Using metamemory measures and memory tests to estimate eyewitness free recall performance

Memory

ISSN: 0965-8211 (Print) 1464-0686 (Online) Journal homepage: https://www.tandfonline.com/loi/pmem20

Using metamemory measures and memory tests to estimate eyewitness free recall performance

Renan Benigno Saraiva, Lorraine Hope, Robert Horselenberg, James Ost, James D. Sauer & Peter J. van Koppen

MEMORY
2020, VOL. 28, NO. 1, 94–106
https://doi.org/10.1080/09658211.2019.1688835

Routledge
Taylor & Francis Group

Using metamemory measures and memory tests to estimate eyewitness free recall performance

Renan Benigno Saraiva [a,b], Lorraine Hope[a], Robert Horselenberg[b], James Ost[a], James D. Sauer[c] and Peter J. van Koppen[b,d]

[a]Department of Psychology, Faculty of Science, University of Portsmouth, Portsmouth, UK; [b]Department of Criminal Law and Criminology, Maastricht University, Maastricht, Netherlands; [c]Division of Psychology, University of Tasmania, Hobart, Australia; [d]Department of Criminal Law and Criminology, VU University Amsterdam, Amsterdam, Netherlands

ABSTRACT

Using a mock witness methodology, we investigated the predictive value of metamemory measures and objective memory tests as indicators of eyewitness free recall performance. Participants (*n* = 208) first completed a metamemory assessment that included assessments of self-rated memory capacity, memory development and use of strategies. In a separate session, participants watched a mock-crime video and provided a free recall account, followed by one out of four independent memory tests (i.e., free recall, cued recall, face recognition and general knowledge). Accuracy, amount of details reported, confidence and over/underconfidence in the eyewitness free recall were the main dependent variables. Results indicated three main findings: (1) subjective assessments of memory capacity were not related to eyewitness free recall performance; (2) although individual confidence and over/underconfidence was somewhat stable across different memory tests, accuracy was less stable; and (3) individuals with higher self-rated memory capacity had a slightly stronger confidence-accuracy relation in free recall. These results are discussed with respect to metamemory assessments and performance stability across memory tests of different domains.

ARTICLE HISTORY
Received 10 May 2019
Accepted 23 October 2019

KEYWORDS
Eyewitness testimony; metamemory; free recall; confidence; memory

Eyewitness statements are often critical in criminal investigations and maybe the only available source of information about a crime when physical evidence is absent. Correct information provided by eyewitnesses can help investigators identify and trace suspects, but incorrect information can impede the investigative process, wasting valuable time and resources. Therefore, the ability to distinguish between accurate and inaccurate information is critical when evaluating witness's statements. Psychological research has identified some factors that can help discriminate eyewitness identification accuracy, such as confidence (Sauer, Brewer, Zweck, & Weber, 2010; Sporer, Penrod, Read, & Cutler, 1995; Wixted & Wells, 2017), decision time (Weber, Brewer, Wells, Semmler, & Keast, 2004), and decision processes (Sauerland & Sporer, 2007; Weber et al., 2004). However, only a small number of studies have investigated factors that may assist with the estimation of accuracy in free reports (Dahl, Allwood, Scimone, & Rennemark, 2015; Odinot, Wolters, & van Giezen, 2013; Sauer & Hope, 2016; Weber & Brewer, 2008). In the current research, we aimed to further investigate the relationship between accuracy and confidence in eyewitness free recall. We tested whether metamemory instruments and distinct memory tests can be used to estimate eyewitness accuracy, amount of details reported, confidence and over/underconfidence in a free recall task pertaining to the witnessed event.

The eyewitness confidence-accuracy relationship

Confidence assessments may be present in many instances of criminal investigations, occurring whenever a police officer, lawyer or another law practitioner asks a witness if they are sure about a given account or identification. Lay people and practitioners in the criminal justice system often regard eyewitness confidence as a strong indicator of eyewitness credibility (Brewer & Burke, 2002; Penrod & Cutler, 1995). Many studies have shown that confidence can be a valid indicator of eyewitness identification performance, especially when confidence statements are collected soon after identification decisions (Brewer & Wells, 2006; Brewer, Keast, & Rishworth, 2002; Sauer et al., 2010; Wixted & Wells, 2017; Wixted, Mickes, Clark, Gronlund, & Roediger, 2015). However, eyewitnesses may exaggerate their confidence in an identification or their confidence may be inflated, particularly if encoding conditions are poor or if biased lineup procedures are used, making confidence statements in such circumstances less reliable estimates of accuracy (Douglass & Jones, 2013;

CONTACT Renan Benigno Saraiva renan.saraiva@port.ac.uk
Supplemental data for this article can be accessed at https://doi.org/10.1080/09658211.2019.1688835

Leippe, Eisenstadt, & Rauch, 2009; Lindsay, Read, & Sharma, 1998; Penrod & Cutler, 1995).

In eyewitness identification research, confidence-accuracy relationship is typically examined by adopting calibration procedures (e.g., Brewer & Wells, 2006; Palmer, Brewer, Weber, & Nagesh, 2013). In this procedure, eyewitness accuracy is often plotted against specific confidence categories (e.g., proportion of accurate responses given with 10% confidence). Perfect calibration is represented by a linear function in which eyewitness accuracy is equivalent to each confidence category (e.g., 10% accuracy for 10% confidence responses, 20% accuracy for 20% confidence responses, and so on). Calibration analysis allows for different ways of assessing the eyewitness confidence-accuracy relationship, including three commonly adopted statistics: Calibration (C); Over/underconfidence (O/U) and Adjusted Normalized Resolution Index (ANRI). Calibration (C) represents how far a given calibration curve is from a perfect calibration. It ranges from 0 (perfect calibration) to 1, and lower values indicate better calibration. Over/underconfidence (O/U) indicates if a curve strays more above or below the perfect calibration line, with values ranging from −1 (very underconfident) to 1 (very overconfident). The Adjusted Normalized Resolution Index (ANRI) represents how well confidence discriminates accurate from inaccurate responses, with higher values indicating better discrimination. C, O/U and ANRI are informative and complementary statistics. That is because C provides an estimate on how far a given calibration curve strays from a perfect calibration, but it does not indicate whether weak confidence-accuracy relationships are due to overconfidence or underconfidence (which is provided by O/U). ANRI is a resolution statistic that provides an estimate on how well confidence can be used to distinguish accurate from inaccurate responses.

Most research on eyewitness confidence-accuracy (CA) relationship has focused on suspect identifications, but witnesses' freely recalled memories are also highly relevant in forensic contexts, since most of the information provided by eyewitnesses comprise descriptions of the perpetrator and the event (Van Koppen & Lochun, 1997). Some studies suggest that free recall paradigms generate stronger CA relations when compared to forced-response or recognition memory paradigms (Robinson & Johnson, 1996; Robinson, Johnson, & Herndon, 1997; Robinson, Johnson, & Robertson, 2000). Studies focusing on free recall of staged crimes, for example, have found CA correlations of around .60 (Odinot et al., 2013; Odinot & Wolters, 2006; Robinson & Johnson, 1996). In comparison, a meta-analysis on the eyewitness confidence-accuracy relationship for identification tasks has found a weighted effect size of $r = .37$ for choosers and $r = .12$ for nonchoosers (Sporer et al., 1995). The apparently stronger CA relation observed in free recall tasks may be accounted for by its "free" component, because in such procedures witnesses tend to choose which information to report based on their confidence, consequently increasing metamemory realism (Allwood, Jonsson, & Granhag, 2005). In fact, theoretical frameworks of memory reporting suggest that people balance the demands for informativeness and accuracy during cued-recall, withholding details that fall below a pre-set criterion of probable accuracy (Koriat & Goldsmith, 1996; Robinson et al., 2000; Robinson & Johnson, 1996). However, important aspects of the CA relation in eyewitness free recall remain to be examined, particularly those related to individual differences in self-perceived memory capacity and memory functioning.

Metamemory and the CA relationship

Metamemory refers to an individual's knowledge and awareness of his or her own memory capabilities, based on previous experiences and beliefs (Dunlosky & Bjork, 2008). This introspective knowledge is used to monitor and control encoding, retrieval, reporting processes, and to provide information about memory confidence (Koriat & Goldsmith, 1996; Nelson & Narens, 1994). During confidence assessments individuals may rely mostly on memory trace strength, typically providing higher confidence ratings to stronger memory traces (Koriat, Goldsmith, & Pansky, 2000). Notwithstanding, some authors argue that memory confidence may not only be related to memory strength, but also to intrinsic, heuristic and self-credibility cues (Jonsson & Allwood, 2003; Leippe & Eisenstadt, 2014). Intrinsic cues have been found to influence other types of metamemory judgements such as judgements of learning (Koriat, Bjork, Sheffer, & Bar, 2004) and feeling of knowing (Koriat, 2000). In legal settings, eyewitnesses might use the difficulty of the task as a cue to report confidence, showing less confidence for correct answers to hard questions when compared to correct answers to easy questions (Howie & Roebers, 2007; Stankov, 2000).

Heuristic cues comprise one's beliefs about external factors that can help or impair memory encoding and retrieval, for example when eyewitnesses put more effort on recall tasks, even if it does not lead to changes in accuracy (Shaw & Zerr, 2003). These self-credibility cues can also be derived from people's beliefs about their overall memory performance. Some could overestimate their ability to recall events and show overconfidence, others may underestimate their memory ability and show underconfidence (Leippe, Eisenstadt, Rauch, & Stambush, 2006; Olsson & Juslin, 1999). Traits of self-confidence have also been linked to performance accuracy and other metacognitive factors (Kleitman & Stankov, 2007), although memory self-credibility cues have yet to be explored in relation to eyewitness free recall memory.

Memory self-credibility cues may have important implications not only on how individuals report confidence during free recall accounts, but also on the quantity and quality of information disclosed. Evans and Fisher (2011) argue that eyewitness free recall reporting is, in its essence, a metacognitive control process in which individuals strategically withhold uncertain responses, or choose to provide accurate responses that are likely to be imprecise (Koriat & Goldsmith, 1996). However, it is still unclear whether self-assessments of memory capacity play a role when witnesses are considering what information to disclose in terms of accuracy, quantity and precision. Koriat, Nussinson, Bless, and Shaked (2008) propose that memory monitoring processes may be partly based on the learning experience (i.e., the experience of processing and remembering the learning material) or on domain-specific beliefs (e.g., "I do not have a very good memory"). Unlike situational learning experiences, domain-specific beliefs are enduring and have been constructed on the basis of previous experiences and implicit theories and schemas about memory (e.g., perceived effects of ageing on memory, perceived memory stability across domains; Hertzog & Dixon, 1994; Hertzog, McGuire, & Lineweaver, 1998). If eyewitness confidence statements and memory traces are partially based on domain-specific beliefs, it might be expected that self-ratings of memory capacity are predictive of performance in eyewitness free recall tasks.

Eyewitness recall and performance on objective memory tests

Subjective self-ratings of memory capacity may be useful predictors of memory performance, but performance in objective memory tests might be more informative with respect to eyewitness accuracy, confidence and over/underconfidence. Some studies, for example, show that accuracy in different face memory tests are predictive of eyewitness identification accuracy (Bindemann, Brown, Koyas, & Russ, 2012; Morgan et al., 2007). However, to our knowledge, no research has examined the relation between eyewitness free recall performance and performance in unrelated memory tests, such as free-recall and cued-recall for a non-criminal event, or face recognition and general knowledge tests. On one hand, some evidence shows that different memory systems are rather independent (Melby-Lervåg & Hulme, 2013; Simons et al., 2016), suggesting that performance in one memory test may not be predictive of performance in other tests. On the other hand, a few studies indicate some stability in performance for different memory tests due to individual differences (Jaschinski & Wentura, 2002; Zhu et al., 2010). It may be expected that tasks more closely related to an eyewitness free recall paradigm (e.g., unrelated free recall or cued recall tests) are more predictive of eyewitness free recall performance than more distantly related tasks (e.g., face recognition and general knowledge). In the present experiment, we investigated the predictive utility of different objective memory tests (i.e., free recall, cued recall, general knowledge and face recognition) for eyewitness free recall performance, in order to examine stability in accuracy, confidence, and over/underconfidence across different memory domains. A recognition test for unrelated faces was included as a test more distantly related to the eyewitness free recall, in order to better examine the extent to which memory performance stability may change depending on similarities between memory domains.

Current research

The idea that confidence measures are affected not only by the availability of memory traces, but also by different intrinsic, heuristic and self-credibility cues is critical to the understanding and practical utility of confidence in forensic contexts. Previous research has shown that self-assessments of memory capacity can be used to estimate the diagnostic value of eyewitnesses' confidence in identification tasks (Olsson & Juslin, 1999; Searcy, Bartlett, & Memon, 2000). In this study, we tested the use of metamemory self-assessment instruments, and objective memory tests as predictors of eyewitness accuracy, completeness, confidence and over/underconfidence in a free recall task. We hypothesised that self-reported metamemory measures would be predictive of eyewitness accuracy (H1), amount of details reported (H2), confidence (H3); and over/underconfidence (H4). These predictions are based on theoretical frameworks suggesting that memory monitoring processes are influenced not only by the experiences of the quality of memory processes at the time the metacognitive judgment is made, but also by domain-specific beliefs (e.g., metamemory traits; Evans & Fisher, 2011; Koriat et al., 2008). Moreover, these predictions are based on some initial evidence that metamemory self-assessments are related to identification confidence, although mixed results can be found regarding the relation between metamemory and identification accuracy (Olsson & Juslin, 1999; Searcy et al., 2000).

Estimating eyewitness confidence using metamemory components relates to an important theoretical question, concerning the role of intrinsic cues on reports of confidence during free recall tasks. It was expected that individuals with higher metamemory scores (e.g., higher contentment with one's own memory) would have a stronger confidence-accuracy relation than individuals with lower metamemory scores (H5). Finally, it was hypothesised that performance in different objective memory tests would be predictive of eyewitness free recall performance (H6), and tests more closely related to the free recall test would be more predictive than tasks distantly related to this paradigm (H7).

Method

Participants and design

A sample of 208 participants was recruited from the local and student community. Most participants were recruited from the student community (77%) in relation to other local participants (23%). Participants were aged between

18 and 70 years old, M age = 23.25, SD = 9.33 and 81% were female. The required sample size was estimated using power analysis conducted for a Multivariate Regression Analysis with $f = 0.05$, alpha = .05 and power = 0.95. The projected sample size needed for this effect size was approximately $N = 205$. Participants either received course credits or a £5 compensation for their time.

All participants completed a metamemory assessment and a free recall test of a mock-crime video. In a between-subjects design, participants were assigned to one of four possible objective memory test conditions: free recall, cued recall, general knowledge and face recognition. The dependent variables were accuracy, amount of details reported, confidence and over/underconfidence in the eyewitness free recall task.

Materials

Multifactorial Metamemory Questionnaire (MMQ; Troyer & Rich, 2002*).* The MMQ is an instrument with 57 items comprising three distinct factors: Contentment (i.e., affect related to memory abilities, $\alpha = 0.91$), Ability (i.e., frequency of memory problems in different situations, $\alpha = 0.89$), and Strategy (i.e., use of memory strategies in everyday life, $\alpha = 0.84$). The factor Contentment consists of 18 items (e.g., "my memory is worse than most other people my age") rated on a scale from 1 (*strongly agree*) to 5 (*strongly disagree*), with higher scores indicating higher memory contentment. The factor Ability has 20 items which require respondents to indicate how often they experienced memory mistakes over the last two weeks (e.g., "how often do you forget an appointment?"), on a scale ranging from 1 (*all the time*) to 5 (*never*). Higher scores in Ability indicate fewer (self-reported) memory problems in daily situations. The factor Strategy has 19 items related to the use of different memory strategies (e.g., "how often do you create a story to link together information you want to remember?") and respondents indicate the frequency with which each strategy was used over the last 2 weeks using a scale that ranges from 1 (*never*) to 5 (*all the time*). Higher scores in Strategy indicate a more frequent use of memory strategies.

Squire Subjective Memory Questionnaire (SSMQ; Squire, Wetzel, & Slater, 1979*;* Van Bergen, Brands, & Jelicic, 2010*).* The SSMQ is an instrument that assesses subjective memory functioning and consists of one single factor reflecting people's beliefs about their own memory functioning ($\alpha = 0.93$). The instrument includes 18 items (e.g., "My ability to reach back in my memory and recall what happened a few minutes ago is"), rated on nine-point scales that range from −4 (*worse than ever before*) to 4 (*better than ever before*). Higher scores in SSMQ indicate a higher self-perceived memory functioning.

*Facial recognition and general memory skill assessment (*Olsson & Juslin, 1999*).* Developed by Olsson and Juslin (1999), this instrument includes two items that assess self-reported general memory skill (e.g., "give an estimate of your general memory ability, compared to other people's general memory ability") and two items that assess self-reported facial recognition skill (e.g., "give an estimate of your ability to remember faces as compared to other people's ability to remember faces"). Participants indicate their ability in comparison to the normal population on a 11-point scale that ranges from −5 (*much worse*) to 5 (*much better*).

Mock-crime stimulus. The stimulus event for the eyewitness free recall task was a short film (2:30min) depicting a theft. In the film, two perpetrators (a man and a woman) follow a young man into his house. One of the perpetrators pretends to be lost and asks the victim for directions, and while the victim is distracted the other perpetrator steals his laptop, phone and keys.

Objective memory tests

Free-recall test. In the free-recall condition, participants completed a second free-recall test for a stimulus that was not associated with the eyewitness paradigm. In this condition, participants watched a short film depicting a cleaning routine in a house (2.5min), then completed an unrelated filler task (5min) before completing a free-recall test concerning this cleaning stimulus film.

Cued-recall test. In the cued-recall condition, participants watched the same film of the free-recall test condition depicting a cleaning routine in a house and then completed an unrelated filler task (5min). Next participants completed 17 cued-recall questions about the video (e.g., what did the woman do in the TV room?). Each question was followed by a confidence scale ranging from 0% ("not at all certain") to 100% ("totally certain").

Face recognition test. Forty-five adult male faces with no unusual identifying features were selected from a database of faces (Thomaz & Giraldi, 2010). Faces were standardised in size, resolution, and background colour. During the training phase, 30 faces were displayed to participants in random order on a computer screen. Each face was presented for three seconds, with a three-second inter-stimulus interval. After completing a 5min filler task participants took part in the testing phase, in which a second set of 30 faces was presented, including 15 faces from the training phase and 15 new faces. Each face was shown individually. Participants were instructed to indicate whether or not the face had been seen before and give a confidence judgement using a scale that goes from 0% ("not at all certain") to 100% ("totally certain"). Participants had unlimited time to make their decision and proceed to the next face.

General knowledge memory test. A pool of 38 general knowledge questions (e.g., "What country is known as the Land of Rising Sun?") was generated and pilot tested. Eight questions were considered too easy or too difficult and were removed from the final pool. Participants in the General Knowledge condition were asked to answer each one of the 30 questions and rate their confidence in a scale that ranges from 0% ("not at all certain") to 100%

("totally certain"). All questions in this were open-ended and no answer alternatives were provided.

Procedure

First, using the online platform Qualtrics, participants completed the full set of metamemory measures (MMQ, SSMQ, and facial recognition and general memory skill assessment, Olsson & Juslin, 1999). Twenty-four hours after completing the metamemory measures, participants took part in a lab session. Not all participants who completed the online session ($N = 287$) signed up to the lab session ($N = 208$), so participant recruitment continued until we achieved our stipulated sample size (minimum of $N = 205$). In the lab session, participants first watched the mock-crime stimulus film and then completed an unrelated filler task (5 min). Immediately after participants received the free recall test about the mock-crime. The instructions were as follows:

> In the space provided, report all details that you can remember about the video, including the sequence of actions and events, and the people that were involved. If you recall information or specific details out of the order in which they happened, report these details as they come to mind (i.e., do not leave out any details.) Do not guess about details that you cannot remember. Feel free to use full sentences or bullet points – but please make sure your report is as complete and accurate as possible.

When participants finished the task, the researcher read through the report and marked each detail for which participants should now report a confidence judgement (e.g., A caucasian (1) man (1) in a burgundy (1) hoody (1) stole (1) the bike (1) = 6 details). Participants were then asked to read through the statement and indicate their degree of confidence for each detail, using a scale that ranged from 0% ("not at all certain") and then 10, 20, 30, … to 100% ("totally certain"). After completing their confidence assessment, participants took part in the objective memory test corresponding to their randomly allocated experimental condition (i.e., free recall, cued recall, general knowledge and face recognition). Ten per cent of all eyewitness recall transcripts were coded independently by two raters and Intraclass Correlation Coefficients were calculated for number of details = .97 (95% CI = .93, .99), and correctness = .96 (95% CI = .92, .98)

Results

Our analysis focused on four dependent measures in the eyewitness free recall task: accuracy, amount of details reported, confidence and over/underconfidence. Accuracy was defined as the proportion of accurate responses reported, amount of details as the total amount of accurate details, and confidence as the mean of all confidence statements provided for each detail. In calibration research, over/underconfidence can be computed as a statistic that relates to how well-calibrated participants are in their confidence-accuracy relationship, ranging from −1 (very underconfident) to 1 (very overconfident; see Brewer & Wells, 2006 for calculation details). Similar scores were computed for all objective memory tests (i.e., free recall, cued recall, general knowledge and face recognition). Scores on the objective memory tests served as predictors of eyewitness free recall performance, instead of being treated as dependent variables (see Table 1).

Metamemory scale scores and eyewitness free recall performance

First, we fitted a multivariate regression model including the scale scores on the metamemory assessments as predictors of free recall accuracy, amount of details reported, confidence and over/underconfidence. One outlier in the amount of details variable was highly influential in the test parameters (Cook's Distance = 0.13) and was capped to the upper limit to avoid biased results. QQ-plots revealed that the errors distribution in the accuracy and confidence models were not normally distributed. Therefore, we fitted these models using gamma distributions in order to obtain more robust estimates (Manning, Basu, & Mullahy, 2005). None of the metamemory scales were predictive of accuracy, amount of details reported, confidence or over/underconfidence in the eyewitness free recall test (smallest p value = .17). Results from four different multivariate tests (Pillai, Wilks, Hotelling-Lawley, and Roy) suggest that the coefficients for MMQ-Contentment, MMQ-Strategy, MMQ-Ability, SSMQ, and the facial recognition and general memory skill assessment by Olsson and Juslin (1999) do not seem to be statistically different from 0 for any of the dependant measures (see Table S1 in the supplemental materials). Bayesian multiple regression models were fitted to further examine evidence for the null model, using a standard Jeffreys–Zellner–Siow prior (see Table 2). Results revealed that marginal inclusion probabilities for all coefficients in the accuracy, amount of details reported, confidence and over/underconfidence models were negligible (highest inclusion probability = .07; Kruschke, 2015). Therefore, subjective assessments

Table 1. Means and standard deviations of the main memory performance variables.

	Proportion Accuracy *M* (*SD*)	Amount of details *M* (*SD*)	Confidence *M* (*SD*)	Over/underconfidence *M* (*SD*)
Eyewitness free recall ($n = 208$)	0.97 (0.04)	40.7 (11.4)	0.92 (0.58)	−0.04 (0.06)
Free recall ($n = 55$)	0.95 (0.05)	30.2 (13.2)	0.93 (0.52)	−0.02 (0.06)
Cued recall ($n = 50$)	0.69 (0.11)	–	0.64 (0.14)	−0.05 (0.13)
Face recognition ($n = 50$)	0.45 (0.08)	–	0.43 (0.92)	−0.01 (0.15)
General Knowledge ($n = 55$)	0.41 (0.15)	–	0.39 (0.18)	−0.02 (0.14)

Table 2. Bayesian Multiple linear regression models including metamemory scales as predictors of eyewitness free recall performance.

Coefficient	*M*	*SD*	$BF_{inclusion}$	95% Credible Interval	
				Lower	Upper
Free recall accuracy					
MMQ-Contentment	−1.033e −4	9.148e −4	0.049	0.000	0.000
MMQ-Ability	7.298e −5	9.946e −4	0.045	0.000	0.000
MMQ-Strategies	−1.113e −4	9.914e −4	0.049	0.000	0.000
SSMQ	−5.650e −7	4.521e −4	0.042	0.000	0.000
General Memory Ability	4.057e −6	3.028e −4	0.042	0.000	0.000
Face Memory Ability	−1.961e −5	2.760e −4	0.044	0.000	0.000
Free recall accuracy					
MMQ-Contentment	−0.018	0.273	0.047	0.000	0.000
MMQ-Ability	−0.043	0.383	0.052	−0.110	0.000
MMQ-Strategies	0.071	0.423	0.065	0.000	0.611
SSMQ	0.016	0.177	0.049	0.000	0.000
General Memory Ability	−0.002	0.105	0.045	0.000	0.000
Face Memory Ability	0.012	0.105	0.052	−0.085	0.000
Free recall confidence					
MMQ-Contentment	−0.003	0.033	0.186	−0.100	0.068
MMQ-Ability	0.055	0.088	0.596	0.000	0.261
MMQ-Strategies	−0.007	0.035	0.193	−0.135	0.021
SSMQ	0.012	0.032	0.302	−0.011	0.090
General Memory Ability	0.012	0.026	0.418	−0.007	0.090
Face Memory Ability	0.018	0.027	0.678	−0.002	0.077
Free recall over/underconfidence					
MMQ-Contentment	2.426e −4	0.002	0.064	−8.178e −4	3.309e −4
MMQ-Ability	−0.002	0.005	0.132	−0.017	0.000
MMQ-Strategies	1.985e −4	0.002	0.061	−8.374e −5	0.002
SSMQ	1.333e −4	0.001	0.059	−3.329e −4	5.241e −4
General Memory Ability	5.799e −5	8.374e −4	0.057	−1.340e −4	5.694e −4
Face Memory Ability	−4.606e −5	6.504e −4	0.056	−2.721e −4	2.763e −4

of memory capacity were not related to eyewitness free recall performance.

Objective memory tests and eyewitness free recall performance

We used simple regressions to test the relationship between eyewitness free recall performance and performance on the four different memory tests (see Table 3). Eyewitness free recall accuracy was predicted by cued recall accuracy ($\beta = 0.38$), but not by accuracy in the other memory tests. Confidence in the eyewitness free recall was predicted by confidence expressed in the free recall ($\beta = 0.75$), cued recall ($\beta = 0.54$) and general knowledge memory tests ($\beta = 0.30$). A similar pattern of results was observed for over/underconfidence in the free recall test, which was predicted by over/underconfidence in the free recall ($\beta = 0.63$), cued recall ($\beta = 0.56$) and general knowledge ($\beta = 0.49$) tests. Performance in the face recognition test did not relate to any of the performance variables in the eyewitness free recall. Therefore, although individual confidence and over/underconfidence was somewhat stable across different memory tests, accuracy was less stable.

Metamemory and confidence-accuracy calibration analysis

Calibration analyses were conducted to test the relation between the metamemory scale scores and eyewitness

Table 3. Regression models using performance in the different memory tests as predictors of eyewitness free recall performance.

	B (*SE*)	*B* CI	*β*	*t*	*p*
Eyewitness recall amount of details					
Free recall amount of details	0.57 (0.07)	[0.41, 0.72]	0.71	7.44	<.001
Eyewitness recall accuracy					
Free recall accuracy	0.12 (0.10)	[−0.08, 0.34]	0.16	1.23	.22
Cued recall accuracy	0.14 (0.05)	[0.04, 0.25]	0.38	2.92	<.01
General Knowledge accuracy	−0.04 (0.02)	[−0.07, 0.01]	−0.21	−1.59	.12
Face recognition accuracy	−0.04 (0.04)	[−.12, 0.05]	−0.12	−0.84	.40
Eyewitness recall confidence					
Free recall confidence	0.83 (0.09)	[0.63,1.03]	0.75	8.34	<.001
Cued recall confidence	0.25 (0.05)	[0.13, 0.36]	0.54	4.49	<.001
General knowledge confidence	0.09 (0.04)	[0.01, 0.17]	0.30	2.27	.02
Face recognition confidence	−0.05 (0.08)	[−0.23, 0.12]	−0.08	−0.60	.54
Eyewitness recall OU					
Free recall OU	0.70 (0.11)	[0.46, 0.94]	0.63	5.84	<.001
Cued recall OU	0.29 (0.06)	[0.17, 0.42]	0.56	4.71	<.001
General knowledge OU	0.24 (0.06)	[0.12, 0.36]	0.49	4.03	<.001
Face recognition OU	0.08 (0.06)	[−0.04, 0.21]	0.17	1.26	.21

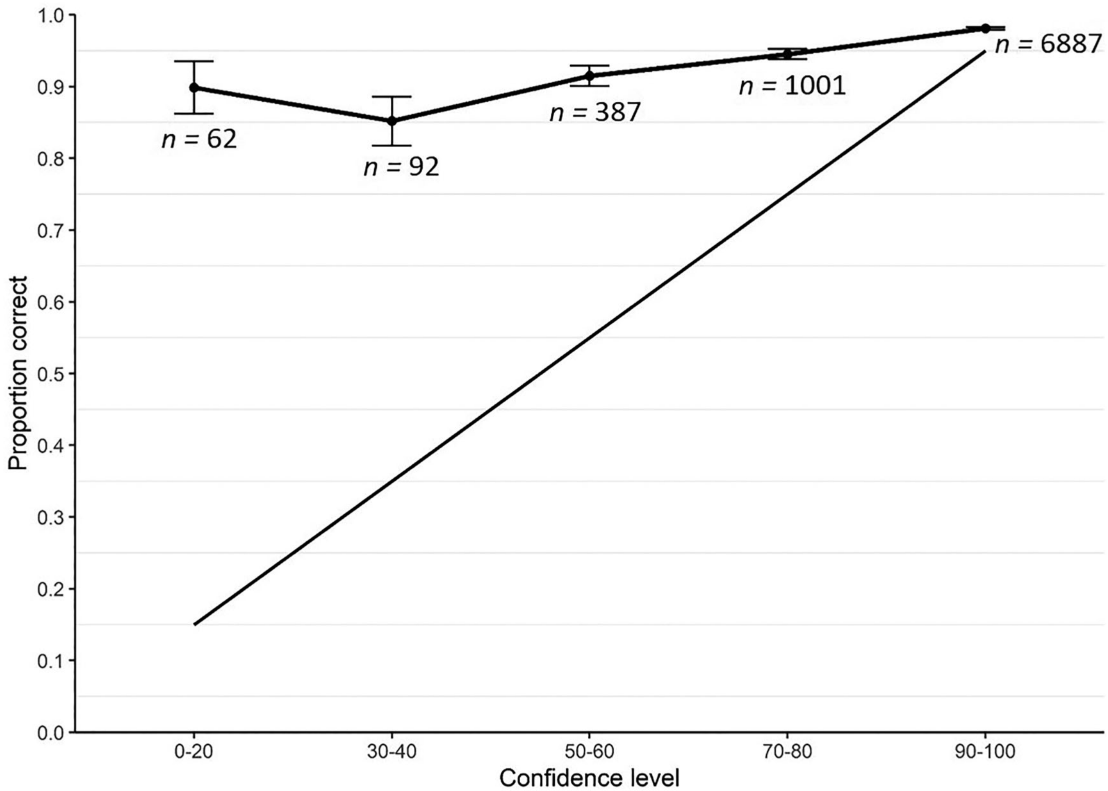

Figure 1. Calibration curves of eyewitness free recall. Error bars are 95% confidence intervals. n indicates the total amount of observations per confidence level.

confidence-accuracy relationship in the eyewitness recall test. The legalPsych package in R was used for all calibration calculations (Van Boeijen & Saraiva, 2018). We first produced calibration curves for the eyewitness free recall test (see Figure 1) and the other memory tests (see Figure 2). The diagonal line represents perfect calibration, such that each level of confidence is equivalent to the level of accuracy for decisions made with that level of confidence. Observations above this line indicate underconfidence, and observations below this line indicate overconfidence. Three calibration statistics were computed: calibration index, over/underconfidence and resolution. Calibration (*C*) represents how far a given calibration curve is from a perfect calibration. It ranges from 0 (perfect calibration) to 1, and lower values indicate better calibration. Over/underconfidence (*O/U*) indicates if a curve strays more above or below the perfect calibration line, with values ranging from −1 (very underconfident) to 1 (very overconfident). The Adjusted Normalized Resolution Index (*ANRI*) represents how well confidence discriminates accurate from inaccurate responses, with higher values indicating better discrimination (see Brewer and Wells (2006)). Following Palmer et al. (2013), we used a jackknife procedure to compute standard errors for each calibration statistic, which were then converted to 95% inferential confidence intervals (Tryon, 2001). If the confidence intervals do not overlap, that indicates a significant difference (see Table 4). The *C* statistic pointed to a reasonably strong calibration for the eyewitness free recall test ($C = .02$). However, examining the calibration curves it can be observed that most information disclosed was accurate, even those reported with low levels of confidence, so that confidence and accuracy in the eyewitness free recall did not co-vary systematically (see Figure 1). Therefore, the calibration statistics (C, O/U and ANRI) should be interpreted taking into account the low variability in accuracy for the eyewitness free recall test. For example, the C statistic points to a strong calibration for the eyewitness free recall test ($C = .02$) mainly due to a high number of correct responses given with high confidence levels (e.g., 90–100% confident). Performance in the face recognition was the most distant from perfect calibration, presenting underconfidence for lower levels of confidence and overconfidence for higher levels of confidence.

Next, we compared calibration statistics between high and low scorers on each of the metamemory measures. Following a procedure similar to the one used by Olsson and Juslin (1999), individuals above the 66th percentile were selected as high scorers and individuals below the 33th percentile were selected as low scorers (see Figure 3). Inspection of the confidence intervals suggests that high scorers in the MMQ-Contentment, MMQ-Ability, SSMQ, and General Memory Skill were less underconfident and slightly better calibrated than low scorers in those components. There were no observable differences between

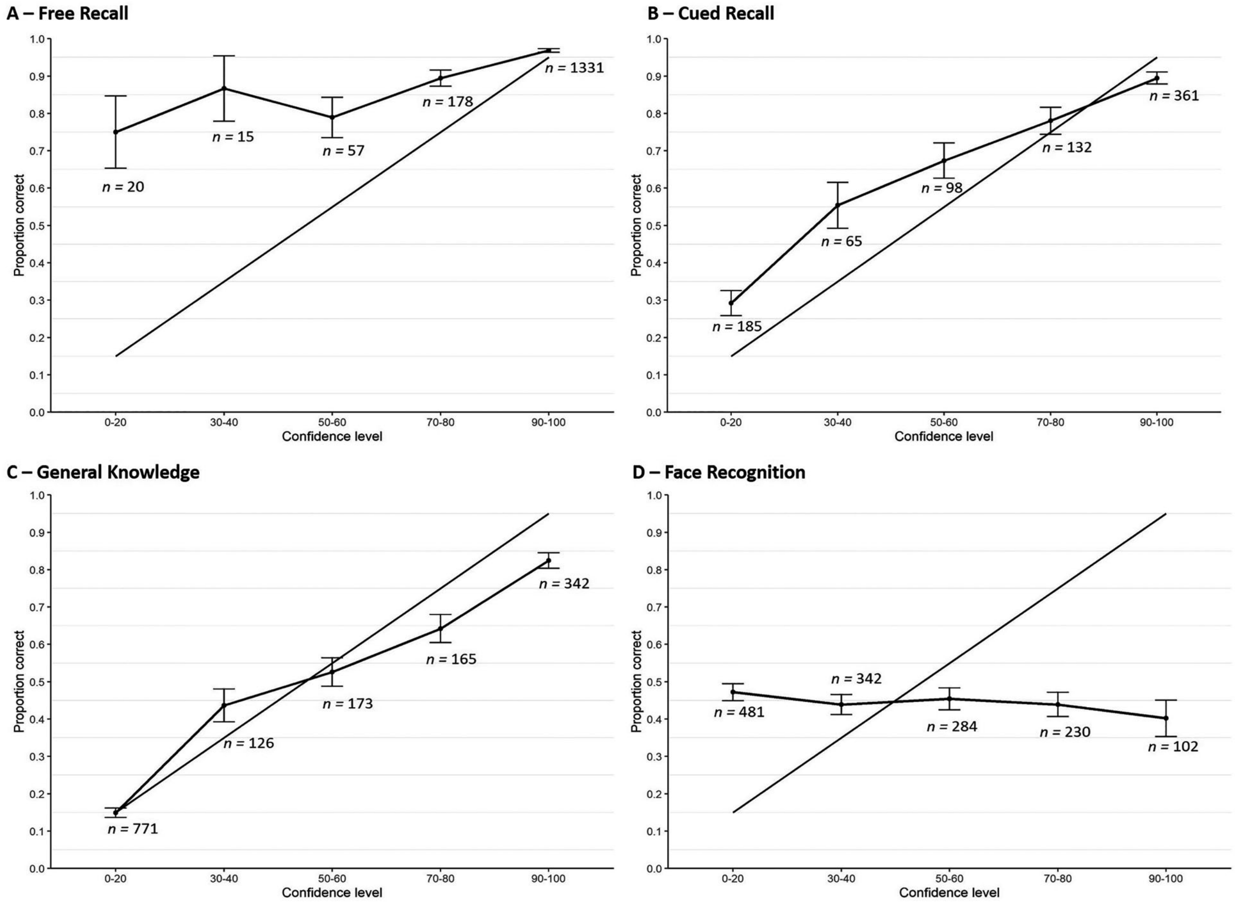

Figure 2. Calibration curves of the objective memory tests. Error bars are 95% confidence intervals. n indicates the total amount of observations per confidence level.

the resolution scores of high and low metamemory scorers. Therefore, individuals with higher self-rated scores in some metamemory factors had a slightly stronger confidence-accuracy relation in the eyewitness free recall.

Discussion

Recently there has been great interest in the predictive utility of subjective and objective memory assessments for eyewitness performance (e.g., Baldassari, Kantner, & Lindsay, 2019; Russ, Sauerland, Lee, & Bindemann, 2018). In the current study testing the use of metamemory measures and memory tests as predictors of eyewitness free recall performance, we contribute three key findings: (1) Contrary to some of our hypotheses, the metamemory scales examined had no relation with accuracy, amount of details reported, confidence or over/underconfidence in eyewitness free recall; (2) Individuals with high metamemory scores presented a slightly stronger confidence-accuracy calibration in eyewitness free recall; and, (3) Eyewitness free recall confidence and over/underconfidence was predicted by performance in objective memory tests. These findings extend our understanding of how eyewitness performance relates to subjective self-assessments of memory ability (Bornstein & Zickafoose, 1999; Evans & Fisher, 2011) and objective memory performance in different tasks (Bindemann et al., 2012; Morgan et al., 2007).

We predicted that self-assessments of memory capacity and other metamemory components would be related to eyewitness accuracy, amount of details reported, confidence and over/underconfidence. Previous studies have found positive associations between eyewitness free recall performance and individual differences such as working memory capacity, intelligence, and temperament (Chae & Ceci, 2005; Jaschinski & Wentura, 2002; Zhu et al., 2010). In the current study, we adopted diverse

Table 4. Calibration statistics for each memory test with inferential confidence intervals (ICI).

	C [ICI]	*OU* [ICI]	*ANRI* [ICI]
Eyewitness recall	0.02 [0.01,0.02]	−0.05 [−0.05, −0.04]	0.04 [0.02, 0.06]
Free recall	0.01 [0.01,0.01]	−0.02 [−0.03, −0.01]	0.09 [0.02, 0.15]
Cued recall	0.02 [0.01, 0.03]	−0.05 [−0.08, −0.02]	0.25 [0.19, 0.31]
General Knowledge	0.01 [0.01, 0.02]	−0.01 [−0.03, 0.01]	0.35 [0.30, 0.40]
Face recognition	0.07 [0.06, 0.09]	−0.03 [−0.06, 0.00]	0.01 [−0.01,0.01]

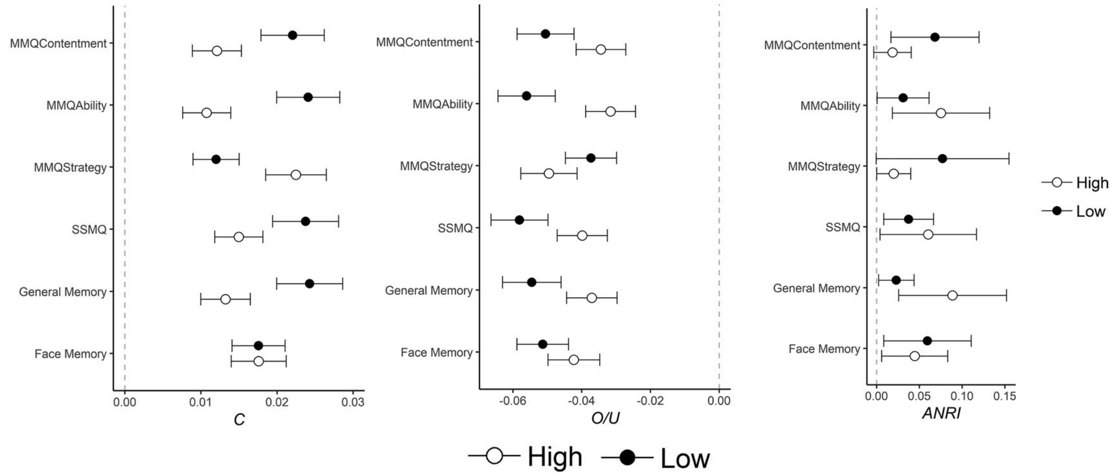

Figure 3. Inferential confidence intervals of calibration statistics for high and low scorers in each metamemory measure.

measures assessing different components of metamemory, such as memory ability, endorsement of memory strategies, and memory development over time. Surprisingly, there is no evidence that the metamemory scales used related to eyewitness free recall performance, suggesting that intrinsic cues of general memory capacity may have no relation with the amount of details reported or quality of eyewitness memory reports. The metamemory scales are also unrelated to the confidence or over/underconfidence in the eyewitness free recall, suggesting that confidence expressed for freely recalled information may originate mostly from memory trace strength, rather than self-credibility cues.

Some theoretical frameworks propose that individuals freely recalling a witnessed event monitor their memories and control what they report in order to achieve an acceptable accuracy (Evans & Fisher, 2011; Koriat & Goldsmith, 1996). Koriat et al. (2008) argue that monitoring process may be based on domain-specific beliefs (e.g., "I do not have a very good memory") or on the learning experience (i.e., the experience of processing and remembering the learning material). Our findings demonstrate that learning experience may be more influential than domain-specific beliefs when eyewitnesses are monitoring which and how much information to disclose during free recall. If replicated, these findings can be an indication that individual differences in metamemory are of little value in estimating eyewitnesses' disclosure of information. However, given that witnesses in the current study demonstrated high levels of accuracy which can also explain why learning experience was more influential than domain-specific beliefs, these relationships need to be further examined by testing witnessing conditions with higher variability in encoding difficulty. Furthermore, our metamemory assessment focused on broader memory self-efficacy domains (e.g., "I am unhappy with my general memory ability"), so future studies may benefit from examining metamemory assessments that are more specifically related to eyewitness domains (e.g., Bobak, Mileva, & Hancock, 2019; Saraiva et al., 2019).

Apart from the finding that metamemory measures are not predictive of free recall over/underconfidence on the individual level, we observed that individuals who score high on some metamemory components have a stronger confidence-accuracy relationship than low scorers in the group level. Specifically, individuals who claim to have higher memory contentment, self-rated memory ability, and memory development over time are slightly less underconfident than those individuals with lower scores in the same components. So, at least at the group level, individuals with higher levels of contentment with their own general memory ability may present confidence statements that better reflect their probable accuracy in an eyewitness free recall task. In other words, individuals who are not content with their own memories may present lower confidence statements even if they are probably accurate. That finding is consistent with previous results by Olsson and Juslin (1999), who observed that individuals who rated themselves as good face recognisers demonstrated a more diagnostic confidence-accuracy relation. It is important to note, however, that this effect was mostly associated to lower level of confidence, given that responses with high levels of confidence were almost invariantly correct. That is, high levels of confidence in the free recall tasks were almost always associated with correct information. It should also noted that participants demonstrated high levels of accuracy in the eyewitness free recall task, so the observed confidence-accuracy relation may differ in situations where witnesses do not present high free recall accuracy. Furthermore, the associations between those metamemory measures (i.e., memory contentment, self-rated memory ability, and memory development) and eyewitness over/underconfidence were only observed on the group level (i.e., low raters vs high raters), and no direct relation was observed when testing the same measures as predictors of over/underconfidence.

In contrast to subjective self-ratings of memory ability, stable individual differences in objective memory performance may be a better indicator of eyewitness free recall performance. In fact, in some studies it was found that performance in face recognition tests is somewhat predictive of eyewitness lineup performance (e.g., Baldassari et al., 2019; Russ et al., 2018). In the current study, we predicted that performance in different memory tests would be related to performance in an eyewitness free recall task. The results show that amount of details reported in an eyewitness free recall was closely related to the amount of details reported in an unrelated free recall. This finding seems to indicate some stability in the reporting of accurate information across different testing situations, which may be explained by individual differences related to attentional and cognitive resources, engagement with the task, or a combination of both (Aslan & Bäuml, 2011; Melby-Lervåg, Redick, & Hulme, 2016).

We also observed that confidence and over/underconfidence in the free recall, cued recall and general knowledge tests were predictors of confidence and over/underconfidence in the eyewitness free recall test. This finding corroborates previous findings showing that individuals express confidence in a somewhat stable manner across different testing conditions, including eyewitness memory domain (Bornstein & Zickafoose, 1999; Jonsson & Allwood, 2003; Kelemen, Frost, & Weaver, 2000; Mengelkamp & Bannert, 2010). Jonsson and Allwood (2003), for example, found some individual stability in confidence judgements for word knowledge and logical/spatial ability tasks, while Bornstein and Zickafoose (1999) found that overconfidence, calibration, and resolution in a general knowledge and cued recall test were positively correlated. We expand on those findings by presenting evidence of individual stability not only for confidence, but also for over/underconfidence in eyewitness free recall and unrelated tests (i.e., free recall, cued recall and general knowledge). Interestingly, this association was stronger for memory tests more closely related to the eyewitness free recall task (e.g., free recall, followed by cued recall and general knowledge). Complementarily, there was no stability between the eyewitness free recall and face recognition task, a finding in line with models outlining independent systems for face, episodic and semantic memory (Yovel & Kanwisher, 2004). Such consistency in individual's confidence and overconfidence for similar memory domains may be explained by stable metacognitive cues (e.g., previous experience in that memory domain; Koriat et al., 2008). Taken together, our results support frameworks that propose domain-specific memory self-efficacy, defined as an individual's appraisal of his or her usual ability in a given memory domain (Beaudoin & Desrichard, 2011; Hertzog & Dixon, 1994). Furthermore, these findings align with previous studies suggesting that self-confidence traits are meaningfully related to cognitive abilities (Kleitman & Stankov, 2007).

Eyewitness statements are commonly sought to aid investigations or to obtain incriminatory or exculpatory evidence. We found no indication for a relation between eyewitness free recall performance and metamemory self-assessments, including self-rated memory ability, endorsement of memory strategies and memory development over time. This finding may contribute to two theoretical predictions requiring further research: (1) individuals may have limited insight on their own *general* memory ability (Beaudoin & Desrichard, 2011; Perfect, 2004); or (2) individuals may have accurate insight on their *general* memory ability, but such intrinsic cues have little influence on the disclosure of witnessed events (Evans & Fisher, 2011; Koriat & Goldsmith, 1996). From a practical perspective, it is problematic to question the credibility of eyewitness reports based on general self-assessments of memory capacity. As argued by Evans and Fisher (2011), contrary to common belief eyewitnesses who say "I don't know" more often may be more accurate than other witnesses, given that they are better monitoring their reporting to provide accurate information. Additionally, we observed that on the group level individuals who distrust their own memories have a weaker confidence-accuracy relationship, presenting lower confidence statements even if they are likely accurate. If replicated, this finding could offer a basis for metamemory assessments to better discriminate overconfident and underconfident eyewitnesses. Finally, we found some stability between eyewitness free recall performance and performance in related memory tasks (i.e., unrelated free recall, cued recall and general knowledge). This result lends some support to the concept of an objective test designed to estimate eyewitness free recall performance, an approach that has been tested for eyewitness identification settings (Baldassari et al., 2019; Bindemann et al., 2012; Morgan et al., 2007).

There are a number of limitations that should be taken into consideration. First, the mock-crime video was seen in the same session as the eyewitness free recall test. This procedure was adopted to guarantee the feasibility of the study, but in more naturalistic contexts eyewitnesses commonly report what they have seen after a longer delay. In fact, it is unlikely that eyewitnesses would provide their accounts immediately after the critical event, given that officers who are deployed to the crime event may be faced with responsibilities that take precedence over interviewing witnesses (e.g., securing and preserving the crime scene). Importantly, previous studies have found differential effects of self-capacity measures and stability in memory tests that were dependant on task difficulty (Howie & Roebers, 2007; Mengelkamp & Bannert, 2010; Stankov, 2000). It may be the case, for example, that the relation between objective memory tests and eyewitness free recall performance is weaker if there is significant memory decay for information about the crime. Furthermore, we observed high accuracy and high confidence in the eyewitness free recall reports, which may have been due to the nature of the task or because the interval between encoding and retrieval was short making the task easy. The high accuracy observed for the eyewitness free recall test also limits the interpretation of calibration findings, given that the reported calibration statistics (C, O/U, ANRI) represent average point estimates. Therefore, the procedures adopted to guarantee the feasibility of the study should be taken into account when generalising the observed results to real crime situations. We highlight the need for research that investigates associations between measures of self-efficacy, performance in objective memory tests, and eyewitness free recall performance under varying levels of difficulty (e.g., longer retention intervals; Sauer et al., 2010). A second limitation is that the metamemory assessment always occurred 24 h before the eyewitness paradigm. In naturalistic contexts, such metamemory assessment would realistically occur after an eyewitness account was obtained. In determining our procedure, we reasoned that exposure to the eyewitness paradigm and other memory tests before the completion of the metamemory measures would have affected witnesses' self-assessments of their memory abilities to a greater degree than completing the assessments would affect eyewitness performance. While this was perhaps appropriate for an initial test of relationships between the tasks, future studies should examine the use of metamemory measures after the eyewitness task.

Conclusions

In sum, our results allow for three main inferences. First, metamemory assessments appear to have little value in estimating individual performance in eyewitness free recall settings. With reference to utility in the applied context, this finding suggests that dismissing or questioning the credibility of eyewitness reports based on self-assessments of memory capacity may be unwarranted. Second, at the group level individuals with higher self-ratings of memory capacity had a slightly stronger confidence-accuracy calibration. This is initial evidence that free recall confidence is a better predictor of accuracy among individuals that are not very doubtful about their own memory performance. Third, we find stability in confidence and over/underconfidence measures across eyewitness free recall and other memory tests of similar domain. If this relationship is replicated in future research it may indicate that individual differences or intrinsic metamemory cues (e.g., experience with memory issues) partly explain levels of realism for confidence judgements in memory tasks.

Acknowledgements

We thank Joseph Finch, Heidi Mcintosh, and Leigha West for their assistance with data collection and data coding.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported by a fellowship award from the Erasmus Mundus Joint Doctorate Program the House of Legal Psychology (EMJD-LP) with Framework Partnership Agreement (FPA) 2013–0036 and Specific Grant Agreement (SGA) 2013-0036.

Data Availability Statement

All datasets and R analysis scripts for this study can be found at https://osf.io/cywtk/?view_only=69155f972cf84e3bbad4321cfece650d

ORCID

Renan Benigno Saraiva http://orcid.org/0000-0001-9461-7594

References

Allwood, C. M., Jonsson, A. C., & Granhag, P. A. (2005). The effects of source and type of feedback on child witnesses' metamemory accuracy. *Applied Cognitive Psychology*, *19*, 331–344. doi:10.1002/acp.1071

Aslan, A., & Bäuml, K. H. T. (2011). Individual differences in working memory capacity predict retrieval-induced forgetting. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *37*, 264–269. doi:10.1037/a0021324

Baldassari, M. J., Kantner, J., & Lindsay, D. S. (2019). The importance of decision bias for predicting eyewitness lineup choices: Toward a lineup skills test. *Cognitive Research: Principles and Implications*, *4* (2), 1–13. doi:10.1186/s41235-018-0150-3

Beaudoin, M., & Desrichard, O. (2011). Are memory self-efficacy and memory performance related? A meta-analysis. *Psychological Bulletin*, *137*, 211–241. doi:10.1037/a0022106

Bindemann, M., Brown, C., Koyas, T., & Russ, A. (2012). Individual differences in face identification postdict eyewitness accuracy. *Journal of Applied Research in Memory and Cognition*, *1*, 96–103. doi:10.1016/j.jarmac.2012.02.001

Bobak, A. K., Mileva, V. R., & Hancock, P. J. (2019). Facing the facts: Naive participants have only moderate insight into their face recognition and face perception abilities. *Quarterly Journal of Experimental Psychology*, *72*, 872–881. doi:10.1177/1747021818776145

Bornstein, B. H., & Zickafoose, D. J. (1999). "I know I know it, I know I saw it": The stability of the confidence–accuracy relationship across domains. *Journal of Experimental Psychology. Applied*, *5*, 76–88. doi:10.1037/1076-898X.5.1.76

Brewer, N., & Burke, A. (2002). Effects of testimonial inconsistencies and eyewitness confidence on mock-juror judgments. *Law and Human Behavior*, *26*, 353–364. doi:10.1023/A:1015380522722

Brewer, N., Keast, A., & Rishworth, A. (2002). The confidence-accuracy relationship in eyewitness identification: The effects of reflection and disconfirmation on correlation and calibration. *Journal of Experimental Psychology. Applied*, *8*(1), 44–56. doi:10.1037/1076-898X.8.1.44

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

Chae, Y., & Ceci, S. J. (2005). Individual differences in children's recall and suggestibility: The effect of intelligence, temperament, and self-perceptions. *Applied Cognitive Psychology*, *19*, 383–407. doi:10.1002/acp.1094

Dahl, M., Allwood, C. M., Scimone, B., & Rennemark, M. (2015). Old and very old adults as witnesses: Event memory and metamemory. *Psychology, Crime & Law*, *21*, 764–775. doi:10.1080/1068316X.2015.1038266

Douglass, A. B., & Jones, E. E. (2013). Confidence inflation in eyewitnesses: Seeing is not believing. *Legal and Criminological Psychology*, *18*, 152–167. doi:10.1111/j.2044-8333.2011.02031.x

Dunlosky, J., & Bjork, R. A. (2008). The integrated nature of metamemory and memory. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 11–28). New York, NY: Psychology Press.

Evans, J. R., & Fisher, R. P. (2011). Eyewitness memory: Balancing the accuracy, precision and quantity of information through metacognitive monitoring and control. *Applied Cognitive Psychology*, *25*, 501–508. doi:10.1002/acp.1722

Hertzog, C., & Dixon, R. A. (1994). Metacognitive development in adulthood and old age. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 227–251). Cambridge, MA: MIT Press.

Hertzog, C., McGuire, C. L., & Lineweaver, T. T. (1998). Aging, attributions, perceived control, and strategy use in a free recall task. *Aging, Neuropsychology, and Cognition*, *5*, 85–106. doi:10.1076/anec.5.2.85.601

Howie, P., & Roebers, C. M. (2007). Developmental progression in the confidence-accuracy relationship in event recall: Insights provided by a calibration perspective. *Applied Cognitive Psychology*, *21*, 871–893. doi:10.1002/acp.1302

Jaschinski, U., & Wentura, D. (2002). Misleading postevent information and working memory capacity: An individual differences approach to eyewitness memory. *Applied Cognitive Psychology*, *16*, 223–231. doi:10.1002/acp.783

Jonsson, A.-C., & Allwood, C. M. (2003). Stability and variability in the realism of confidence judgments over time, content domain, and gender. *Personality and Individual Differences*, *34*, 559–574. doi:10.1016/S0191-8869(02)00028-4

Kelemen, W. L., Frost, P. J., & Weaver, C. A. (2000). Individual differences in metacognition: Evidence against a general metacognitive ability. *Memory & Cognition*, *28*, 92–107. doi:10.3758/bf03211579

Kleitman, S., & Stankov, L. (2007). Self-confidence and metacognitive processes. *Learning and Individual Differences*, *17*, 161–173. doi:10.1016/j.lindif.2007.03.004

Koriat, A. (2000). The feeling of knowing: Some metatheoretical implications for consciousness and control. *Consciousness and Cognition*, *9*, 149–171. doi:10.1006/ccog.2000.0433

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General*, *133*, 643–656. doi:10.1037/0096-3445.133.4.643

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

Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory accuracy. *Annual Review of Psychology*, *51*, 481–537. doi:10.1146/annurev.psych.51.1.481

Koriat, A., Nussinson, R., Bless, H., & Shaked, N. (2008). Information-based and experience-based metacognitive judgments: Evidence from subjective confidence. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 117–136). New York, NY: Psychology Press.

Kruschke, J. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. London: Academic Press.

Leippe, M. R., & Eisenstadt, D. (2014). Eyewitness confidence and the confidence-accuracy relationship in memory for people. In R. C. Lindsay, D. G. Ross, J. D. Read, & M. P. Toglia (Eds.), *The handbook of eyewitness psychology* (pp. 377–425). New York, NY: Psychology Press.

Leippe, M. R., Eisenstadt, D., & Rauch, S. M. (2009). Cueing confidence in eyewitness identifications: Influence of biased lineup instructions and pre-identification memory feedback under varying lineup conditions. *Law and Human Behavior*, *33*, 194–212. doi:10.1007/s10979-008-9135-y

Leippe, M. R., Eisenstadt, D., Rauch, S. M., & Stambush, M. A. (2006). Effects of social-comparative memory feedback on eyewitnesses' identification confidence, suggestibility, and retrospective memory reports. *Basic and Applied Social Psychology*, *28*, 201–220. doi:10.1207/s15324834basp2803_1

Lindsay, D. S., Read, J. D., & Sharma, K. (1998). Accuracy and confidence in person identification: The relationship is strong when witnessing conditions vary widely. *Psychological Science*, *9*, 215–218. doi:10.1111/1467-9280.00041

Manning, W. G., Basu, A., & Mullahy, J. (2005). Generalized modeling approaches to risk adjustment of skewed outcomes data. *Journal of Health Economics*, *24*, 465–488. doi:10.1016/j.jhealeco.2004.09.011

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*, 270–291. doi:10.1037/a0028228

Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of "far transfer": Evidence from a meta-analytic review. *Perspectives on Psychological Science*, *11*, 512–534. doi:10.1177/1745691616635612

Mengelkamp, C., & Bannert, M. (2010). Accuracy of confidence judgments: Stability and generality in the learning process and predictive validity for learning outcome. *Memory & Cognition*, *38*, 441–451. doi:10.3758/MC.38.4.441

Morgan, C. A., 3rd, Hazlett, G., Baranoski, M., Doran, A., Southwick, S., & Loftus, E. (2007). Accuracy of eyewitness identification is significantly associated with performance on a standardized test of face recognition. *International Journal of Law and Psychiatry*, *30*, 213–223. doi:10.1016/j.ijlp.2007.03.005

Nelson, T. O., & Narens, L. (1994). Why investigate metacognition. Metacognition: Knowing about knowing, 1–25. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 1–26). Cambridge, MA: MIT Press.

Odinot, G., & Wolters, G. (2006). Repeated recall, retention interval and the accuracy–confidence relation in eyewitness memory. *Applied Cognitive Psychology*, *20*, 973–985. doi:10.1002/acp.1263

Odinot, G., Wolters, G., & van Giezen, A. (2013). Accuracy, confidence and consistency in repeated recall of events. *Psychology, Crime & Law*, *19*, 629–642. doi:10.1080/1068316X.2012.660152

Olsson, N., & Juslin, P. (1999). Can self-reported encoding strategy and recognition skill be diagnostic of performance in eyewitness identifications? *The Journal of Applied Psychology*, *84*, 42–49. doi:10.1037/0021-9010.84.1.42

Palmer, M. A., Brewer, N., Weber, N., & Nagesh, A. (2013). The confidence-accuracy relationship for eyewitness identification decisions: Effects of exposure duration, retention interval, and divided attention. *Journal of Experimental Psychology: Applied*, *19*, 55–71. doi:10.1037/a0031602

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

Perfect, T. J. (2004). The role of self-rated ability in the accuracy of confidence judgements in eyewitness memory and general knowledge. *Applied Cognitive Psychology*, *18*, 157–168. doi:10.1002/acp.952

Robinson, M. D., & Johnson, J. T. (1996). Recall memory, recognition memory, and the eyewitness confidence–accuracy correlation. *The Journal of Applied Psychology*, *81*, 587–594. doi:10.1037/0021-9010.81.5.587

Robinson, M. D., Johnson, J. T., & Herndon, F. (1997). Reaction time and assessments of cognitive effort as predictors of eyewitness memory accuracy and confidence. *The Journal of Applied Psychology*, *82*, 416–425. doi:10.1037/0021-9010.82.3.416

Robinson, M. D., Johnson, J. T., & Robertson, D. A. (2000). Process versus content in eyewitness metamemory monitoring. *Journal of Experimental Psychology. Applied*, *6*, 207–221. doi:10.1037/1076-898X.6.3.207

Russ, A. J., Sauerland, M., Lee, C. E., & Bindemann, M. (2018). Individual differences in eyewitness accuracy across multiple lineups of faces. *Cognitive Research: Principles and Implications*, *3*(1), 1–17. doi:10.1186/s41235-018-0121-8

Saraiva, R. B., van Boeijen, I. M., Hope, L., Horselenberg, R., Sauerland, M., & van Koppen, P. J. (2019). Development and validation of the Eyewitness Metamemory Scale. *Applied Cognitive Psychology*, *33*, 964–973. doi:10.1002/acp.3588

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

Sauer, J., & Hope, L. (2016). The effects of divided attention at study and reporting procedure on regulation and monitoring for episodic recall. *Acta Psychologica*, *169*, 143–156. doi:10.1016/j.actpsy.2016.05.015

Sauerland, M., & Sporer, S. L. (2007). Post-decision confidence, decision time, and self-reported decision processes as postdictors of identification accuracy. *Psychology, Crime & Law*, *13*(6), 611–625. doi:10.1080/10683160701264561

Searcy, J., Bartlett, J. C., & Memon, A. (2000). Influence of post-event narratives, line-up conditions and individual differences on false identification by young and older eyewitnesses. *Legal and Criminological Psychology*, *5*, 219–235. doi:10.1348/135532500168100

Shaw, J. S., & Zerr, T. K. (2003). Extra effort during memory retrieval may be associated with increases in eyewitness confidence. *Law and Human Behavior*, *27*, 315–329. doi:10.1023/A:1023487908640

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. L. (2016). Do "brain-training" programs work? *Psychological Science in the Public Interest*, *17*, 103–186. doi:10.1177/1529100616661983

Sporer, S. L., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin*, *118*(3), 315–327. doi:10.1037/0033-2909.118.3.315

Squire, L. R., Wetzel, C. D., & Slater, P. C. (1979). Memory complaint after electroconvulsive therapy: Assessment with a new self-rating instrument. *Biological Psychiatry*, *14*, 791–801.

Stankov, L. (2000). Complexity, metacognition, and fluid intelligence. *Intelligence*, *28*, 121–143. doi:10.1016/S0160-2896(99)00033-1

Thomaz, C. E., & Giraldi, G. A. (2010). A new ranking method for principal components analysis and its application to face image analysis. *Image and Vision Computing*, *28*, 902–913. doi:10.1016/j.imavis.2009.11.005

Troyer, A. K., & Rich, J. B. (2002). Psychometric properties of a new metamemory questionnaire for older adults. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *57*, 19–27. doi:10.1093/geronb/57.1.P19

Tryon, Warren W. (2001). Evaluating statistical difference, equivalence, and indeterminancy using inferential confidence intervals: An integrated alternative method of conducting null hypothesis statistical tests. *Psychological Methods*, *6*(4), 371–386. doi:10.1037//1082-989X.6.4.371

Van Bergen, S., Brands, I., & Jelicic, M. (2010). Assessing trait memory distrust: Psychometric properties of the Squire Subjective Memory Questionnaire. *Legal and Criminological Psychology*, *15*, 373–384. doi:10.1348/135532509X471960

Van Boeijen, I. M., & Saraiva, R. (2018). legalPsych: A tool for calculating calibration statistics in eyewitness research. GitHub: https://github.com/IngerMathilde/legalPsych

Van Koppen, P. J., & Lochun, S. K. (1997). Portraying perpetrators: The validity of offender descriptions by witnesses. *Law and Human Behavior*, *21*, 661–685. doi:10.1023/A:1024812831576

Weber, N., & Brewer, N. (2008). Eyewitness recall: Regulation of grain size and the role of confidence. *Journal of Experimental Psychology: Applied*, *14*, 50–60. doi:10.1037/1076-898X.14.1.50

Weber, N., Brewer, N., Wells, G. L., Semmler, C., & Keast, A. (2004). Eyewitness identification accuracy and response latency: The unruly 10–12-second rule. *Journal of Experimental Psychology. Applied*, *10*, 139–147. doi:10.1037/1076-898X.10.3.139

Wixted, J. T., Mickes, L., Clark, S. E., Gronlund, S. D., & Roediger, H. L., 3rd. (2015). Initial eyewitness confidence reliably predicts eyewitness identification accuracy. *The American Psychologist*, *70*, 515–526. doi:10.1037/a0039510

Wixted, J. T., & Wells, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest*, *18*, 10–65. doi:10.1177/1529100616686966

Yovel, G., & Kanwisher, N. (2004). Face perception: Domain specific, not process specific. *Neuron*, *44*, 889–898. doi:10.1016/j.neuron.2004.11.018

Zhu, B., Chen, C., Loftus, E. F., Lin, C., He, Q., Chen, C., … Dong, Q. (2010). Individual differences in false memory from misinformation: Cognitive factors. *Memory*, *18*, 543–555. doi:10.1080/09658211.2010.487051