic strength. However, witnesses, not independent experts, make the diagnosis. Eyewitness beliefs and expectations about lineups are likely to inform how they approach the recognition task and how effectively they use memory and metamemorial information. Ultimately, witness metacognitions have a substantial impact on the procedure's diagnostic strength. The reliability of results from even the most tightly controlled procedures can be decreased by eyewitness preconceptions about memory retrieval and evaluation (e.g., Brewer & Palmer, 2010).

If law enforcement follow recommendations to consider identification confidence expressed at the time of the test as part of diagnosing suspect guilt, many low-confidence errors might have no serious consequences (Wixted, Mickes, & Fisher, 2018). Even so, eyewitness memory could still be improved and identification procedures could take better advantage of witnesses' cognitive and metacognitive capacity. Improved procedures that control metacognitive effects and stimulate better memory use would increase the evidence-value of test results. But to build these procedures, we need to start from theoretical models that explain the underlying processes. We extended a continuous dual-process model of recognition (Wixted & Mickes, 2010) to develop novel lineup procedures and instructions. In four experiments we tested these modifications designed to get eyewitnesses to make better use of memory and metacognitive resources to give more informative lineup responses. Below, we review the literature on eyewitness metacognition and outline the rationale for our modified identification procedures.

Identification bias

A misconception shared by many eyewitnesses is that lineups are held to 'get the guy' (Malpass & Devine, 1981; Wells, 1984). Identification evidence is highly valued and this often biases eyewitnesses toward making an identification and against deciding the offender is not present. Biased eyewitnesses neglect critical qualities of retrieved memories that indicate an identification is unwarranted (Meissner, Tredoux, Parker, & MacLin, 2005, Palmer, Brewer, McKinnon & Weber, 2010). In this way, bias raises the risk of false identification for innocent suspects resembling the offender (Brewer & Palmer, 2010).

A key feature of simultaneous lineups that may increase the effect of identification bias is also a feature of associative recognition memory lab tasks: stimuli selected for high similarity. In both tasks, a previously-seen target must be discriminated from highly-similar unstudied stimuli. Unstudied stimuli seem familiar due to their similarity to studied targets and are often mistakenly recognised (Cook, Marsh, & Hicks, 2005; Malmberg, 2008). Continuous dual-process theories of recognition propose that both general familiarity and more detailed recollection can inform recognition in associative tasks (Wixted & Mickes, 2010). However, if participants focus on evaluating positive familiarity-based evidence of match, they underestimate the importance of recollections that might rule out a match (i.e., recall-to-reject). For example, an unstudied face that closely resembles a studied face is likely to be falsely recognised if the participant is not aware that test stimuli are highly similar and that specific target details must be recollected to distinguish them.

In simultaneous lineups, the best match for the target will be the lineup member most similar to the memorial representation of the offender and must be found among multiple similar faces (Wells & Lindsay, 1980). Associative memory research suggests that participants focusing on positive match to find the best-match fail to attend to mismatching details unless told recollection is critical, or recollection is prompted by the task (e.g.,

 doi:10.20944/preprints201812.0012.v1

enforced response delay, Malmberg & Xu, 2007). However, when task structure or instructions cue use of recollection, false recognition can be reduced, and the familiarity of studied targets supplemented with recollective detail (Ingram, Mickes, & Wixted, 2012). In simultaneous lineup procedures, however, there is no explicit structural cue for recollection (cf Wixted & Mickes, 2014). Further, Malmberg and Xu (2007) found that familiarity-based false recognition increased when test-lists included both high and low similarity unstudied stimuli. The presence of less similar unstudied test stimuli 'tricks' participants to judge that effortful recollection is not needed to differentiate foils from targets, and highly-similar stimuli are then misrecognised. As lineups commonly include some less similar members, eyewitnesses are likely to neglect recollection, raising false identification risk for innocent suspects.

Two design modifications would target these error-prone processes: structural cues for recollection, and metacognitive instructions about using recollection. In four experiments, we investigated the effects of 1) elimination and delayed-choice lineup procedures structured to prompt recollection; and, 2) task-related and phenomenological instructions for memory use.

Lineup Structure

Several theories of simultaneous lineups propose identification decisions are made in two steps (Clark, 2003; Pozzulo & Lindsay, 1999; Wixted et al., 2018). First, while memorial evidence accumulates in response to lineup members, a best-match is found (*best-match selection*). Second, the witness decides whether the best-match is the offender (*target-presence judgment*). To find the best-match, eyewitnesses can rely on the strength of familiarity without considering recollected details. These details can be more difficult to recollect than the more automatic familiarity associated with a face, particularly for weak memories (Malmberg, 2008). Once established, familiarity-based recognition processes are then likely to be used for the second target-presence judgment, unless the decision strategy is re-set (Alban & Kelley, 2012). If not re-set, the witness might not effortfully recollect critical details that would count against misidentification or increased evidence for accurate identification. Therefore, witnesses should be encouraged to avoid familiarity-based processes, and to retrieve and use recollection.

To discourage similarity-based choosing, Wells (1984) influentially proposed sequential lineup presentation. However, sequential presentation discourages liberal choosing without improving discriminability (Clark, 2012; Wixted, Mickes, Dunn, Clark, & Wells, 2016). Simultaneous presentation offers potential benefits for memory retrieval, such as comparisons between lineup members that could highlight useful features for diagnosing guilt (Goodsell, Gronlund, & Carlson, 2010; Wixted & Mickes, 2014; Wixted, Vul, Mickes, & Wilson, 2018). Therefore, we retained simultaneous presentation in two novel lineup procedures.

We tested two-step simultaneous lineup procedures requiring witnesses to give separate responses for the best-match and offender-presence judgments. In the *elimination lineup* (Pozzulo & Lindsay, 1999), witnesses indicate the best-match (even if not thought to be the offender), and then indicate whether the best-match is the offender. In the *delayed-choice lineup*, eyewitnesses first indicate whether or not the offender is in the lineup (without selecting a lineup member), and later indicate which lineup member is the offender (if thought present). Evidence from lab tasks (Alban & Kelley, 2012) suggests that by separating these judgments, eyewitnesses will be cued to re-set their recognition strategy to that perceived to be most appropriate for each judgment. Thus, eyewitnesses would be prompted to retrieve evidence both for and against identification, even if best-match selection was familiarity-based.

Previous tests of elimination lineups (Pozzulo & Lindsay, 1999) were promising, and we aimed to test the procedure in a more powerful experimental paradigm. In addition, previous findings suggested the delayed-choice lineup would encourage recollection in two ways. First, the reversed order of best-match and offender-presence judgments re-frames the approach to the lineup and counters eyewitness bias toward premature focus on the amount of positive match associated with the best-match. Instead, eyewitnesses would be tasked to weigh up evidence for and against the offender being present. Second, it incorporates a response lag, enforcing greater time for recollection to accumulate (Malmberg & Xu, 2007).

However, a wealth of evidence suggests that familiarity-based lineup decision strategies are difficult to disrupt (Brewer & Palmer, 2010) to more strongly cue recollection and monitoring. Therefore, we tested these two-step procedures with and without metacognitive instructions about lineup tasks and diagnostic qualities of accurate memories shown to increase recollection in recognition tasks (Lane, Roussel, Villa, & Morita, 2007). To our knowledge, this method had not previously been tested with lineups.

Metacognitive Instructions

Recognition is affected by metacognitive knowledge of the task and memory processes (McDermott & Roediger, 1998). Metacognitive beliefs vary in accuracy and task fit, and both retrieval and memory monitoring are shaped by decision-making strategies (Gallo, 2004). Poor belief-task fit is associated with less than optimal recognition strategies (Lane et al., 2007; Malmberg & Xu, 2007). For example, over-estimating the likelihood that the offender is in the lineup encourages an inappropriately liberal identification criterion (Wells & Seelau, 1995). However, these recognition findings also indicated that instructions can align witnesses' metacognitive knowledge with task requirements, suggesting that they might also improve identification reliability in lineup recognition.

Giving participants information about lineup task features in isolation from their implications for memory use does not consistently improve memory retrieval and monitoring (Gallo, 2004). For example, unbiased lineup instructions that the culprit may or may not be in the lineup encourage more conservative decision-making without better use of memory (Clark, 2005; Mickes et al., 2017). Metacognitive instructions derived from several dual-process theories of recognition to improve recollection target two key mechanisms: recollection retrieval; and monitoring diagnostic features of memories (Lane et al., 2007). Two types of instructions that target these mechanisms are potentially useful for lineups (Blank & Launay, 2014; Lane et al., 2007): task-related information about common memory errors (e.g., familiarity-based false recognition) and phenomenological markers of memory accuracy (e.g., vividness and clarity). These metacognitive interventions increase recollection and decrease false recognition and recall in various recognition tasks (Lane, Roussel, Starns, Vila, & Alonzo, 2008). Therefore, they were expected to improve recollection in lineup recognition, especially those structured to counter familiarity-based identification judgments (i.e., elimination and delayed-choice lineups). As multiple mechanisms contribute to false recognition and its reduction in memory tasks, the same is likely true for lineup tasks. Therefore, we tested both task-related and phenomenological instructions.

Task-related. When knowledge about how memory works in a given task is not fit for purpose, task-related 'how-to' instructions can often be used effectively. For example, attention can be switched to source evidence when misinformation has been received (Johnson, Hashtroudi, & Lindsay, 1993), and recollection can be effortfully retrieved in associative recognition (Gallo, Roediger, & McDermott, 2001; Starns, Lane, Alonzo, & Roussel, 2007). While task-related instructions are most effective when given before encoding (Gallo, 2010) they have also proven effective when given before retrieval in tasks with high target-foil similarity (Gallo et al., 2001; Starns et al., 2007). For lineups, this is crucial, as instructions always follow crime exposure.

Task-related instructions are more specific to memory and more directly linked to metacognitive action than abstract unbiased lineup instructions that warn the suspect might not be present but give no information about how to avoid memory errors. More specific information about potential errors is likely to preferentially target potential false identifications, rather than generally discourage choosing (Blank & Launay, 2014; Higham, Blank, & Luna, 2017). Effective lineup task-related instructions might best include information about lineup filler selection and lineup similarity, familiarity-based decision-making, and recollection.

Phenomenological. Task-related instructions are often supplemented with phenomenological information about the experience of remembering (Lane et al., 2007; 2008). This information cues attention to qualities associated with accurate memories in terms of relative vividness, clarity, and detail (Mather, Henkel, & Johnson, 1997). While task-related instructions alone have increased discriminability, Lane and colleagues' (2007; m2008) results with tasks featuring high target-foil similarity suggest phenomenological information contributed to this effect.

Phenomenological information drawing attention to the experience of remembering contrasts with more abstract descriptions of familiarity and recollection. Experiences of relative vividness and clarity correspond to experiences that arise from both recollection and familiarity, but participants attending to these qualities differentiate effectively between true and false recognition even when false recognition is subjectively compelling (Schooler Gerhard, & Loftus, 1996). By highlighting the diagnostic value of these qualities in differentiating between more accurate (informed by recollection) and less accurate (based on familiarity) memories, phenomenological instructions reduce false recognition, and have been effective in tasks for which task-related instructions were not consistently effective (Lane et al., 2007).

The Experiments

In four face recognition lineup experiments, we tested task-related and phenomenological instructions in standard, elimination, and delayed-choice simultaneous lineups. Participants studied target photos and made identification decisions for a four-person lineup for each target. This paradigm allowed collection of multiple data points per participant and stimulus to evaluate regular response patterns. In these preliminary tests we aimed to maximise generalisability across stimuli before proceeding to more realistic identification paradigms. Mansour and colleagues (2017) found minimal effect of completing multiple trials on lineup responses. Identification paradigms using only one or two crime stimuli require large samples and have limited generalisability across cases, despite greater ecological validity.

For standard lineups, participants either made a positive identification or indicated the target was not present. For elimination lineups, participants identified the best-match to memory for the target then indicated whether the best-match was the target. For delayed-choice lineups, participants indicated whether the target was in the lineup, then identified the target (if they had previously said that the target was present). Each lineup type was tested with and without metacognitive instructions.

We had three major research questions. First, would structural modifications in simultaneous lineups cue greater discriminability than in standard procedures? We expected bias toward identification would be countered by separating lineup judgments, and discriminability would be greater than for standard lineups. Second, would metacognitive instructions improve discriminability? We predicted metacognitive instructions would increase retrieval and use of recollection to improve discriminability. Third, would familiarity-based responding in standard lineups render instructions less effective than in modified lineups? We anticipated participants completing standard lineups would be less able to take advantage of instructions than those completing elimination or delayed-choice lineups and show less improvement in discriminability.

Experiment 1

We tested the effect of the delayed-choice lineup and metacognitive instructions on discriminability, predicting greater discriminability in delayed-choice than standard lineups, and with instructions than when none were provided.

Method

Except where indicated, all experiments followed the same method.

Design. Participants were randomly assigned to the between-subjects manipulation of lineup (standard; delayed-choice; delayed-choice-instructions). Target-presence was manipulated within subjects.

Participants. We aimed to collect data from 30 participants per condition to achieve sufficient power to test hypotheses with aggregate signal detection measures and mixed effects regression models (Judd, Westfall, & Kenny, 2017). Participants were undergraduate students participating for payment or course credit. In Experiment 1, 93 students (65 female, age: M = 23.30, SD = 6.56 years) participated. All collected data were analysed.

Materials. We used a lineup collection of 40 sets of five description-matched colour photos (*citation obscured for blind review*). Photos showed a front view of the head and neck with neutral facial expression and background (200 × 200 pixels) and were presented on an 18” monitor (resolution: 1920 × 1080 pixels). In each set, we randomly designated a target, target-replacement (for target-absent lineups), and three lineup fillers. Sets were used in two blocks of 20. In each block, half of the trials were target-present (target plus fillers) and half target-absent (target-replacement plus fillers). Lineups were presented as a horizontally centred row of four photos with response buttons below. A text cue (name or occupation, by block) centred horizontally above photos was presented with each target (study) and lineup (test) to provide associative information so that participants were looking for a specific studied target. Photo order in lineups, trial order in blocks, cue and target-presence assignment, were randomly determined for each participant. Name cue randomisation was constrained by gender.

Procedure. After giving informed consent, participants completed the computerised experiment in individual cubicles using a mouse to click on-screen buttons in response to written prompts. After four practice trials, participants completed two blocks of trials. Block order and block cue type were counterbalanced across participants. Blocks included a study phase, a visual distracter task, and a test phase. In the study phase, target photos were presented sequentially in the centre of the screen. Following exposure times established in stimulus piloting (*citation obscured for blind review*), cues were presented 1000 ms before target presentation and throughout target exposure (1000 ms). Following the study phase, participants worked on a spatial memory task for 3 min. In each test phase, participants completed 20 trials. Participants were told lineups may or may not include a target (unbiased instructions, Malpass & Devine, 1981). Lineup responses and latencies were recorded.

In the *standard* lineup condition, participants gave a single response to identify a lineup member as the studied target by clicking on the photo or clicked the *Not Present* button.

In the *delayed-choice* lineup condition, participants gave two identification responses to each lineup. The two decisions were blocked to prevent decision type subsequently affecting decision strategy. Participants gave their first response to each lineup in two blocks by indicating the studied face was *Present* or *Not Present*. To prevent guessing, participants were told they would make two responses for each lineup if they indicated the target was *Present*. After all trials were completed, lineups given a *Present* response were presented again in series, and participants clicked on the face identified as the target.

In the *delayed-choice-instructions* lineup condition, participants followed the same procedure as in the delayed-choice condition, but with additional instructions provided before test phases. Participants were given *task-related instructions* and *phenomenological instructions (*full instructions are included with supplementary materials). In task-related instructions, participants were told that lineups were description-matched, leading to high similarity lineups that include some relatively less similar fillers. They were further told that previous research showed that due to high lineup similarity "…*people sometimes mistakenly decide that the studied face is present when all four photographs show faces that have not been studied. When people focus on the similarity between test faces and the studied face, they pay less attention to evidence that the test faces are not the studied face."* In phenomenological instructions, we told participants that relatively clear, vivid memories, with recollected detail, are diagnostic of accuracy: "*Research has shown that accurate decisions about the studied face being present are likely to be accompanied by more vivid and clear memories of studying the face than when an unstudied test face is inaccurately and mistakenly recognised. These can include memories of how the studied face looked, memories of how the name/occupation and the studied face are related, and other thoughts and feelings that were experienced when the face was studied."* Phenomenological instructions were repeated when participants were instructed to use this information to improve the accuracy of their lineup judgments and midway through test blocks: "*We would like you to avoid mistakenly and inaccurately recognising a test face that is similar to the studied face but was not studied."*.

 doi:10.20944/preprints201812.0012.v1

Following the final test phase in all experiments, all participants completed a manipulation check questionnaire including three open-ended items, one option-ranking item, and 17 items with 7-point Likert response scales. Questions tested recall of instructions and self-reported decision-making (complete questionnaire included with supplementary materials). As the modified recognition tasks were complex and required multiple recognition responses, we did not measure decision confidence or use ratings of familiarity or recollection in these experiments for two reasons. First, decision confidence for different recognition responses would not be directly comparable. Second, we wanted to avoid any effect of providing decision confidence or familiarity and recollection ratings on participants' metacognitive approach to test lineups in this multiple trial paradigm.

Results & Discussion

Overview and measures. We used an alpha level of .05 for planned comparisons. The main dependent measures were 'suspect' identifications (of targets and target-replacements), and overall choosing from a lineup (identify any lineup member). Aggregate descriptive statistics for identification responses and signal detection measures of discriminability (*d'*), response bias (*C*), and Cohen's *d* for key comparisons of discriminability are shown in Table 1.

For datasets including multiple participants' responses to multiple stimuli aggregate analyses can be biased and do not appropriately account for variability (Judd, Westfall, & Kenny, 2012). Therefore, we made planned comparisons of response bias and discriminability with mixed effects logistic regression models predicting choosing or 'suspect' identifications, equivalent to variance (ANOVA) models of *C* and *d'* (Wright & London, 2009). Each model indicated response bias (predictor: lineup); discriminability (predictor: target-presence); and differences in discriminability between experimental conditions (e.g., between lineup conditions, predictor: lineup × target-presence interaction term, Wright & London, 2009). Coefficients for each predictor represent the log odds ratio (*lnOR)* of the outcomes. Following convention, we report odds ratios (*OR*) which indicate effect size. For significant effects, the 95% confidence interval (95% CI_{OR}) excludes 1.

In models analysing choosing (including filler selections), we evaluated differences between target-absent and target-present lineups, and between lineups. As filler identifications are largely familiarity-based (due to similarity to the target), these models indicated whether differences between conditions in rates of accurate identification reflected overall accuracy (as a result of better recollection), or a different pattern of errors (e.g., a greater proportion of filler selections).

Regression models included all fixed factors of interest. Random effects and model fit statistics are reported in supplementary materials. Data were analysed using the lme4 package (Bates, Maechler, Bolker, & Walker, 2014) for R (R Core Team, 2014).

Delayed-choice. We compared standard and delayed-choice lineups to evaluate whether delayed-choice procedures produced greater discriminability of targets from other lineup members (Table 2, Figure 1). In models predicting choosing overall, there were no significant differences between lineup types in response bias (predictor: lineup) or choosing patterns within target-present or target-absent lineups (predictor: lineup × target presence), though response bias tended to be lower. Looking only identifications of target and target-replacements from lineups (excluding filler selections) there was significantly lower response bias for delayed-choice lineups (lineup, *OR* = 0.79) but no significant difference in discriminability (predictor: lineup × target presence). Results showed participants completing delayed-choice lineups were less likely to identify studied targets and target-replacements, than those completing standard lineups, but not better at discriminating between targets and target-replacements.

We expected the delayed-choice task to encourage recollection. However, results suggest that delayed-choice influenced decision strategies or memory use without improving recollection. One explanation is that delayed-choice did not disrupt participants' metacognitive approach to the lineup but reduced accuracy in best-match selection. Here, it is possible that separating the component lineup judgements (target-presence; best-match selection) in the multiple trial paradigm, involving greater exposure to lineup members at test, created interference that made it difficult for participants to accurately distinguish the source of familiarity of filler lineup members and targets. As a result, accurate target selections were more difficult. This interpretation draws some support from examination of first-decision accuracy (as shown in models of choosing overall), in which response bias was less affected by lineup type. Alternatively, type of recognition response did not affect multiple-choice decision-making for recognition arrays. Or, recollection might have been used in both conditions, with lineup type affecting only criterion placement, or best-match selection accuracy. In summary, results did not support the idea that dividing and reversing lineup judgments would promote better recollection use.

Metacognitive instructions. We tested the effectiveness of metacognitive instructions in delayed-choice lineups, proposing instructions would reduce familiarity-based decision-making and promote recollection. Responses to the manipulation check questionnaire in all experiments were generally consistent with participants attending to instructions (descriptive statistics and pairwise comparisons for all experiments are reported in the online supplement). Comparison of standard and instructions lineups showed significantly lower response bias for instructions lineups (lineup: target identifications, OR = 0.09; choosing, OR = 0.89), with no significant difference in discriminability (predictor: lineup × target presence,

Table 2, Figure 1). Participants given metacognitive instructions in delayed-choice lineups were less likely to choose than those completing standard lineups but did not more accurately differentiate targets from target-replacements. This effect was largely due to fewer false identifications (i.e., toward greater discriminability).

Comparing patterns of target/target-replacement identifications in delayed-choice lineups with and without instructions, there was significantly greater discrimination with instructions (predictor: lineup × target presence , *OR* = 2.16, Table 2, Figure 1). Importantly, there was no significant difference in accurate target identifications, but instructions helped delayed-choice lineup participants to avoid false identifications more effectively than when no instructions were given. This finding is notable given the already lower response bias in delayed-choice than standard lineups. Even with this lower risk of false identification, instructions further lowered the risk for innocent 'suspects'.

Experiment 2

Eyewitness identification studies show that unbiased lineup instructions and strong warnings against false identification produce conservative responding (Mickes et al., 2017). Task-related instructions in Experiment 1 might have similarly highlighted this risk, producing a warning effect that overshadowed the potential benefit of phenomenological information. Therefore, in Experiment 2 we cut task-related instructions to focused on phenomenology. We also tested the revised metacognitive instructions with standard lineups, predicting significantly better discriminability with instructions than without.

We replicated the between-subjects comparison of delayed-choice and standard lineups from Experiment 1 and added within-subjects tests of instructions with both lineups. Participants completed 40 test trials in either the standard or delayed-choice condition, then a further 40 test trials with instructions in the same lineup type. We created additional lineups

 doi:10.20944/preprints201812.0012.v1

to increase stimulus variability and generalizability of results. In additional face sets, target appearance varied between study and test, providing a strong test of manipulations (Hancock, Bruce, & Burton, 2000).

Method

The method was as in Experiment 1, except as indicated below.

Design. Participants were randomly assigned between-subjects to lineup (standard; delayed-choice). Target-presence and instructions (instructions; none) were manipulated within subjects.

Participants. We tested 64 participants (47 female, age: M = 21.97, SD = 6.11 years).

Materials. We used the original collection of lineups, supplemented by a second collection of 40 sets constructed using the same method. In new lineups, test photos of targets varied from study photos. As in original lineups, target photos in lineups showed a front view of the head and neck with neutral facial expression. However, target photos at study showed either a front view of the head and neck with a different facial expression (smiling), or a three-quarter view of the head and neck with neutral expression. Study-test variation makes face recognition more challenging (Hancock et al., 2000), providing a stronger test of modified procedures, and more closely approximating encoding-test variations of eyewitnesses.

Procedure. After giving informed consent, participants completed four blocks of 20 lineup trials.

Lineup. Lineup type was manipulated on two levels between-subjects (standard; delayed-choice) and instructions on two levels within-subjects (instructions; none). Block order of lineup collections was counterbalanced within lineup and instructions conditions. *Standard* and *delayed-choice* lineup conditions were as in Experiment 1, but for the additional within-subjects manipulation of instructions.

Instructions. Instructions were varied on two levels within-subjects (metacognitive; standard*)*. In the first two blocks, participants followed the procedure for their respective lineup condition. Then, before the test phases in two further blocks of lineup trials, all participants were given metacognitive instructions as in Experiment 1 but with *task-related* references to lineup structure and similarity-based identification errors removed. *Phenomenological* instructions were unchanged. Participants were instructed to use this information to make accurate identification decisions: "*If you think the test photograph is similar to your memory for the studied face, but your memory is not clear, vivid or detailed, you should response Not Present."* That is, one instruction to avoid similarity-based positive identification decisions remained.

Results & Discussion

Descriptive statistics are presented in Table 1. To replicate the Experiment 1 comparison of delayed-choice and standard simultaneous lineups, we compared response patterns in these conditions *without* instructions and found no significant differences in response bias or discriminability (Table 3, Figure 2). However, in each model, the magnitude of the coefficient for lineup was not trivial, and CIs largely included values below zero, suggesting meaningfully lower response bias with delayed-choice than standard lineups, as in Experiment 1.

Delayed-choice lineups with instructions were then compared separately with delayed-choice (no instructions) and standard (no instructions) lineup conditions (Table 3, Figure 2). Participants completing delayed-choice lineups were significantly less likely to make an identification (accurate or not) from the lineup when instructions were given than when they were not (lineup: target identifications, $OR = 0.50$; choosing overall, $OR = 0.58$) or than when standard lineups procedures were followed without instructions (lineup: target identifications,

 doi:10.20944/preprints201812.0012.v1

OR = 0.41; choosing overall, *OR* = 0.39). However, comparison with standard lineups showed no significant difference in discriminability. Contrary to expectations, delayed-choice lineups with revised instructions did not encourage better recollection than standard lineups, though choosing was less biased.

In contrast, and following Experiment 1 results, discriminability was meaningfully improved by instructions in delayed-choice lineups (Table 3, Figure 2). While this effect was not significant (predictor: lineup × target presence) the coefficient was substantial (choosing overall: *b* = 0.29; target identifications: *b* = 0.49), with *OR* CIs biased strongly toward values falling above 1. This indicated that participants in the delayed-choice condition differentiated guilty from innocent suspects meaningfully better when given metacognitive instructions, although the effect was not statistically significant. Notably, as the effect was within-subjects, instructions had meaningfully improved memory use after participants had earlier established less optimal familiarity-based decision-making in trials completed without instructions. While practice effects were a potential threat (as we could not sensibly counterbalance instructions conditions), there was no equivalent effect of instructions in standard lineups, as discussed below.

To evaluate the effectiveness of instructions with standard lineup procedures, we compared standard lineup identification performance with and without instructions (Table 3, and Figure 2). There was significantly lower response bias with instructions (lineup: choosing, *OR* = 0.55; target identifications, *OR* = 0.41), and discriminability of targets from target-replacements was significant lower when instructions were provided (predictor: lineup × target presence, *OR* = 0.46). Lower response bias with instructions in standard lineups limited the scope for further reductions in false identification more than it reduced the scope for further reductions in target identification. Therefore, reduced discriminability might have been an artefact of floor effects. However, the lack of differences between target-absent and target-present lineups in choosing or not (other than lower bias for both) suggests instead that participants were making worse identification decisions in target-present lineups (i.e., a greater proportion of filler selections with instructions), rather than only identifying fewer targets. Regardless, instructions did not significantly improve discriminability in standard lineups but reduced choosing. As target picks were reduced more than filler selections from target-present lineups, and the pattern of responses differed between instructions conditions in delayed-choice lineups, this was unlikely to be a practice effect.

In sum, instructions reduced response bias in standard lineups without improving discriminability. The pattern of effects across lineup conditions broadly replicated between-subjects results in Experiment 1 observed for more heavily task-related instructions.

Experiment 3

In Experiments 1 and 2, we found meaningfully greater discriminability when instructions were given with delayed-choice lineups than without instructions, but the effect was not significant. The design provided adequate power to detect a non-trivial effect (Judd et al., 2017), therefore in a third experiment we revised the instructions to remove remaining references to lineup similarity and false identification, leaving only phenomenological information about memory accuracy. In Experiment 3, lineup type was varied between-subjects on three levels: standard; *standard-instructions* (standard lineup with phenomenological instructions)*;* and *recollection* (delayed-choice lineup with phenomenological instructions). Varying lineup between-subjects allowed us to test whether lower response bias with instructions in Experiment 2 was due to instructions being given following blocks completed without instructions, rather than to their specific content. We expected the impact of revised instructions to be greater with delayed-choice procedures when participants had not established familiarity-based decision-making, than with standard procedures.

Method

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

Materials and Procedure. We used both lineup collections. After giving informed consent, each participant completed four blocks of 20 lineups. Lineup was manipulated on three levels between-subjects (standard; standard-recollection; recollection). The order of lineup collections was counterbalanced within lineup and instructions blocks.

The *standard* lineup condition was as in Experiment 1, but for the additional two blocks of trials.

The *standard-recollection* condition was identical to the standard lineup condition, except that participants were also given *phenomenological instructions* before test phases. Instructions were identical to those provided in Experiment 2, except the explicit instruction to avoid similarity-based identifications was removed (i.e., "*If you think the test photograph is similar to your memory for the studied face, but your memory is not clear, vivid or detailed, you should response Not Present.")*.

The *recollection* lineup condition was as for the delayed-choice condition with instructions identical to those provided in the *standard-recollection* condition.

Results & Discussion

Recollection lineups (delayed-choice with instructions) showed significantly greater discriminability than standard lineups (target presence × lineup: target identifications, *OR*= 0.53; OR; choosing, *OR* = 0.44, Table 4, Figure 3). False identification was significantly less likely in recollection lineups (odds of false identification $_{Recollection}$ = 0.33; odds $_{Standard}$ = 0.67), but there was no significant difference between the lineups in target identifications (odds $_{Recollection}$ = 1.70; odds $_{Standard}$ = 2.23). As predicted, witnesses made significantly better use of memory when given instructions in delayed-choice lineups than in standard lineups without instructions.

In stark contrast, comparison of standard and standard-instructions lineups revealed no significant differences in identification performance (see Table 4). Providing instructions in standard lineups had no significant effect on either response bias or discriminability when compared with standard lineups without instructions (false identifications, *OR*= 1.00; target identifications, *OR*= 0.96).

Summary of Results for Experiments 1-3

Contrary to evidence from recognition memory paradigms (e.g., Blank & Launay, 2014; Gallo, 2010; Starns et al., 2007), task-related information highlighting lineup mechanics did not scaffold better use of memory, and we found a smaller effect of instructions in Experiments 1 and 2 than in Experiment 3 when task-related instructions were not included with phenomenological information. In Experiment 3, phenomenological instructions with delayed-choice lineup procedure (the recollection lineup) facilitated more effective discrimination of targets from target-replacements than in standard simultaneous lineups. Results are consistent with continuous dual-process accounts of familiarity-based recognition (Wixted & Mickes, 2010). Countering best-match decision-making and encouraging better use of recollection in the recollection lineup improved lineup reliability.

However, results also showed that neither the delayed-choice lineup or phenomenological instructions independently improved discriminability. Only in combination did these modifications reframe participants' metacognitive approach to improve memory use.

Experiment 4

To further test the dual-process explanation, we tested elimination lineups with and without the phenomenological instructions from Experiment 3. Prior research (Pozzulo & Lindsay, 1999) suggested that separating best-match and offender-presence judgements would prevent participants from carrying over familiarity-based memory processes used to select a best-match from the lineup to their judgment about whether the best-match was the offender. We predicted elimination lineups would be associated with greater discriminability than standard simultaneous lineups. In Experiments 1 to 3 we found no evidence that separating component judgments in delayed-choice lineups (without metacognitive instructions) had improved recollection. Therefore, if elimination lineups improved recollection, reordering judgments had interfered with any benefit of separating them.

We originally proposed that phenomenological instructions would be less effective when given to participants in lineup procedures that prioritise familiarity-based best-match selection than lineups countering this approach. Accordingly, we expected instructions in elimination lineups to more effectively increase recollection than in standard lineups. Results from Experiments 1-3 showed instructions were effective when provided with delayed-choice procedures requiring serial lineup judgements, even when these procedures were not effective without instructions. Therefore, we expected instructions would also be more effective in elimination lineups than standard lineups, even though the order of component judgments was unchanged.

In Experiment 4, we varied lineup type (standard; elimination) and instructions (phenomenological; none) in a fully between-subjects design. If separating component

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2018 doi:10.20944/preprints201812.0012.v1

judgments allowed participants to use memory better, we expected greater discriminability when instructions were provided with elimination lineups than in standard lineups (no instructions). Further, we expected instructions would produce greater discriminability in elimination than when provided with standard lineups.

Method

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

Materials and procedure. We used original lineup sets. Lineup condition was manipulated on three levels between-subjects (standard; elimination; elimination-instructions).

The *standard* lineup condition was as in Experiment 1.

The *elimination* lineup condition was a two-step procedure comprising a *best-match* judgment and a *target presence* judgment. Participants made both decisions for each lineup before proceeding to the next lineup. For the best-match judgment, participants clicked on the lineup member best matching the studied face. Once selected, remaining faces disappeared from the screen, leaving the best-match on-screen for an additional 1500 ms. For the second judgment, participants indicated whether the best-match selection was the target (*Yes; No)*.

The *elimination-instructions* condition was identical to the *elimination* condition except that before each test phase participants were given the *phenomenological instructions* used in Experiment 3.

Results & Discussion

Descriptive statistics are presented in Table 1. We found no significant difference in response bias or discriminability between standard and elimination lineups (Tables 5, Figure

4). Thus, the elimination lineup structure had no significant effect on the quality of lineup decision-making. Comparisons of standard lineups with and without instructions also provided no evidence that instructions improved memory use (Table 5, Figure 4). There was no significant effect of lineup on response bias or discriminability. Likewise, comparison of elimination lineups with and without instructions (Table 5, Figure 4) showed no significant difference in response bias or discriminability. Effect size indices suggested a small detriment to discriminability, if anything, in the elimination-instructions condition. In sum, we found no significant differences in choosing or discriminability between any conditions, and no evidence that separating judgments in elimination lineups facilitated the efficacy of phenomenological instructions.

General Discussion

We extended a continuous dual-process model of recognition (Wixted & Mickes, 2014) to lineup tasks to explain the effects of task structure and eyewitness metacognition on identification decision-making. Specifically, we argued that eyewitness bias to choose from the lineup is associated with reliance on familiarity-based positive match to memory and neglect of critical recollection that would rule out false identification. Across four experiments we found evidence that the recollection lineup featuring a serial decision structure and instructions about memory phenomenology reduced false identification from simultaneous lineups without significantly reducing accurate identifications.

Metacognitive instructions improved identification performance

These results provide the first demonstration of an effective metacognitive intervention in lineups. Participants informed about the qualities of accurate remembering discriminated more effectively between target and unstudied faces in recollection lineups, demonstrating a similar contribution of metamemory as in other recognition tasks requiring recollection, but only when the task followed a delayed-choice procedure. Two findings are important for lineups: eyewitness memory can be enhanced with metamemorial instructions; and the 'recollection lineup', combining delayed-choice with phenomenological instructions, did not only encourage more cautious choosing, it improved discriminability.

Instruction Type. Results also provided the first indication of the type of metacognitive instructions that are effective in lineups. Discriminability was increased when participants were given phenomenological instructions about the qualities of accurate memories (i.e., relatively greater vividness, clarity, and detail). In contrast, task-related information led participants only to choose more cautiously. Greater discriminability with phenomenological instructions was evident only with delayed-choice, not with elimination or standard simultaneous lineups.

Taken together, these findings are consistent with the idea that familiarity-based best-match selection in simultaneous lineups is resistant to intervention, and witnesses neglect the role of recollection. Task-related information highlighting lineup mechanics and increasing awareness of the risk of false identification may have reinforced eyewitness bias toward a singular focus on finding and evaluating the best-match in the lineup and worked against greater recollection. If so, metacognitive instructions would only be effective in procedures that do not strongly cue familiarity-based decision-making by prioritising best-match selection. Alternatively, participants might not have attended to more common-sense phenomenological information when relatively novel task-related elements were included in instructions (e.g., Kahnemann, 2011).

The question of which retrieval or post-retrieval mechanisms contributed to the recollection lineup effect has implications for applications of phenomenological instructions in diverse cases. Phenomenological instructions would allow eyewitnesses to better evaluate

 doi:10.20944/preprints201812.0012.v1

familiarity and recollection if 1) either lineup procedure or instructions cued recollection and/or 2) instructions improved memory monitoring. If substantial recollection was not needed (e.g., the distinctiveness heuristic, Schacter, Israel & Racine, 1999) recollection lineups might be effective even when recollection is not available. However, when phenomenological instructions were provided in lineup procedures that did not encourage recollection (elimination and standard simultaneous lineups), improved monitoring might not have been effective as mainly familiarity had been retrieved. If so, and if monitoring improvements required recollection (e.g., Guerin, Robbins, Gilmore, & Schacter, 2012), these instructions would not be useful under poor encoding or retention conditions. We argue that multiple mechanisms contribute to recognition errors and their prevention, and the best mechanisms for improving memory will vary across offenders, lineups, and witnesses. Such questions are an important focus for future investigation.

The recollection lineup

Features of the recollection lineup (combining delayed-choice procedures with phenomenological instructions) are promising but need further testing and investigation. It represents a starting point for the development of further modifications to identification tests that might more tightly control witness variables and use witnesses' cognitive capacity to increase reliability. However, implicit metacognitions will be more highly salient and potentially influential in real lineups. Therefore, rigorous testing of novel tests in identification paradigms would be needed. Similarly, the effects of extraneous cues of familiarity-based decision-making (e.g., administrator or co-witness bias) must be examined with any novel procedure. A procedure that relies on metacognitive influence for its efficacy will not be valuable if known influences on metacognition are not also controlled.

Implications for recognition and lineup theory

Our results have important implications for theories of eyewitness identification tasks and task development. Findings support the idea of familiarity-based best-match selection processes in simultaneous lineups with associated neglect of recollection and add to evidence that task-type and metacognitive interventions can improve memory use in identification decision-making. While the recollection lineup itself has potential applications, other interventions in diverse test formats should also be investigated based on further theoretical development (e.g., Wixted et al., 2018). Overall, our results reaffirm the value of extending basic memory theory directly to complex applied lineup tasks.

We also extended research on metacognitive instructions by demonstrating that providing more accurate information about a task is not always better than not providing any task information. Blank and Launay (2014) suggested the effectiveness of task-related instructions increased with specificity. However, we found greater specificity affected only decision criterion and not discriminability. More specific instructions might sometimes counteract the benefits of otherwise enhanced metacognitive awareness, while process-focused phenomenological instructions were effective.

Conclusions. Results were consistent with a dual-process account of simultaneous lineup decision-making. The novel recollection lineup reduced false identification and increased discriminability by diverting participants from relying on familiarity and enhancing metacognitive awareness of how memory quality can be used to diagnose memory accuracy. The effectiveness of the recollection lineup shows that basic recognition memory theory can explain eyewitness memory use and inform designs for improved identification tests. We also demonstrated that providing accurate task information could nevertheless trigger error-prone decision-making processes. The effectiveness of metacognitive instructions depended on how they interacted with participants' implicit understandings and task-related decision processes. Overall, we demonstrated that the delayed-choice structure and phenomenological instructions did not independently improve discriminability in simultaneous lineups but were effective in combination. Hence, decision structure and instructions must work in concert for interventions targeting eyewitness metacognition to increase the reliability of identification decisions.

References

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

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

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

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

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

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

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

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

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

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

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

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

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

Higham, P. A., Blank, H., & Luna, K. (2017). Effects of postwarning specificity on memory performance and confidence in the eyewitness misinformation paradigm. Journal of experimental psychology: applied, 23(4), 417. doi: 10.1037/xap0000140

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

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

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

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. Annual Review of Psychology, 68, 601-625. doi: 10.1146/annurev-psych-122414-033702

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

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

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

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

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

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

Mansour, J. K., Beaudry, J. L., & Lindsay, R. C. L. (2017). Are multiple-trial experiments appropriate for eyewitness identification studies? Accuracy, choosing, and confidence across trials. Behavior research methods, 49(6), 2235-2254. doi: 10.3758/s13428-017-0855-0

Mather, M., Henkel, L. A., & Johnson, M. K. (1997). Evaluating characteristics of false memories: Remember/know judgments and memory characteristics questionnaire compared. *Memory & Cognition*, *25*(6), 826-837. doi: 10.3758/bf03211327

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

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

Mickes, Laura, et al. "ROC s in Eyewitness Identification: Instructions versus Confidence Ratings." Applied Cognitive Psychology 31.5 (2017): 467-477.

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

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

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

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

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

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

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

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

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

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

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

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

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

Wixted, J. T., Mickes, L., Dunn, J. C., Clark, S. E., & Wells, W. (2016). Estimating the reliability of eyewitness identifications from police lineups. Proceedings of the National Academy of Sciences, 113(2), 304-309..

Wixted, J. T., Mickes, L., & Fisher, R. P. (2018). Rethinking the reliability of eyewitness memory. Perspectives on Psychological Science, 13(3), 324-335.

Wixted, J. T., Vul, E., Mickes, L., & Wilson, B. M. (2018a). Models of lineup memory. Cognitive psychology, 105, 81-114.

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

Table 1

Descriptive Statistics for Lineup Responses (Accurate Identification (ID) rates (Hit rates), Filler ID Rates, Correct Rejection Rates, False ID Rates, Overall Accuracy) and Aggregate Signal Detection Measures (d', c) by Experiment and Lineup Condition, Experiments 1-4*

			Response Proportion									Signal Detection Measures		Lineup Comparison (vs S)
			Target-present				Target-absent				All Trials			
Expt	Lineup	Statistic	Hits	Filler IDs	Incorrect Rejections	*n*	Correct Rejections	Filler IDs	False IDs	*n*	Overall accuracy	*d'*	*C*	Cohen's *d* for *d'*
1	Standard (S)	*M*	.65	.12	.23	30	.64	.24	.13	30	0.65	1.51	0.37	-
		SD	.16	.11	.14		.24	.16	.11		0.20	0.81	0.14	-
		95% CI	.64, .66	.11, .13	.22, .24		.62, .65	.22, .25	.12, .14		.64, .66	1.23, 1.80	0.32, 0.42	-
	Delayed-choice (DC)	*M*	.50	.18	.32	30	.69	.19	.12	30	0.60	1.12	0.59	0.48
		SD	.19	.13	.13		.20	.15	.09		0.20	0.81	0.11	-

doi:10.20944/preprints201812.0012.v1

		95% CI	.48, .52	.17, .19	.31, .33		.67, .70	.18, .21	.11, .13		.59, .61	0.83, 1.51	0.55, 0.63	-
	Instructions	*M*	.57	.10	.33	33	.79	.13	.08	33	0.68	1.58	0.61	0.09
	(I)	*SD*	.19	.11	.17		.16	.14	.07		0.18	0.79	0.11	-
		95% CI	.56, .59	.09, .11	.31, .34		.78, .80	.12, .14	.08, .09		.67, .69	1.30, 1.86	0.57, 0.63	-
2	S	*M*	.51	.15	.34	32	.63	.26	.11	32	0.57	1.25	0.31	-
		SD	.19	.10	.18		.21	.17	.09		0.20	0.54	0.11	-
		95% CI	.50, .53	.14, .16	.32, .35		.61, .65	.24, .27	.10, .12		.56, .58	1.05, 1.44	0.27, 0.35	-
	SI	*M*	.43	.13	.44	32	.70	.24	.06	32	0.57	1.40	0.24	0.30
		SD	.21	.09	.24		.25	.22	.06		0.17	0.47	0.11	-
		95% CI	.41, .45	.12, .14	.42, .46		.68, .72	.22, .26	.05, .06		.56, .58	1.16, 1.64	0.20, 0.28	-
	DC	*M*	.47	.13	.40	32	.71	.21	.08	32	0.59	1.34	0.28	0.10
		SD	.15	.09	.16		.15	.13	.07		0.15	0.55	0.08	-
		95% CI	.46, .48	.12, .13	.39, .41		.70, .73	.20, .22	.07, .08		.58, .60	1.15, 1.53	0.25, 0.31	-
	DCI	*M*	.45	.10	.45	32	.79	.16	.05	32	0.62	1.49	0.25	0.48
		SD	.19	.11	.21		.17	.14	.05		0.18	0.45	0.10	-

doi:10.20944/preprints201812.0012.v1

		95% CI	.43, .46	.09, .11	.44, .47		.77, .80	.15, .17	.05, .06		.61, .63	1.33, 1.65	0.22, 0.28	-
3	S	*M*	.49	.18	.32	32	.59	.30	.12	32	0.54	1.24	0.31	-
		SD	.15	.12	.16		.17	.14	.07		0.16	0.51	0.09	-
		95% CI	.48, .50	.18, .19	.32, .33		.58, .59	.29, .33	.12, .12		.53, .55	1.06, 1.42	0.28, 0.34	-
	DCI	*M*	.50	.12	.37	32	.72	.19	.09	32	0.61	1.42	0.32	0.19
		SD	.14	.09	.14		.16	.11	.06		0.15	0.48	0.08	-
		95% CI	.50, .51	.12, .13	.37, .38		.71, .73	.18, .19	.09, .09		.60, .62	1.25, 1.59	0.29, 0.35	-
	SI	*M*	.52	.17	.31	32	.58	.29	.12	32	0.55	1.29	0.30	0.10
		SD	.13	.14	.11		.17	.14	.08		0.15	0.53	0.07	-
		95% CI	.51, .53	.16, .18	.31, .32		.58, .59	.29, .33	.12, .13		.54, .56	1.11, 1.47	0.28, 0.32	-
4	S	*M*	.43	.16	.40	45	.73	.20	.07	45	0.58	1.29	0.25	-
		SD	.17	.14	.15		.20	.16	.07		0.19	0.67	0.08	-
		95% CI	.42, .44	.15, .17	.39, .41		.72, .75	.19, .21	.07, .08		.57, .59	1.09, 1.49	0.23, 0.27	-
		M	.39	.13	.48	33	.71	.21	.07	33	0.55	1.09	0.23	0.32

Elimination	*SD*	.17	.13	.19		.20	.15	.07		0.19	0.57	0.10	-
(E)	95% CI	.38, .40	.12, .14	.46, .49		.70, .73	.20, .22	.07, .08		.54, .56	0.90, 1.28	0.20, 0.26	-
EI	*M*	.40	.14	.46	34	.70	.23	.08	34	0.55	1.16	0.24	0.18
	SD	.19	.15	.19		.20	.17	.08		0.20	0.75	0.09	-
	95% CI	.39, .41	.13, .15	.45, .48		.68, .71	.21, .24	.07, .09		.54, .56	0.91. 1.41	0.21, 0.27	-

* Note: Inferential analysis of aggregate data give biased estimates of effects when both participant and stimuli are random factors (Judd, Westfall, & Kenny, 2012). Refer to mixed effects models for tests of experimental factors.

Table 2

Models Predicting Identifications (Lineup Rejection=0; Identification=1) by Lineup Comparison and Model, Experiment 1

		Lineup Comparisons								
		#Standard vs DC			#Standard vs Instructions			#DC vs Instructions		
Outcome	Predictor	*b* (95% CI)	*OR* (95% CI)	SE_b	*b* (95% CI)	*OR* (95% CI)	SE_b	*b* (95% CI)	*OR* (95% CI)	SE_b
Choosing overall (filler selections included)	Intercept	-0.70			-0.70			-0.91		
	Lineup	-0.23 (-0.82, 0.36)	0.79 (0.44, 1.43)	0.30	-0.89* (-1.49, -0.28)	0.41* (0.23, 1.32)	0.31	-0.63* (-1.16, -1.03)	0.53* (0.31, 2.80)	0.27
	TP	1.88 * (1.51, 2.47)	6.55* (4.53, 11.82)	0.25	2.03* (1.55, 2.52)	7.61* (4.71, 12.43)	0.25	1.74* (1.26, 2.22)	5.70* (3.53, 9.21)	0.24
	Lineup × TP	-0.24 (-0.93, 0.45)	0.79 (0.39, 1.57)	0.35	0.34 (-0.34, 1.04)	1.40 (0.71, 2.83)	0.35	0.58 (-0.10, 1.25)	1.79 (0.90, 3.49)	0.34
Target ID (filler selections excluded)	Intercept	-0.70			-0.70			-0.92		
	Lineup	-0.23* (-0.81, 0.35)	0.79* (0.44, 1.42)	0.30	-0.90* (-1.11, -0.28)	0.41* (0.33, 1.32)	0.31	-0.63* (-1.16, -1.02)	0.53* (0.31, 2.77)	0.27
	TP	1.83*	6.23*	0.27	1.86*	6.42*	0.27	1.38*	3.97*	0.27

	(1.30, 2.36)	(3.67, 10.59)		(1.31, 2.39)	(3.71, 10.91)		(0.84, 1.91)	(2.32, 6.75)	
Lineup ×	-0.44	0.64	0.38	0.35	1.42	0.39	0.77*	2.16*	0.38
TP	(-1.19, 0.31)	(0.30, 1.36)		(-0.42, 1.11)	(0.66, 3.03)		(0.02, 1.52)	(0.98, 4.57)	

Note: * indicates significant predictors, # indicates the reference lineup condition. Number of observations, fillers included (fillers excluded):

Standard vs DC = 2400 (2220), n = 60 (60); Standard vs Instructions = 2520 (2384), n = 63 (63); DC vs Instructions = 2520 (2344), n = 63 (63)

Table 3

Models predicting Identifications (Lineup Rejection=0; Identification=1) by Lineup Comparison and Model, Experiment 2

		Lineup Comparisons											
		#Standard vs Delayed-Choice			#Standard vs Delayed-Choice-Instructions			#Standard vs Standard-Instructions			#Delayed-Choice vs Delayed-Choice-Instructions		
Outcome	Predictor	*b* (95% CI)	*OR* (95% CI)	SE_b	*b* (95% CI)	*OR* (95% CI)	SE_b	*b* (95% CI)	*OR* (95% CI)	SE_b	*b* (95% CI)	*OR* (95% CI)	SI
Target ID (filler selections excluded)	Intercept	-2.18			-2.16			-2.07			-2.61		
	Lineup	-0.55 (-1.16, 0.06)	0.58 (0.32, 1.05)	0.31	-0.89* (-1.54, -0.24)	0.41 (0.21, 0.79)	0.33	-0.90* (-1.39, 0.41)	0.41 (0.25, 1.49)	0.25	-0.70* (-1.27, -0.13)	0.50 (0.28, 0.87)	0.2
	TP	2.63* (2.18, 3.08)	13.87 (8.85, 21.76)	0.23	2.65* (2.22, 3.08)	14.15 (9.12, 21.98)	0.22	2.54* (2.09, 2.99)	12.68 (8.08, 19.89)	0.23	2.74* (2.29, 3.19)	15.49 (9.87, 24.29)	0.2
	Lineup × TP	0.26 (-0.33, 0.85)	1.30 (0.73, 2.34)	0.30	0.37 (-0.24, 0.98)	1.45 (0.79, 2.66)	0.31	-0.78* (-0.19, -1.37)	0.46 (0.25, 0.83)	-.30	0.49 (-0.08, 1.06)	1.63 (0.57, 1.06)	0.2
Choosing overall (filler selections included)	Intercept	-0.62			-0.64			-0.67			-1.04		
	Lineup	-0.41 (-0.86, 0.04)	0.66 (0.42, 1.04)	0.23	-0.93* (-1.46, -0.40)	0.39 (0.36, 0.77)	0.27	-0.59* (-1.00, -0.18)	0.55 (0.37, 0.83)	0.21	-0.54* (-0.85, -0.23)	0.58 (0.43, 0.79)	0.1
	TP	1.36* (1.03, 1.69)	3.90 (2.77, 5.42)	0.17	1.41* (1.04, 1.78)	4.10 (2.83, 5.93)	0.19	1.41* (1.02, 1.80)	4.10 (1.65, 6.05)	0.20	1.46* (1.13, 1.79)	4.31 (3.86, 5.93)	0.1

Lineup × TP	0.12 (-0.33, 0.57)	1.13 (0.71, 1.79)	0.23	0.34 (-0.17, 0.85)	1.40 (0.84, 2.34)	0.26	-0.11 (-0.56, 0.34)	0.90 (0.70, 1.79)	0.23	0.29 (-0.08, 0.66)	1.34 (0.92, 1.92)	0.

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

doi:10.20944/preprints201812.0012.v1

Table 4

Models Predicting Identifications (Lineup Rejection=0; Identification=1) by Lineup Comparison and Model, Experiment 3

		Lineup Comparisons					
		#Standard vs Delayed-Choice			#Standard vs Standard-Instructions		
Outcome	Predictor	*b* (95%CI)	*OR* (95%CI)	SE_b	*b* (95%CI)	*OR* (95%CI)	SE_b
Target ID (filler selections excluded)	Intercept	-0.40			-0.39		
	Lineup	*-0.71 (-1.11, -0.30)	0.49 (.033, .74)	0.21	-0.00 (-0.39, -0.39)	1.00 (0.68. 1.48)	0.20
	TP	*0.83 (0.47, 1.18)	3.32 (2.44, 4.48)	0.18	0.83 (0.52, 1.14)	3.29 (2.48, 4.31)	0.16
	Lineup × TP	*0.53 (0.08, 0.98)	1.55 (1.06, 2.29)	0.23	0.06 (-0.34, 0.48)	1.04 (0.75, 1.46)	0.21
Choosing overall (filler selections included)	Intercept	-0.40			-0.40		
	Lineup	-0.71* (-1.11, -0.30)	0.49 (0.33, 0.74)	0.21	0.00 (-0.39, 0.39)	1.00 (0.68. 1.48)	0.20
	TP	1.20* (0.89, 1.50)	2.29 (1.60, 3.25)	0.16	1.19* (0.91, 1.46)	2.29 (1.68. 3.13)	0.14
	Lineup × TP	0.44* (0.06, 0.83)	1.70 (1.08. 2.66)	0.20	0.04 (-0.29, 0.38)	1.06 (0.71, 1.62)	0.17

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

Table 5

Models Predicting Identifications (Lineup Rejection=0; Identification=1) by Lineup Comparison and Model, Experiment 4

		Lineup Comparisons								
		#Standard vs Elimination-Instructions			#Elimination vs Elimination-Instructions			#Standard vs Elimination		
Outcome	Predictor	*b* (95%CI)	*OR* (95%CI)	SE_b	*b* (95%CI)	*OR* (95%CI)	SE_b	*b* (95%CI)	*OR* (95%CI)	SE_b
Target IDs (excluding filler selections)	Intercept	-2.46			-2.91			-2.44		
	Lineup	0.01 (-0.64, 0.65)	1.01 (0.53, 1.92)	0.33	0.39 (-0.30, 1.07)	1.48 (0.74, 2.92)	0.35	-0.55 (-1.29, 0.19)	0.58 (0.28, 1.21)	0.38
	TP	2.57* (1.91, 3.22)	13.07 (6.75, 25.03)	0.34	2.72* (2.14, 3.30)	15.18 (8.50, 27.11)	0.30	2.41* (1.82, 3.00)	11.13 (6.17. 20.09)	0.30
	Lineup × TP	-0.32 (-1.19, 0.54)	0.73 (0.30, 1.72)	0.44	-0.38 (-1.16, 0.39)	0.68 (0.31, 1.48)	0.39	0.38 (-0.44, 1.20)	1.46 (0.64, 3.32)	0.42
Choosing (filler selections included)	Intercept	-1.10			-1.13			-1.08		
	Lineup	0.10 (-0.43, 0.63)	1.11 (0.65, 1.88)	0.27	0.14 (-0.41, 0.69)	1.15 (0.66, 1.99)	0.28	-0.05 (-0.57, 0.46)	0.95 (0.57, 1.58)	0.27
	TP	1.64* (1.19, 2.09)	5.16 (3.29, 8.08)	0.23	1.28* (0.90, 1.67)	3.60 (2.46, 5.31)	0.20	1.55* (1.17, 1.92)	4.71 (3.22, 6.82)	0.19

doi:10.20944/preprints201812.0012.v1

Lineup × TP	-0.47 (-1.07, 0.14)	0.63 (0.34, 1.15)	0.31	-0.12 (-0.66, 0.42)	0.89 (0.52, 1.52)	0.28	-0.27 (-0.79, 0.25)	0.76 (0.45, 1.28)	0.27

Note: * indicates significant predictors, # indicates the reference lineup condition. Number of observations, fillers included (fillers excluded):

Standard vs BMF = 2220 (1809), *n* = 78 (78); Standard vs BMFI = 2260 (1831), *n* = 79 (79); BMF vs BMFI = 2680 (2204), *n* = 67 (67)

doi:10.20944/preprints201812.0012.v1

 doi:10.20944/preprints201812.0012.v1

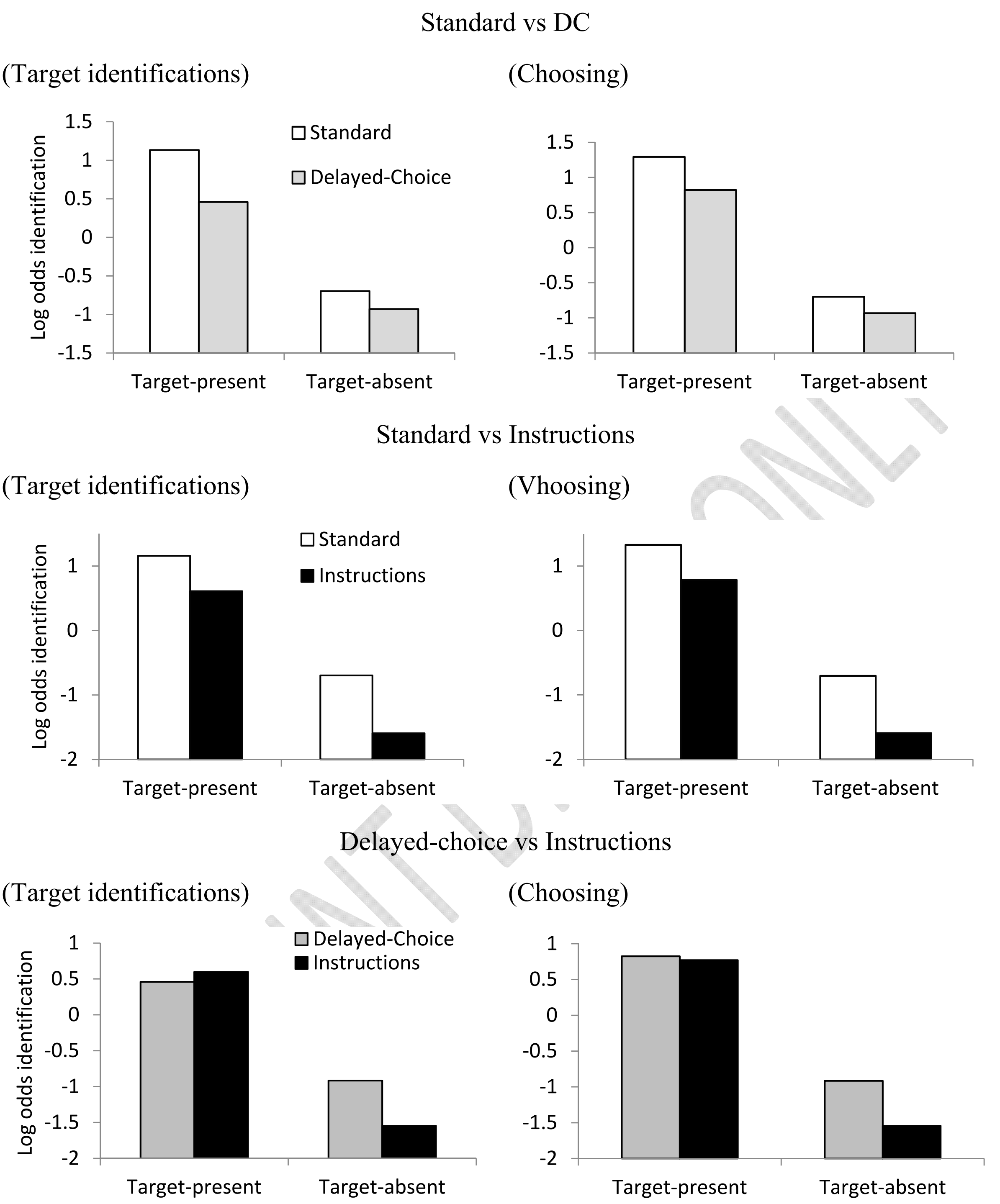
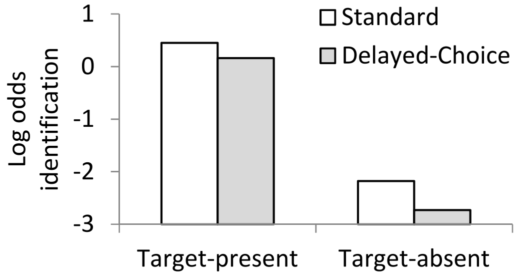


Figure 1. Estimated log odds of identification by lineup type and target presence, Experiment 1. Panels on the left show models predicting target identifications and panels on the right show models predicting choosing (including filler selections). Coefficient standard errors presented in Table 2.

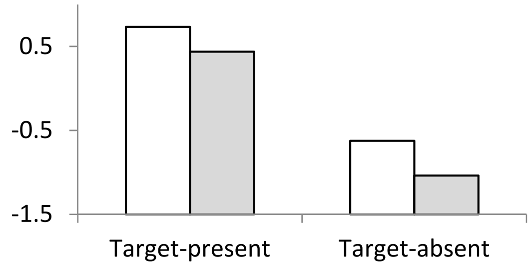
 doi:10.20944/preprints201812.0012.v1

Standard vs Delayed-Choice

(Target identifications)

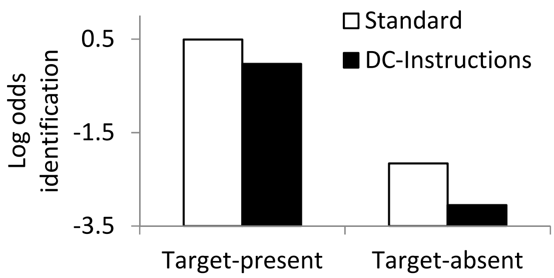


(Choosing)

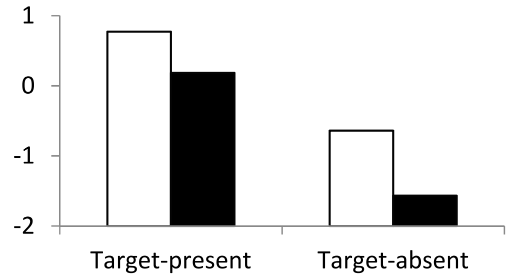


Standard vs Delayed-Choice-Instructions

(Target identifications)

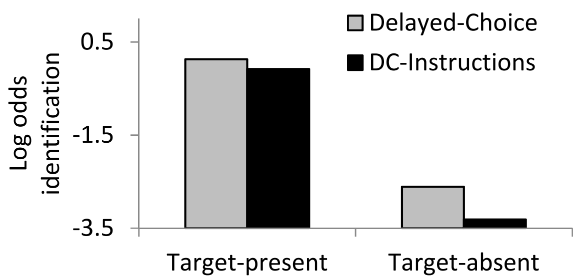


(Choosing)

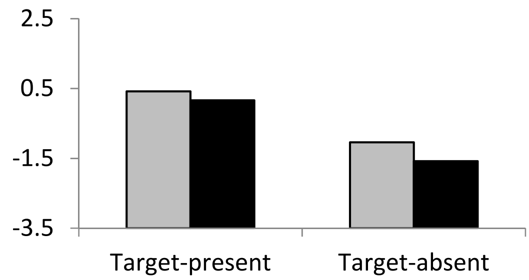


Delayed-Choice vs Delayed-Choice-Instructions

(Target identifications)

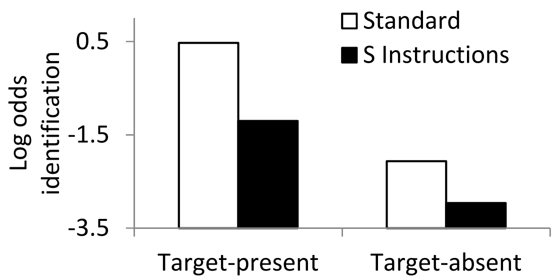


(Choosing)



Standard vs Standard-Instructions

(Target identifications)



(Choosing)

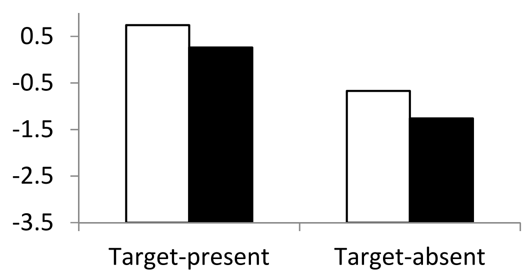


Figure 2. Estimated log odds of identification by lineup type and target presence, Experiment 2. Panels on the left show models predicting target identifications and panels on the right show models predicting choosing (including filler selections). Coefficient standard errors shown in Table 3.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2018 doi:10.20944/preprints201812.0012.v1

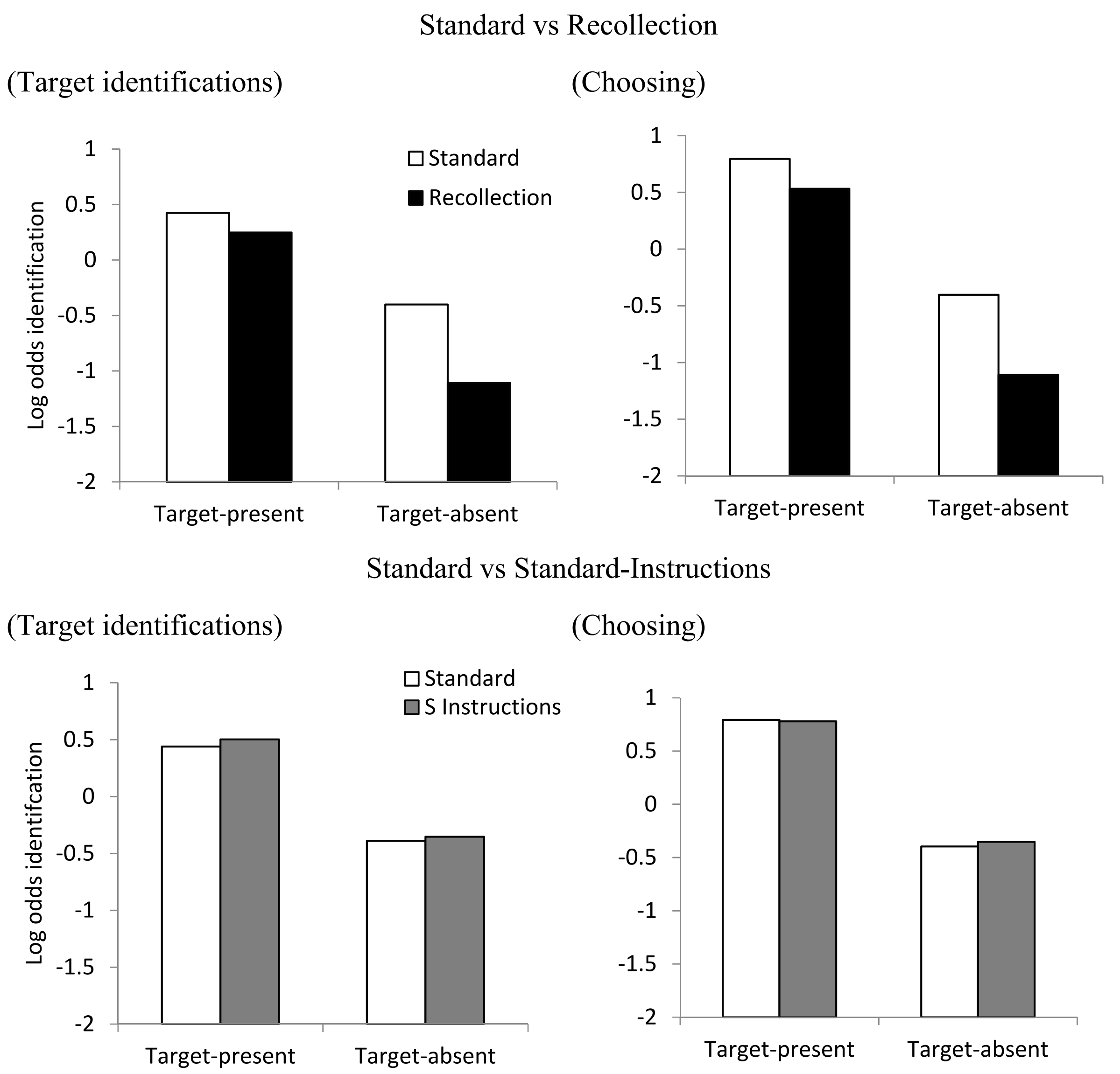


Figure 3. Estimated log odds of identification by lineup type and target presence, Experiment 3. Panels on the left show models predicting target identifications and panels on the right show models predicting choosing (including filler selections). Coefficient standard errors are presented in Table 4.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2018 doi:10.20944/preprints201812.0012.v1

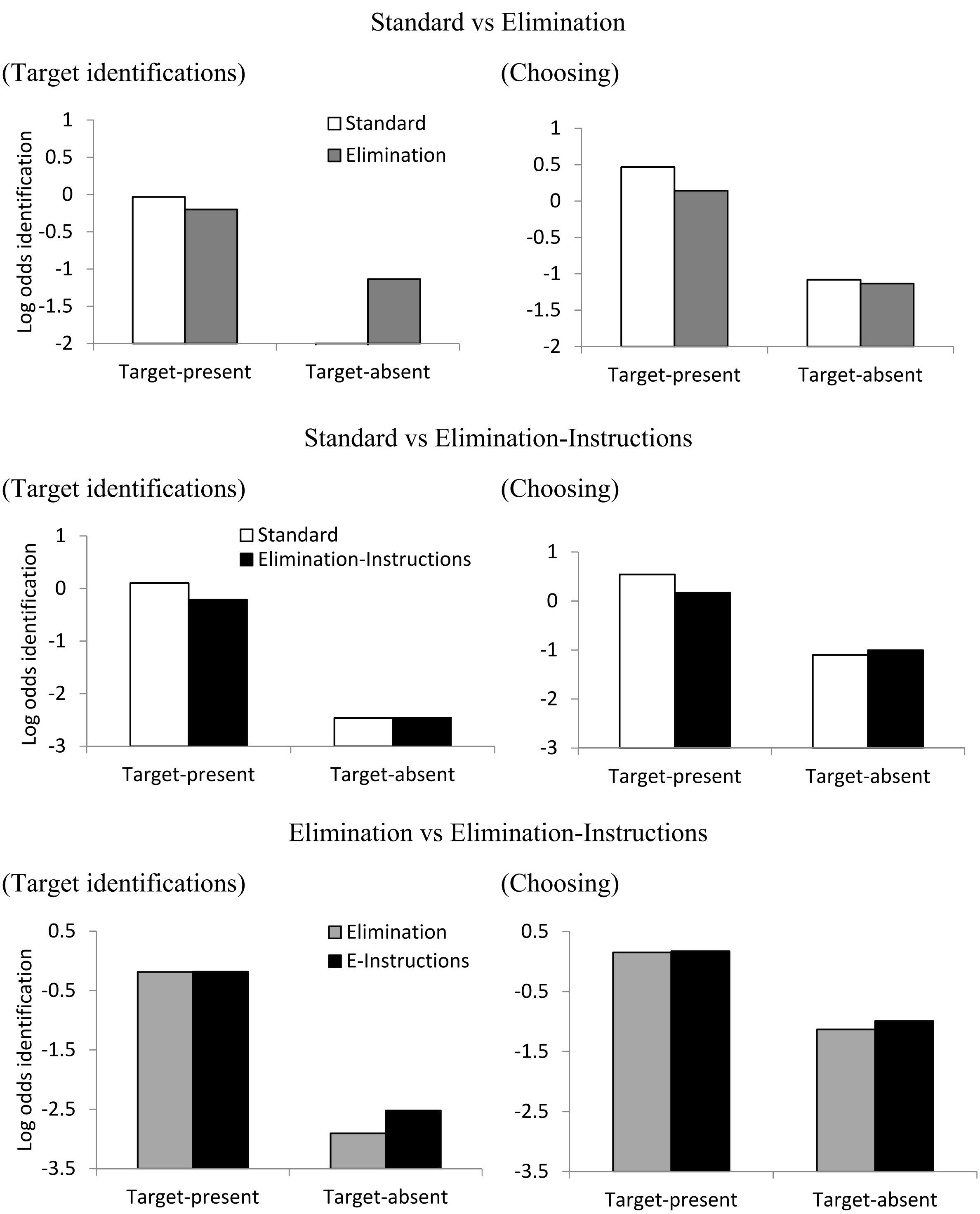


Figure 4. Estimated log odds of identification by lineup type and target presence, Experiment 4. Panels on the left show models predicting target identifications and panels on the right show models predicting choosing (including filler selections). Coefficient standard errors for estimates are displayed in Table 5.

 doi:10.20944/preprints201812.0012.v1

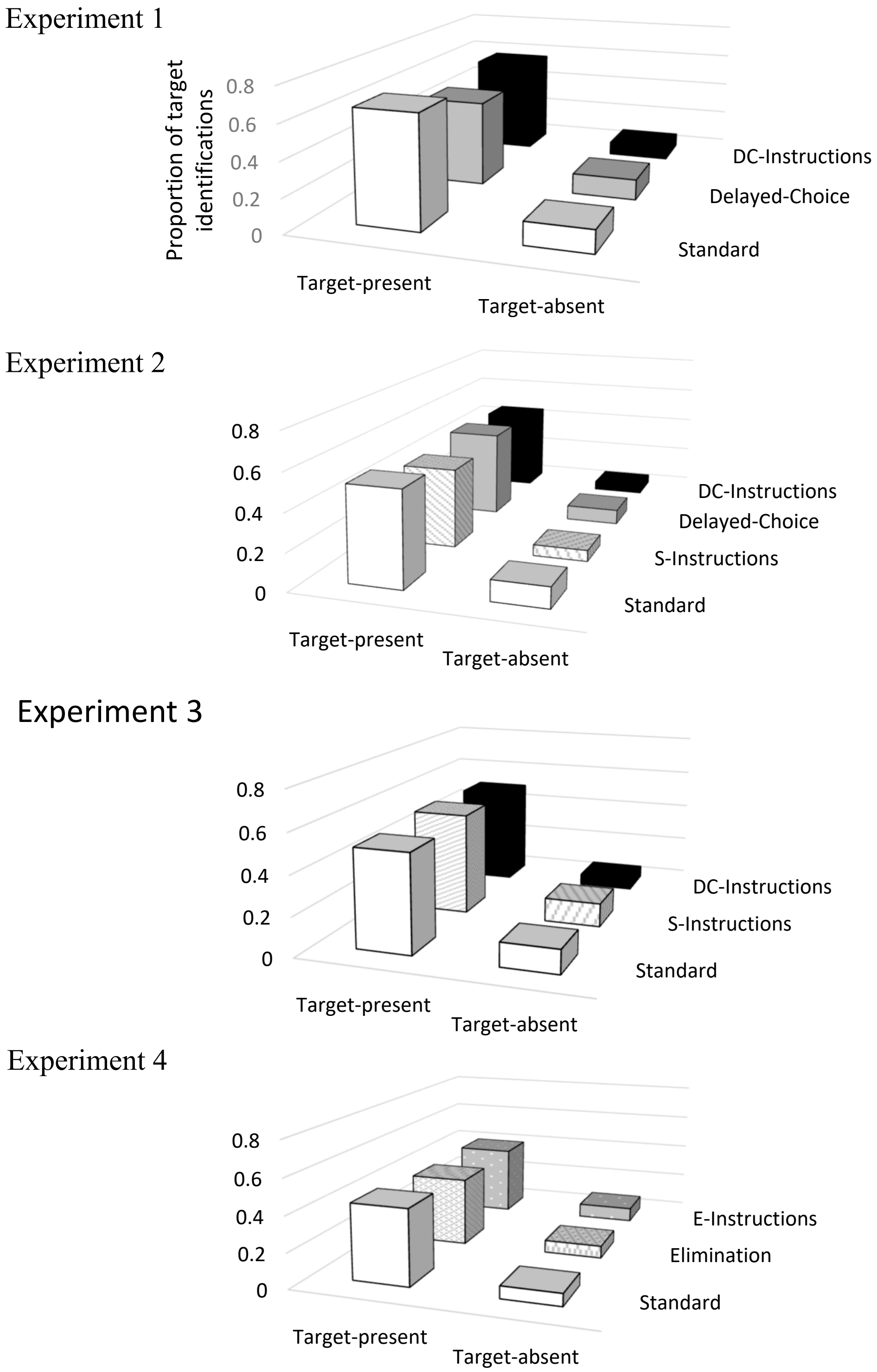


Figure 5. Proportion of target/target-replacement ('suspect') identifications by lineup type and target presence, Experiments 1-4

 doi:10.20944/preprints201812.0012.v1

Supplemental Materials

Here, we report:

1. Random effects standard deviations for mixed-effects models reported in the manuscript (see Tables 1-5). Standard deviations provide an estimate of the variability of the random effect.
2. Metacognitive instructions scripts for Experiments 1-4
3. Instructions manipulation check questionnaire, Experiments 1-4.
4. Instructions manipulation check questionnaire results.

1. Random Effects for Mixed-Effects Models Reported for Experiments 1-4

Table 6

Random Effects Standard Deviations for Mixed-Effects Model Predicting Identification Decisions (Lineup Rejection; Positive Identification) by Lineup Type in Experiment 1

	Lineup Comparison					
	#Standard vs DC		#Standard vs DCI		#DC vs DCI	
Random Effect	Filler Identifications Included	Filler Identifications Excluded	Filler Identifications Included	Filler Identifications Excluded	Filler Identifications Included	Filler Identifications Excluded
Participant	1.01	1.00	1.01	1.00	0.87	0.87
TP|Participant	1.08	1.27	1.10	1.26	1.05	1.25
Stim	0.29	0.27	0.32	0.31	0.18	0.16
TP|Stim	0.17	0.00	0.26	0.32	0.49	0.38
Lineup|Stim	0.46	0.03	0.57	0.58	0.40	0.42
Lineup*TP|Stim	0.62		0.60	0.69	0.72	0.62

doi:10.20944/preprints201812.0012.v1

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

doi:10.20944/preprints201812.0012.v1

Table 7

Random Effects Standard Deviations for Mixed-Effects Models Predicting Identification Decisions (Lineup Rejection; Positive Identification) by Lineup Type in Experiment 2

	Lineup Comparisons							
	[#]Standard vs DC		[#]Standard vs DCI		[#]Standard vs SI		[#]DC vs DCI	
Random Effect	Filler Identifications Included	Filler Identifications Excluded	Filler Identifications Included	Filler Identifications Excluded	Filler Identifications Included	Filler Identifications Excluded	Filler Identifications Included	Filler Identifications Excluded
Participant (ID)	0.78	0.82	0.93	0.86	0.96	0.98	0.65	0.55
TP|ID	0.61	0.58	0.71	0.53	0.84	0.78	0.51	0.50
Lineup|ID					0.80	0.43	0.39	0.66
Lineup*TP|ID					0.79	0.29		0.22
Stim	0.47	0.84	0.46	0.93	0.46	0.69	0.43	0.94
TP|Stim	0.52	1.02	0.55	0.93	0.51	0.60	0.65	0.98
Lineup|Stim	0.02	0.59	0.08	0.43	0.06			
Lineup*TP|Stim	0.22		0.08	0.34	0.13			

doi:10.20944/preprints201812.0012.v1

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

doi:10.20944/preprints201812.0012.v1

 doi:10.20944/preprints201812.0012.v1

Table 8

Random Effects Standard Deviations for Mixed Effects Model Predicting Identification Decisions (Lineup Rejection; Positive Identification) by Lineup Type in Experiment 3

	Lineup Comparisons			
	#Standard vs PR		#Standard vs SI	
Random Effect	Filler Identifications Included	Filler Identifications Excluded	Filler Identifications Included	Filler Identifications Excluded
Participant	0.74	0.74	0.72	0.70
TP|Participant	0.59	0.74	0.48	0.65
Stim	0.43	0.42	0.42	0.00
TP|Stim	0.77	0.92	0.74	0.68
Lineup|Stim	0.05		0.14	0.43
Lineup*TP|Stim			0.19	0.25

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

Table 9

Random Effects Standard Deviations for Mixed Effects Model Predicting Identification Decisions (Lineup Rejection; Positive Identification) by Lineup Type in Experiment 4

	Lineup Comparisons					
	#Standard vs Elimination (E)-I		#E vs EI		#Standard vs E	
Random Effect	Filler Identifications Included	Filler Identifications Excluded	Filler Identifications Included	Filler Identifications Excluded	Filler Identifications Included	Filler Identifications Excluded
ID	0.98	0.94	0.98	0.96	0.94	0.99
TP|ID	1.01	1.41	0.80	1.12	0.61	1.06
Stim	0.09	0.35	0.29	0.99	0.10	0.35
TP|Stim	0.53	0.98	0.47	0.80	0.38	1.14
Lineup|Stim	0.29	0.10	0.45	0.79	0.47	0.86
Lineup*TP|Stim	0.33	0.09	0.66	0.63	0.85	1.17

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

2. Metacognitive Instructions in Experiments 1-4

Experiment 1: Task-Related and Phenomenological Instructions

Before you begin this first series of recognition tests, we would like to give you some more information about how to do the task. Each group of photographs in this recognition test is made up of faces that match the same description as the studied face in gender, face shape, hair colour and style, eye colour, skin colour and complexion, and age. However, some faces in the group will look more similar to the studied face than others. Sometimes the studied face is present in the group of photographs and sometimes it is not.

Because some of the faces in the group are very similar to the studied face, people sometimes mistakenly decide that the studied face is present when all four photographs show faces that have not been studied. When people focus on the similarity between test faces and the studied face, they pay less attention to useful evidence that the test faces are not the studied face.

Research has shown that accurate decisions about the studied face being present are likely to be accompanied by more vivid and clear memories of studying the face than when an unstudied test face is inaccurately and mistakenly recognised. These can include memories of how the studied face looked, memories of how the [name or occupation] and the studied face are related, and other thoughts and feelings that were experienced when the face was studied.

We would like you to use this information to accurately decide when a studied face is present in the test group of photographs and when it is not. We would like you to avoid mistakenly and inaccurately recognising a test face that is similar to the studied face, but was not studied. When you think that the face you studied with the [name or occupation] is present, think about how clear, vivid, and detailed your memory is. Can you remember the thoughts and feelings you experienced when you studied the face?

If you have a clear and vivid memory of the studied face, you should indicate that the studied face is present. If you think the test photograph is similar to your memory for the studied face, but your memory is not clear, vivid or detailed, you should respond 'Not Present'.

Experiment 2: Task-Related and Phenomenological Instructions

Before you begin this first series of recognition tests, we would like to give you some more information about how memory works in recognition tasks. Each group of photographs in this recognition test is made up of faces that match the same description as the studied face in gender, face shape, hair colour and style, eye colour, skin colour and complexion, and age. Sometimes the studied face is present in the group of photographs and sometimes it is not.

Research has shown that accurate decisions about the studied face being present are likely to be accompanied by more vivid and clear memories of studying the face than when an unstudied test face is inaccurately and mistakenly recognised. These can include memories of how the studied face looked, memories of how the [name or occupation] and the studied face are related, and other thoughts and feelings that were experienced when the face was studied.

We would like you to use this information to accurately decide when a studied face is present in the test group of photographs and when it is not. When you think that the face you studied with the [name or occupation] is present, think about how clear, vivid, and detailed your memory is. Can you remember the thoughts and feelings you experienced when you studied the face?

 doi:10.20944/preprints201812.0012.v1

If you have a clear and vivid memory of the studied face, you should indicate that the studied face is present. If your memory is not clear, vivid, or detailed, you should respond 'Not Present'.

Experiments 3 & 4: Phenomenological Instructions

Before you begin this first series of recognition tests, we would like to give you some more information about how memory works in recognition tasks. Each group of photographs in this recognition test is made up of faces that match the same description as the studied face in gender, face shape, hair colour and style, eye colour, skin colour and complexion, and age. Sometimes the studied face is present in the group of photographs and sometimes it is not.

Research has shown that accurate decisions about the studied face being present are likely to be accompanied by more vivid and clear memories of studying the face than when an unstudied test face is inaccurately and mistakenly recognised. These can include memories of how the studied face looked, memories of how the [name or occupation] and the studied face are related, and other thoughts and feelings that were experienced when the face was studied.

We would like you to use this information to accurately decide when a studied face is present in the test group of photographs and when it is not. When you think that the face you studied with the [name or occupation] is present, think about how clear, vivid, and detailed your memory is. Can you remember the thoughts and feelings you experienced when you studied the face?

3. Instructions Manipulation Check Questionnaire, Experiments 1-4

Instructions

You will now be asked to complete a series of questions about how you used your memory when recognising faces. Each question will be presented on a separate screen. For some questions, you will be asked to type your responses in a text box provided at the bottom of the screen. After you have typed your answer for these questions, click the 'Next' button to proceed to the following screen. For other questions, you will be asked to respond on a rating scale with several response options. Your task will be to click on the option that best describes your response to the question. Once you have clicked on a response option, a 'Next' button will appear. Click the 'Next' button to proceed to the following screen. If you have any questions, please ask the experimenter now. Click 'Next' to proceed to the questions.

Open-ended Questions

Item 1. When one of the four photographs in the test group seemed very similar to your memory of a studied face, what do you think it was important to consider when deciding whether the studied face was present? Please type your answers into the text box presented below. You may type as much as you like. When you have finished typing your answer, click the 'Next' button to proceed to the following question.

Item 2. What is likely to be different between the experience of accurately recognising a studied face and the experience of mistakenly recognising a test photograph that is similar to the studied face? Please type your answers into the text box presented below. You may type as much as you like. When you have finished typing your answer, click the 'Next' button to proceed to the following question.

Item 3. Please describe how you think the test groups of 4 photographs were made up by the experimenters. That is, how do you think the photographs were selected for each group? Please type your answers into the text box presented below. You may type as much as you like. When you have typed every answer, click the 'Next' button to proceed to the following question

Ranking Question

Item 1. Please use the drop-down lists to select 6 features in any order that you think were used to match the 4 faces selected by the experimenters to be in each test group.

 doi:10.20944/preprints201812.0012.v1

Features. Gender; ethnicity; face shape; facial expression; similarity to studied face; hair colour; hair style; eye colour; skin colour; facial hair; nose shape; age; distinctiveness; same description.

Ratings Questions

Instructions. You will now be presented with a series of statements about how likely it is that different kinds of memories are correct. For these statements, you will be presented with 7 response options ranging from "Very unlikely" to "Very likely" to be correct. Your task will be to click on the option that best describes how correct a memory is likely to be when it has this characteristic. If you have any questions, please ask the experimenter now. When you are ready to proceed to the questions, click 'Next' to proceed to the questions.

Item 1. When my memory of a studied face is moderately vivid and the test photograph matches the general description of the studied face, it is___________to be correct.

Item 2. When my memory of a recognised face includes recollection of specific details, it is___________to be correct.

Item 3. When my memory of the studied face is not clear, but one photograph in the test group is much more similar to my memory than the other three, it is___________to be correct.

Item 4.

When I decide the studied face is present by reasoning about my memory for the face and the group of test photographs, I am___________to be correct.

Item 5. When I can remember what I thought or felt when I studied a face, my memory is___________to be correct.

Item 6. When I can remember what I thought about how the displayed name or occupation was connected to the studied face, my memory is___________to be correct.

Item 7. When I can remember how early or late in the study phase a recognised face was studied, my memory is___________to be correct.

Instructions. You will now be presented with a series of statements about how much you relied on different kinds of information when you were deciding whether you had accurately recognised a face or whether it had not been studied. For these statements, you will be presented with 7 response options for how much you relied on the kind of information mentioned in the statement, ranging from “Not at all” to “Always”. Your task will be to click on the option that best describes how much you relied on that kind of information. If you have any questions, please ask the experimenter now. When you are ready to proceed to the questions, click 'Next' to proceed to the questions.

Item 8. (Instructions conditions only) When deciding that a studied face was present, how much did you use the information about memory provided in the instructions?

Item 9. (Instructions conditions only) When deciding that a studied face was not present, how much did you use the information about memory provided in the instructions?

Item 10. When deciding that a studied face was present, how much did you rely on one face being much more similar to the studied face than the other three faces?

Item 11. When deciding that a studied face was not-present, how much did you rely on the overall vividness of your memory of the studied face?

Item 12. When making a decision about whether the studied face was present or not, how much did you rely on how clear your memory of the studied face was?

Item 13. When making a decision about whether the studied face was present or not, how much did you rely on how detailed your memory of the studied face was?

Item 14. When making a decision about whether the studied face was present or not, how much did you rely on remembering thoughts and feelings you had when the face was studied?

Item 15. When deciding that a familiar test face was not the studied face, how much did you pay attention to not having a detailed memory of the studied face?

Item 16. When deciding that a familiar test face had been studied, how much did you rely on reasoning about whether the studied face was likely to be present in that test group?

Item 17. When deciding whether a studied face was present in the test group or not, how much did you rely on finding the face in the test group that was most similar to your memory for the studied face?

Note. For questions with response scales, we used 7-point Likert scales to allow us to capture relatively fine-grained differences in responding (Cook, Heath, Thompson, & Thompson, 2001; Matell & Jacoby, 1972).

References

Cook C., Heath F., Thompson R. L., Thompson B. (2001). Score reliability in web- or Internet-based surveys: Unnumbered graphic rating scales versus Likert-type scales. *Educational and Psychological Measurement,* 61, 697-706. doi: 10.1177/00131640121971356

Matell, M. S., & Jacoby, J. (1972). Is there an optimal number of alternatives for Likert-scale items? Effects of testing time and scale properties. *Journal of Applied Psychology*, *56*, 506. doi: 10.1037/h0033601

4. Instructions Manipulation Check Questionnaire Results

Table 10

Descriptive Statistics and Effect Sizes for Responses to Instructions Manipulation Check Questionnaire Items by Lineup Type and Pairwise Comparison for Experiment 1

			Item Response					
			Lineup Type			Pairwise Comparison		
Item Type	Item	Statistic	Standard *n* = 30	DC *n* = 30	DCI *n* = 33	Statistic	Standard vs DCI	DC vs DCI
Open[#]	1	*M*	0.83	0.77	1.12	*d*	0.41*	0.49*
		SD	0.70	0.73	0.74			
		95% CI	0.58, 1.08	0.50, 1.02	0.87, 1.37	95% CI	0.23, 0.58	0.31, 0.67
	2	*M*	0.07	0.3	058	*d*	0.71*	0.39*
		SD	0.25	0.65	0.79			
		95% CI	-0.02, 0.16	0.07, 0.53	0.31, 0.85	95% CI	0.56. 0.85	0.21, 0.57
	3	*M*	0.87	0.77	1.55	*d*	0.40	0.42
		SD	1.22	1.14	1.75			
		95% CI	0.43, 1.31	0.36, 1.18	0.95, 2.15	95% CI	-0.21, 1.01	0.06, 0.79
Rating	1[^]	*M*	4.30	4.03	4.06	*d*	0.41*	0.03*
		SD	0.99	1.29	0.61			
		95% CI	3.95, 4.65	3.57, 4.49	3.84, 4.28	95% CI	0.21, 0.62	-0.20, 0.26
	2[^]	*M*	4.87	4.40	4.64	*d*	0.16*	0.17
		SD	0.68	1.35	1.48			
		95% CI	4.66, 5.14	3.92, 4.88	4.14, 5.14	95% CI	-0.34, 0.67	-0.18, 0.52

3^	*M*	3.60	3.57	3.06	*d*	0.51*	0.44*
	SD	1.07	1.22	1.09			
	95% CI	3.22, 3.98	3.13, 4.01	2.69, 3.43	95% CI	0.14, 0.89	0.15, 0.73
4^	*M*	3.57	3.27	3.64	*d*	0.06	0.29
	SD	1.25	1.34	1.25			
	95% CI	3.12, 4.02	2.79, 3.75	3.21, 4.07	95% CI	-0.37, 0.49	-0.01, 0.61
5^	*M*	4.57	4.37	4.88	*d*	0.39*	0.48*
	SD	1.55	1.27	0.86			
	95% CI	4.02, 5.12	3.92, 4.82	4.59, 5.17	95% CI	0.10, 0.68	0.21, 0.75
6^	*M*	4.63	4.50	4.85	*d*	0.21*	0.31*
	SD	1.25	1.20	1.06			
	95% CI	4.18, 5.08	4.07, 4.93	4.49, 5.21	95% CI	-0.16, 0.58	0.03, 0.60
7^	*M*	3.43	3.37	3.36	*d*	0.06*	0.00*
	SD	1.38	1.45	1.22			
	95% CI	2.94, 3.92	2.85, 3.89	2.94, 3.78	95% CI	-0.36, 0.48	-0.33, 0.34
8+	*M*	-	-	3.03		-	-
	SD	-	-	1.21			
	95% CI	-	-	2.62, 3.44		-	-
9+	*M*	-	-	3.09		-	-
	SD	-	-	1.38			
	95% CI	-	-	2.62, 3.56		-	-
10+	*M*	3.47	3.47	3.67	*d*	0.15*	0.14*
	SD	1.48	1.50	1.43			
	95% CI	2.94, 4.00	2.93, 4.01	3.18, 4.16	95% CI	-0.35, 0.64	-0.23, 0.50
11+	*M*	3.63	3.77	3.97	*d*	0.57*	0.14*

	SD	1.22	1.17	1.21			
	95% CI	3.19, 4.07	3.35, 4.19	3.56, 4.38	95% CI	0.15, 0.99	-0.22, 0.50
12^+	*M*	4.40	3.87	4.39	*d*	0.01	0.50
	SD	1.07	1.31	0.79			
	95% CI	4.02, 4.78	3.40, 4.34	4.12, 4.66	95% CI	-0.26, 0.28	0.23, 0.77
13^+	*M*	4.43	4.23	4.12	*d*	0.38*	0.10*
	SD	1.07	1.17	1.17			
	95% CI	4.05, 4.81	3.81, 4.65	3.72, 4.52	95% CI	-0.13, 0.68	-0.19, 0.39
14^+	*M*	3.50	3.40	3.91	*d*	0.26*	0.30*
	SD	1.76	1.79	1.63			
	95% CI	2.87, 4.13	2.76, 4.04	3.35, 4.47	95% CI	-0.30, 0.82	-0.12, 0.73
15^+	*M*	3.47	2.97	3.33	*d*	0.11	0.27*
	SD	1.46	1.54	1.22			
	95% CI	2.95, 3.99	2.42, 3.52	2.91, 3.75	95% CI	-0.31, 0.53	-0.08, 0.61
16^+	*M*	3.27	3.03	3.21	*d*	0.21	0.14*
	SD	1.39	1.16	1.49			
	95% CI	2.87, 3.87	2.61, 3.45	2.70, 3.72	95% CI	-0.20, 0.61	-0.20, 0.47
17^+	*M*	3.90	4.07	3.36	*d*	0.38*	0.57*
	SD	1.21	1.01	1.48			
	95% CI	3.47, 4.33	3.71, 4.43	2.86, 3.86	95% CI	-0.14, 0.89	0.25, 0.88

Note. * indicates a difference in the direction predicted if participants attended to instructions. $^{\#}$ indicates open-ended questions, mean number of instructions elements mentioned in response is reported. $^{\wedge}$ indicates questions measured on a scale from 1 (Very unlikely) to 7 (Very likely). $^{+}$ indicates questions measured on a scale from 1 (Not a lot) to 7 (Always).

Table 11

Descriptive Statistics and Effect Sizes for Responses to Instructions Manipulation Check Questionnaire Items by Lineup Type and Pairwise Comparison for Experiment 2

			Item Response						
			Lineup Type				Pairwise Comparison		
Item Type	Item	Statistic	Standard n = 32	SI n = 32	DC n = 32	DCI n = 32	Statistic	Standard vs SI	DC vs DCI
Open#	1	*M*	0.65	1.03	1.00	0.94	*d*	0.64*	-0.10
		SD	0.49	0.71	0.71	0.56			
		95% CI	0.48, 0.82	0.78, 1.28	0.76, 1.24	0.75, 1.13	95% CI	0.13, 1.13	-0.58, 0.39
	2	*M*	0.31	0.22	0.39	0.53	*d*	-0.19	0.22*
		SD	0.47	0.49	0.50	0.72			
		95% CI	0.15, 0.47	0.05, 0.39	0.22, 0.56	0.28, 0.78	95% CI	-0.68, 0.30	-0.26, 0.70
	3	*M*	2.31	2.63	2.33	3.13	*d*	0.18*	0.45*
		SD	1.69	1.74	1.81	1.70			
		95% CI	1.72, 2.90	2.02, 3.22	1.71, 2.95	2.55, 3.71	95% CI	-0.30, 0.67	-0.04, 0.93
Rating	1^	*M*	3.84	3.88	3.91	4.00	*d*	0.03	0.07
		SD	0.72	1.10	1.13	1.30			
		95% CI	3.49, 4.09	3.50, 4.26	3.52, 4.30	3.56, 4.44	95% CI	-0.46, 0.52	-0.41, 0.56
	2^	*M*	4.53	4.53	4.39	4.69	*d*	0.00	0.20
		SD	1.24	1.08	1.66	1.23			
		95% CI	4.10, 4.96	4.16, 4.90	3.82, 4.96	4.27, 5.11	95% CI	-0.49, 0.49	-0.28, 0.68
	3^	*M*	3.16	2.97	3.30	3.53	*d*	-0.20*	0.24
		SD	0.85	1.00	0.95	0.95			

	95% CI	2.87, 3.46	2.62, 3.32	3.01, 3.65	3.21, 3.85	95% CI	-0.69, 0.29	-0.25, 0.72
4^	*M*	3.19	3.31	3.33	3.47	*d*	0.11	0.14
	SD	1.28	0.93	0.95	1.02			
	95% CI	2.75, 3.63	2.99, 3.63	3.01, 3.65	3.12, 3.82	95% CI	-0.38, 0.6	-0.35, 0.62
5^	*M*	3.97	4.03	4.64	4.19	*d*	0.04*	-0.40
	SD	1.45	1.38	0.96	1.49			
	95% CI	3.47, 4.47	3.55, 4.51	4.31, 4.97	3.68, 4.70	95% CI	-0.45, 0.53	-0.84, 0.13
6^	*M*	4.16	3.94	4.21	4.44	*d*	-0.15	0.19*
	SD	1.53	1.41	1.17	1.19			
	95% CI	3.63, 4.69	3.45, 4.43	3.81, 4.61	4.03, 4.85	95% CI	-0.64, 0.34	-0.29, 0.67
7^	*M*	2.88	2.94	2.91	2.72	*d*	0.05	-0.10*
	SD	1.45	1.24	1.42	1.42			
	95% CI	2.38, 3.38	2.51, 3.37	2.43, 3.39	2.24, 3.20	95% CI	-0.44, 0.54	-0.62, 0.35
8[c]	*M*	-	3.00	-	3.47		-	-
	SD	-	1.50	-	1.19			
	95% CI	-	2.48, 3.52	-	3.06, 3.88		-	-
9[+]	*M*	-	3.00	-	3.41		-	-
	SD	-	1.52	-	1.13			
	95% CI		2.47, 3.53	-	3.02, 3.80		-	-
10[+]	*M*	3.28	3.69	3.88	4.28	*d*	0.34*	0.35*
	SD	1.22	1.18	1.27	1.02			
	95% CI	2.86, 3.70	3.28, 4.10	3.45, 4.31	3.92, 4.63	95% CI	-0.16, 0.83	-0.14, 0.83
11[+]	*M*	3.38	4.22	3.88	3.91	*d*	0.68*	0.02*
	SD	1.26	1.21	0.99	1.28			
	95% CI	2.94, 3.82	3.80, 4.63	3.54, 4.22	3.47, 4.35	95% CI	0.17, 1.18	-0.46, 0.51

12+	*M*	4.03	4.09	4.21	4.16	*d*	0.05	-0.1*
	SD	1.23	1.20	0.78	0.95			
	95% CI	3.60, 4.46	3.67, 4.51	3.94, 4.48	3.84, 4.48	95% CI	-0.44, 0.54	-0.55, 0.42
13+	*M*	3.72	3.66	3.79	4.03	*d*	-0.05*	0.19
	SD	1.20	1.38	1.41	1.15			
	95% CI	3.30, 4.14	3.18, 4.14	3.31, 4.27	3.91, 4.69	95% CI	-0.54, 0.44	-0.30, 0.67
14+	*M*	3.53	3.94	3.09	3.53	*d*	0.25*	0.23*
	SD	1.65	1.56	1.97	1.80			
	95% CI	2.96, 4.10	3.40, 4.48	2.42, 3.76	2.92, 4.14	95% CI	-0.24, 0.74	-0.25, 0.71
15+	*M*	2.94	3.00	3.18	3.59	*d*	0.04*	0.32*
	SD	1.44	1.57	1.36	1.19			
	95% CI	2.44, 3.44	2.46, 3.54	2.72, 3.64	3.18, 4.00	95% CI	-0.45, 0.53	-0.17, 0.80
16+	*M*	2.94	2.84	3.30	3.31	*d*	-0.07	0.01*
	SD	1.44	1.11	1.36	1.12			
	95% CI	2.44, 3.44	2.46, 3.22	2.84, 3.76	3.31, 4.07	95% CI	-0.56, 0.42	-0.47, 0.49
17+	*M*	4.00	3.81	4.24	3.69	*d*	-0.15*	-0.5*
	SD	1.14	1.31	1.20	1.12			
	95% CI	3.61, 4.40	3.36, 4.26	3.83, 4.65	3.31, 4.07	95% CI	-0.64, 0.34	-0.96, 0.02

Note. * indicates a difference in the direction predicted if participants attended to instructions. # indicates open-ended questions, mean number of instructions elements mentioned in response is reported. ^ indicates questions measured on a scale from 1 (Very unlikely) to 7 (Very likely). + indicates questions measured on a scale from 1 (Not a lot) to 7 (Always).

 doi:10.20944/preprints201812.0012.v1

Table 12

Descriptive Statistics and Effect Sizes for Responses to Instructions Manipulation Check Questionnaire Items by Lineup Type and Pairwise Comparison for Experiment 3

			Item Response					
			Lineup Type			Pairwise Comparison		
Item Type	Item	Statistic	Standard *n* = 32	SI *n* = 32	DCI *n* = 32	Statistic	Standard vs SI	Standard vs DCI
Open#	1	*M*	0.92	0.95	0.89	*d*	0.05*	-0.06
		SD	0.48	0.62	0.50			
		95% CI	0.72, 1.11	0.76, 1.14	0.70, 1.09	95% CI	-0.44, 0.54	-0.59, 0.47
	2	*M*	0.12	0.19	0.50	*d*	0.20*	0.89*
		SD	0.33	0.40	0.51			
		95% CI	-0.01, 0.25	0.07, 0.31	0.30, 0.70	95% CI	-0.29, 0.69	0.32, 1.44
	3	*M*	1.88	2.00	1.82	*d*	0.06*	-0.04
		SD	1.75	1.90	1.44			
		95% CI	1.17, 2.59	1.43, 2.58	1.26, 2.38	95% CI	-0.40, 0.55	-0.57, 0.49
Rating	1^	*M*	3.54	3.86	4.14	*d*	0.29*	0.58*
		SD	1.07	1.09	1.01			
		95% CI	3.11, 3.97	3.53, 4.19	3.75, 4.53	95% CI	-0.20. 0.78	0.03, 1.12
	2^	*M*	4.58	4.95	5.32	*d*	0.39*	0.82*
		SD	1.03	0.91	0.77			
		95% CI	4.15, 5.00	4.69, 5.23	5.02, 5.62	95% CI	-0.10, 0.88	0.26, 1.37
	3^	*M*	3.23	3.31	3.25	*d*	0.07*	0.02*
		SD	1.42	0.95	1.00			

	95% CI	2.66, 3.80	3.02, 3.60	2.86, 3.64	95% CI	-0.42, 0.55	-0.52, 0.55
4^	*M*	3.42	3.17	4.18	*d*	-0.25*	0.71
	SD	1.10	1.01	1.02			
	95% CI	2.98, 3.86	2.87, 3.48	3.78, 4.58	95% CI	-0.73, 0.25	0.15, 1.25
5^	*M*	4.35	4.10	4.68	*d*	-0.19	0.30*
	SD	1.29	1.28	0.90			
	95% CI	3.83, 4.87	3.71, 4.49	4.33, 5.03	95% CI	-0.68, 0.30	-0.24, 0.83
6^	*M*	4.08	4.26	4.68	*d*	0.14*	0.49*
	SD	1.41	1.31	1.02			
	95% CI	3.51, 4.65	3.86, 4.66	4.28, 5.08	95% CI	-0.35, 0.63	-0.06, 1.03
7^	*M*	2.62	2.86	3.43	*d*	0.18	0.64
	SD	1.10	1.51	1.43			
	95% CI	2.18, 3.06	2.40, 3.32	2.87, 3.99	95% CI	-0.32, 0.67	0.08, 1.17
8+	*M*	-	2.57	2.86		-	-
	SD	-	1.25	1.30			
	95% CI	-	2.19, 2.95	2.35, 3.37		-	-
9+	*M*	-	2.45	2.57		-	-
	SD	-	1.27	1.45			
	95% CI	-	2.07, 2.83	2.01, 3.13		-	-
10+	*M*	3.54	3.81	3.96	*d*	0.21*	0.35*
	SD	1.27	1.25	1.17			
	95% CI	3.03, 4.05	3.43, 4.19	3.51, 4.42	95% CI	-0.28, 0.70	-0.19, 0.88
11+	*M*	3.35	3.86	3.71	*d*	0.42*	0.26*
	SD	1.62	0.90	1.18			
	95% CI	2.70, 4.00	3.59, 4.13	3.25, 4.17	95% CI	-0.08, 0.91	-0.28, 0.79

12+	*M*	3.54	3.79	4.00	*d*	0.19	0.35
	SD	1.53	1.16	1.09			
	95% CI	2.92, 4.16	3.44, 4.14	3.58, 4.42	95% CI	-0.30, 0.68	-0.19, 0.88
13+	*M*	3.85	3.50	4.36	*d*	-0.28*	0.47
	SD	1.22	1.23	0.95			
	95% CI	3.36, 4.34	3.13, 3.87	3.99, 4.73	95% CI	-0.77, 0.21	-0.08, 1.00
14+	*M*	2.81	3.62	3.96	*d*	0.46*	0.67*
	SD	1.94	1.64	1.53			
	95% CI	2.03, 3.59	3.12, 4.12	3.36, 4.56	95% CI	-0.04, 0.95	0.11, 1.20
15+	*M*	2.73	3.10	3.11	*d*	0.25*	0.24*
	SD	1.59	1.41	1.50			
	95% CI	2.09, 3.37	2.67, 3.53	2.53, 3.69	95% CI	-0.25, 0.73	-0.30, 0.78
16+	*M*	3.00	3.17	3.36	*d*	0.13*	0.25*
	SD	1.50	1.21	1.42			
	95% CI	2.39, 3.61	2.80, 3.54	2.81, 3.91	95% CI	-0.37, 0.61	-0.29, 0.78
17+	*M*	3.88	3.95	4.18	*d*	0.05	0.25
	SD	1.42	1.25	0.94			
	95% CI	3.24, 4.38	3.57, 4.33	3.81, 4.55	95% CI	-0.44, 0.54	-0.29, 0.78

Note. * indicates a difference in the direction predicted if participants attended to instructions. # indicates open-ended questions, mean number of instructions elements mentioned in response is reported. ^ indicates questions measured on a scale from 1 (Very unlikely) to 7 (Very likely). + indicates questions measured on a scale from 1 (Not a lot) to 7 (Always).

Table 13

Descriptive Statistics and Effect Sizes for Responses to Instructions Manipulation Check Questionnaire Items by Lineup Type and Pairwise Comparison for Experiment 4

			Item Response					
			Lineup Type			Pairwise Comparison		
Item Type	Item	Statistic	Standard *n* = 45	E *n* = 33	EI *n* = 34	Statistic	Standard vs EI	E vs EI
Open#	1	*M*	1.00	1.04	1.05	*d*	0.33*	0.04*
		SD	0.00	0.20	0.22			
		95% CI	1.00, 1.00	0.97, 1.11	0.98, 1.12	95% CI	-0.12, 0.78	-0.44, 0.52
	2	*M*	1.00	1.09	1.00	*d*	0.00	-0.43
		SD	0.00	0.30	0.00			
		95% CI	1.00, 1.00	0.99, 1.19	1.00, 1.00	95% CI		-0.91, 0.06
	3	*M*	2.87	2.30	3.15	*d*	0.22*	0.65*
		SD	1.36	1.47	1.13			
		95% CI	2.47, 3.27	1.80, 2.80	2.77, 3.53	95% CI	-0.23, 0.67	0.15, 1.13
Rating^	1^	*M*	3.51	3.67	4.00	*d*	0.41	0.344
		SD	1.27	0.85	1.07			
		95% CI	3.14, 3.88	3.38, 3.96	3.64, 4.36	95% CI	-0.04, 0.86	-0.14, 0.82
	2^	*M*	4.51	4.33	4.59	*d*	0.08	0.27
		SD	1.19	1.02	0.89			
		95% CI	4.19, 4.83	3.98, 4.68	4.29, 4.89	95% CI	-0.37, 0.52	-0.22, 0.74
	3^	*M*	2.89	3.24	3.12	*d*	0.19*	-0.11
		SD	1.28	1.20	1.07			

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 December 2018 doi:10.20944/preprints201812.0012.v1

	95% CI	2.52, 3.26	2.83, 3.65	2.76, 3.48	95% CI	-0.26, 0.64	-0.59, 0.37
4^	*M*	3.16	3.30	3.41	*d*	0.24	0.10
	SD	1.04	1.13	1.13			
	95% CI	2.86, 3.46	2.91, 3.69	3.03, 3.79	95% CI	-0.21, 0.68	-0.38, 0.57
5^	*M*	3.40	4.06	3.88	*d*	0.36*	-0.14
	SD	1.42	1.32	1.25			
	95% CI	2.99, 3.82	3.61, 4.51	3.46, 4.30	95% CI	-0.1, 0.8	-0.62, 0.34
6^	*M*	3.82	4.03	3.91	*d*	0.07*	-0.10
	SD	1.34	1.21	1.22			
	95% CI	3.43, 4.21	3.62, 4.44	3.50, 4.32	95% CI	-0.38, 0.54	-0.58, 0.38
7^	*M*	2.91	3.21	3.03	*d*	0.10	-0.15*
	SD	1.18	1.17	1.19			
	95% CI	2.57, 3.24	2.81, 3.61	2.63, 3.43	95% CI	-0.35, 0.54	-0.63, 0.33
8+	*M*	-	-	3.09		-	-
	SD	-	-	1.03			
	95% CI	-	-	2.74, 3.44		-	-
9+	*M*	-	-	3.03		-	-
	SD	-	-	1.06			
	95% CI			2.67, 3.39		-	-
10+	*M*	3.89	3.48	3.59	*d*	-0.23	0.07*
	SD	1.13	1.50	1.50			
	95% CI	3.56, 4.22	2.97, 3.99	3.09, 4.09	95% CI	-0.68, 0.22	-0.41, 0.55
11+	*M*	4.16	3.85	3.38	*d*	-0.69	-0.40
	SD	1.02	1.09	1.26			
	95% CI	3.86, 4.46	3.48, 4.22	2.96, 3.80	95% CI	-1.14, -0.22	-0.87, 0.09

12+	*M*	4.40	3.97	3.85	*d*	-0.55*	-0.11*
	SD	0.99	1.19	0.99			
	95% CI	4.11, 4.69	3.56, 4.38	3.52, 4.18	95% CI	-1.00, -0.10	-0.59, 0.37
13+	*M*	4.13	3.85	4.15	*d*	0.01	0.25
	SD	1.01	1.20	1.21			
	95% CI	3.84, 4.43	3.44, 4.26	3.74, 4.56	95% CI	-0.43, 0.46	-0.24, 0.73
14+	*M*	2.71	3.39	3.26	*d*	0.29*	-0.08
	SD	1.91	1.37	1.83			
	95% CI	2.15, 3.27	2.92, 3.86	2.65, 3.88	95% CI	-0.16, 0.74	-0.56, 0.40
15+	*M*	3.56	3.30	3.24	*d*	-0.23	-0.05
	SD	1.42	1.29	1.42			
	95% CI	3.15, 3.98	2.86, 3.74	2.76, 3.72	95% CI	0.67, 0.22	-0.53, 0.43
16+	*M*	2.87	2.30	3.15	*d*	0.22*	0.65*
	SD	1.36	1.47	1.13			
	95% CI	2.47, 3.27	1.80, 2.80	2.77, 3.53	95% CI	-0.23, 0.67	0.15, 1.13
17+	*M*	3.47	2.94	3.06	*d*	-0.35*	0.09
	SD	1.22	1.46	1.07			
	95% CI	3.11, 3.83	2.44, 3.44	2.70, 3.42	95% CI	-0.80, 0.10	-0.39, 0.57

Note. * indicates a difference in the direction predicted if participants attended to instructions. # indicates open-ended questions, mean number of instructions elements mentioned in response is reported. ^ indicates questions measured on a scale from 1 (Very unlikely) to 7 (Very likely). + indicates questions measured on a scale from 1 (Not a lot) to 7 (Always).