

Research Article

Natalie Martschuk*, Siegfried L. Sporer, Melanie Sauerland

Confidence of Older Eyewitnesses: Is It Diagnostic of Identification Accuracy?

<https://doi.org/10.1515/psych-2018-0010>

Received May 24, 2018; accepted November 24, 2018

Abstract: Since the late 1980s evidence has been accumulating that confidence recorded at the time of identification is a reliable postdictor of eyewitness identification. Nonetheless, there may be noteworthy exceptions. In a re-analysis of a field study by Sauerland and Sporer (2009; $N = 720$; $n = 436$ choosers between 15 and 83 years old) we show that the postdictive value of confidence was reduced for participants aged 40 years or older. Different calibration indices and Bayesian analyses demonstrate a progressive dissociation between identification performance and confidence across age groups. While the confidence expressed following an identification remained unchanged across the lifespan, identification accuracy decreased. Young, highly confident witnesses were much more likely to be accurate than less confident witnesses. With increasing age, witnesses were more likely to be overconfident, particularly at the medium and high levels of confidence, and the postdictive value of confidence and decision times decreased. We conclude that witness age may be an important moderator to take into account when evaluating identification evidence.

Keywords: eyewitness identification; confidence–accuracy relationship; elderly eyewitnesses; lineup

Confidence of Older Eyewitnesses: Is It Diagnostic of Identification Accuracy?

Eyewitness identification is an important determinant of trial outcomes, and mistaken identifications have frequently contributed to miscarriages of justice (Garrett, 2011). This is particularly true for in-court identifications, which are often presented with inflated confidence. Early reviews and meta-analyses of the confidence–accuracy relationship concluded that confidence only modestly predicts the accuracy of eyewitnesses identification (Bothwell, Deffenbacher, & Brigham, 1987; Wells & Murray, 1983). Consequently, many researchers advised against using expressed confidence as an indicator of accuracy (e.g., Penrod & Cutler, 1995). By contrast, when focusing on choosers of the suspect – only those who usually end up testifying in court against the defendant – the meta-analysis by Sporer, Penrod, Read, and Cutler (1995) revealed a *strong* confidence–accuracy correlation (r_{pb} weighted = .37; r_{pb} unweighted = .41).¹ However,

¹ Note that point-biserial correlations of .10, .24, and .37, are equivalent to standardized mean difference d values of .20 (“small”), .50 (“medium”), and .80 (“large”) effects (Cohen, 1988).

Article note: This article is a part of Topical Issue “Strategy Contributions to Cognitive Aging” edited by Beatrice G. Kuhlmann. The research is part of Natalie Martschuk’s doctoral dissertation at the University of Giessen, Germany.

***Corresponding author: Natalie Martschuk**, University of Giessen, Giessen, Germany; now at Charles Sturt University, Manly, Australia, E-mail: nmartschuk@csu.edu.au

Siegfried L. Sporer, University of Giessen

Melanie Sauerland, Maastricht University

Sporer et al. (1995) also noted a large heterogeneity across studies, ranging between $r_{pb} = .10$ and $r_{pb} = .74$. Subsequent studies found even higher confidence–accuracy correlations (e.g., Lindsay, Nilsen, & Read, 2000; Read, 1995). Therefore, it seems worth exploring under which conditions these relationships may be particularly high or not so high.

Recently, in several re-analyses of previous research, Wixted and colleagues concluded that confidence – if measured and recorded at the time of an initial eyewitness identification – was a reliable marker of identification accuracy (Wixted, Mickes, Clark, Gronlund, & Roediger, 2015; Wixted, Read, & Lindsay, 2016). However, this conclusion may only hold under “pristine” conditions (i.e., fair lineup instructions, fair lineups, etc.: Wixted & Wells, 2017). As noted by Palmer, Brewer, Weber, and Nagesh (2013), “it remains an empirical question as to whether there are conditions under which the confidence–accuracy relationship does collapse entirely” (p. 66). Here we draw attention to the possibility that even under “pristine” conditions moderators of the confidence–accuracy relationship do exist and may affect accuracy independently of confidence, thus yielding a dissociation that reduces the probative value of confidence. In particular, we focus on the age of eyewitnesses – from juveniles to the aged – as a potential moderator. While there is a robust effect of age on recognition performance (see the meta-analyses by Fitzgerald & Price, 2015; Kocab, Martschuk, & Sporer, 2018; Martschuk & Sporer, 2018; see also the review by Bartlett & Memon, 2007), much less is known about *confidence* of older witnesses, and in particular the confidence–accuracy relationship.

Metamemory Models of the Confidence–Accuracy Relationship

According to the direct-access view of metacognitive judgments, reported confidence can be understood as an indicator of the strength of a memory trace: The stronger the trace, the more likely a piece of information will be recalled or recognized. Followingly, in an identification decision confidence is expected to be higher when the witness experiences a strong sense of familiarity that the suspect in the lineup matches her or his image of the perpetrator. Signal detection theory models applied to eyewitness identification decisions still rely on this assumption but acknowledge that other types of information (e.g., feedback provided by a police officer) may also affect confidence (for a recent review, see Wixted & Wells, 2017).

An alternative meta-memorial explanation prevalent in the basic literature on memory is *cue utilization theory* (Koriat, 1997), which suggests that people use different types of cues to assign confidence judgments through analytic heuristics. However, metamemory studies have usually been conducted with paradigms that are quite different from those employed in eyewitness identification studies. Many studies have used word list or paired associate paradigms whereby words are either known or have varying degrees of associations with the stimuli-to-be-recalled or recognized. Somewhat closer to eyewitness studies are name-learning paradigms, which demonstrated monitoring accuracy for nouns as targets to be higher than for faces or names (Watier & Collin, 2011). In contrast, in eyewitness identification studies, the target is an unfamiliar person/face that has to be recognized among a list of distractors (fillers) that are usually preselected on the basis of some similarity to the target (as in the study presented here).

In studies on ease of learning, judgments of learning, or feeling of knowing, participants are asked **before** a memory test if particular items will be recalled or recognized (for review, see Dunlosky & Tauber, 2014). In eyewitness identification studies, this would correspond to the police asking a witness before an identification task, if he/she believes to be able to recognize the target in a lineup (referred to as pre-decision confidence vs post-decision confidence which is assessed after a lineup; e.g., Sporer, 1992). There is ample evidence that pre-decision confidence is unrelated to the accuracy of a later identification (see the meta-analysis by Cutler & Penrod, 1989). Hence, in more recent studies, researchers have no longer included pre-decision confidence as an assessment variable.

There is also a large body of research on metacognitive control and memory reporting (e.g. Goldsmith, Pansky, & Koriat, 2014; Koriat & Goldsmith, 1994, 1996), which has been fruitfully applied to eyewitness recall (e.g., Weber & Brewer, 2008). The major difference to identification studies is that in Koriat and

colleagues' studies the focus is on quality vs quantity of recall, not on the accuracy of recognition (but see Perfect & Weber, 2012, for a noteworthy exception).

Thus, while we do not question the importance of these metamemory studies for eyewitness *recall*, we are reluctant to apply the underlying theories to face recognition or person identification paradigms. In person identification studies, there is (usually) only one unfamiliar target-to-be recognized which has been seen once, usually without instructions to commit the face to memory (i.e., an incidental learning paradigm).

When eyewitness researchers applied Koriat's (1997) cue utilization approach to an identification paradigm (Perfect & Weber, 2012), important differences emerged that further support our claim that metamemory theories developed in verbal learning paradigms may not yield comparable results. Whereas the order for "forced reports" and "free reports" did not matter in Koriat and Goldsmith's (1996, Experiment 2) study, for eyewitness identification decisions the order was crucial (Perfect & Weber, 2012). Also, in an identification study by Weber and Perfect (2012) free-report decisions were significantly more diagnostic than forced-report decisions, with no significant reduction in the quantity of correct identifications obtained. That is, there was a benefit to memory accuracy that was accompanied by a negligible cost to memory quantity, in an apparent contradiction of the memory control framework. These data indicate that a simple version of Koriat and Goldsmith's (1996) control framework may not apply to identification decisions when a "Don't know" option is provided. Unfortunately, most identification studies do not offer a "Don't know" option, thus, complicating the comparison between studies in the basic metamemory literature and the eyewitness domain even further (see also the recent discussion on the occurrence of positive, null, and negative confidence–accuracy relationships by Roediger & DeSoto, 2015).

In an attempt to test these rival views within an identification paradigm, Busey, Tunnicliff, Loftus, and Loftus (2000) provided evidence for both the direct access and cue utilization models. When they showed photographs with varying exposure time (Experiment 1) and different levels of illuminations (Experiment 2) the findings were consistent with the trace access memory provided. However, when illumination levels between the exposure and test were varied (Experiment 3), the findings were consistent with the cue utilization approach. Generally, Busey et al. (2000) showed a high confidence–accuracy relationship, unless the conditions were not optimal at the test. This is an example where confidence and accuracy rely on mainly the same information under some circumstances but may also yield a dissociation between confidence and accuracy under different conditions.

The Confidence–Accuracy Relationship across the Lifespan

The confidence–accuracy relationship has been well researched for over 40 years, yet to our knowledge only three studies investigated the issue over the lifespan in an eyewitness identification paradigm in a field setting (Palmer et al., 2013, age range 14–87; Sauer, Brewer, Zweck, & Weber, 2010, age range 15–85; Sauerland & Sporer, 2009, age range 15–84). However, in all three studies the analyses were conducted without separating the age groups. In Sauer et al. (2010) only 6% of the participants ($n = 64$) were over 60 years of age (Sauer, 2016, personal correspondence²). Palmer et al. (2013) conducted additional analyses excluding participants younger than 18 years and older than 65 years. They found no differences in results, except for the influence of exposure time on resolution and of retention interval on calibration. When all participants were included, the resolution and calibration scores were significantly better following a longer exposure duration and a shorter retention interval. By contrast, when adolescents and older participants were excluded from the analyses, no effect of exposure time and retention interval was found. These analyses, however, did not specifically address age-related changes in the confidence–accuracy relationship. In Sauerland and Sporer (2009) more than 9% of the participants (10.8% of choosers) were over the age of 60.

Several laboratory studies that compared different age groups directly suggest age-related changes

² We greatly appreciate the authors for providing us with their age distribution data.

in the confidence–accuracy relationship: In a study with children, confidence was not diagnostic of identification accuracy (Keast, Brewer, & Wells, 2007), and in three studies the confidence of older adults was less diagnostic of accuracy than confidence of young adults (Searcy, Bartlett, & Memon, 1999; Searcy, Bartlett, Memon, & Swanson, 2001; Wright & Stroud, 2002). These results provide some evidence that the confidence–accuracy relationship varies by age, at least in laboratory settings.

A recently published online study compared eyewitness identifications of 890 young (18–30 years), 890 middle-aged (31–59) and 890 older adults (50–95 years) in an online experiment (Colloff, Wade, Wixted, & Maylor, 2017). Participants watched a 30-second video depicting a mock crime and were presented with a fair or unfair target-present or a target-absent lineup. Following their decision they rated their confidence. Analyses revealed an age-related decline in participants' ability to discriminate between guilty and innocent suspects, and to adjust their confidence accordingly using signal detection theory modeling. These results seem to suggest that the confidence–accuracy relationship is stable across the lifespan. However, in an online study the risk of a preselection bias is increased as participants are required to access and use a PC. Therefore, it is possible that a higher proportion of older people as opposed to young people might have not been able to participate (i.e., those who are not fit enough or confident to use a computer). By contrast, participants in the present study were citizens who were around a downtown shopping area during the day, and did not need to have access to a PC (as in Palmer, David, & Fleming, 2014, and Colloff et al., 2017).

Further, the major focus of the Colloff et al. study was on how best to construct fair lineups by modifying all lineup faces when the suspect has a striking mark on his face (e.g. a scar or a large bruising around the eye), using one single male target. As is evident from the literature on face recognition, adding lines to a face or blocking parts of a face disrupts holistic processing (e.g., Ellis, Davies, & Shepherd, 1978; Hancock, Bruce, & Burton, 2000; Richler, Cheung, & Gauthier, 2011; Richler, Mack, Gauthier, & Palmeri, 2009). It is an open question whether the results obtained by Colloff et al. (2017) regarding age effects would replicate with targets without any highly distinguishing marks and without editing lineup fillers to match this mark.

Taken together, there seems to be ample evidence that identification accuracy may decline at old age but we still do not know enough whether there is a parallel decline in the confidence–accuracy relationship. Here we present a re-analysis of the confidence–accuracy relationship from adolescence to old age using an incidental encoding task in a field experiment, originally published by Sauerland and Sporer (2009) without any consideration of participants' age. While the original study investigated identification decision processes independently of age separately for choosers and non-choosers, the present study focused on the influence of age on the confidence–accuracy relationship in choosers only. We refer to results from the earlier study only to the extent that they are contrasted with the present analysis.

Metamemorial Deficits of Older Eyewitnesses

Metamemory, defined as “knowledge and awareness about one's own memory and how memory works more generally” (Castel, McGillivray, & Friedman, 2012, p. 246), refers to both prospective and retrospective metacognitive judgments. The present research focused on retrospective cognitive judgments, namely the expressed confidence that an identification judgment was correct. Two reciprocal processes of metamemory have been suggested: (1) metacognitive *monitoring* and (2) metacognitive *control* (Nelson & Narens, 1990; Price, Mueller, Wetmore, & Neuschatz, 2014). Metacognitive monitoring evaluates the current state of retrieval and provides information about the learning status. Thus, monitoring consists of subjective perceptions about own introspections or mnemonic strength. Metacognitive control is a type of behavior that is used to optimize performance at a meta-level without direct access to the current state, representing an analytic factor of metacognitive judgments. Both monitoring and control are important for metacognitive judgments, and dissociations between behavior and beliefs may become more marked across the lifespan (Palmer et al., 2014).

Previous research demonstrated that while monitoring of well-learned information remained unaffected by aging, metacognitive monitoring of recently encountered information was more likely to be impaired with increasing age (Dodson, Bawa, & Krueger, 2007; Dodson & Krueger, 2006). For instance, Dodson et al. (2007) showed that older participants recognized statements as well as younger participants but made more source monitoring errors and displayed more high-confidence errors (Experiment 1). In Experiment 2,

which used a cued-recall task with word lists also showed more high-confidence errors in older compared to younger adults. Perhaps more relevant to the eyewitness literature, older eyewitnesses also showed more high-confidence suggestibility errors than younger participants after watching a video (Dodson & Krueger, 2006). However, more recent studies using word lists containing related lures suggest that positive, null, and negative confidence–accuracy relationships may be observed as a function of the stimulus material and paradigm (DeSoto & Roediger, 2014; Roediger & DeSoto, 2014). Applied to a lineup task, an innocent suspect strongly resembling the perpetrator may lead to high-confidence false identifications (Roediger & DeSoto, 2015, citing anecdotal case evidence from Buckhout, 1974). Whether this varies as a function of participant age remains an empirical question.

Metacognitive control – which is considered important in an identification task – is impaired in older age (Price *et al.*, 2014). In other words, compared to young adults, older adults are less likely to effectively use available information when making decisions (Hertzog & Hultsch, 2000; Price *et al.*, 2014) and have difficulties retrieving information (Shing, Werkle-Bergner, Li, & Lindenberger, 2009). Cross-sectional metacognitive studies showed that metacognitive efficiency improved during adolescence, reached a plateau during adulthood (age range 11–41 years; Weil *et al.*, 2013), and declined in older adulthood (age range 18–84 years; Palmer *et al.*, 2014). These studies together suggest that, over the lifespan, metacognitive efficiency resembles an inverted U: Adolescents and older adults have a stronger tendency towards overconfidence than young and middle-aged adults. However, most of this research was concerned with learning of word lists or semantic memory, not with face recognition or identification tasks. It has long been known in the memory literature that verbal and visual information, and faces in particular, are encoded and processed differently (e.g. Meissner, Sporer, & Schooler, 2007; Paivio, 1971; Tulving, 1985). Some of these differences may even be more pronounced in older adults (e.g. Jenkins, Myerson, Hale, & Fry, 1999). The present study extended these theoretical notions to the eyewitness identification task by analyzing the confidence–accuracy relationship as a function of age from adolescence to older adulthood.

Decision Time–Accuracy Relationship across the Lifespan

Theoretically, face recognition is considered a holistic process, which occurs rather automatically, even after very short exposures (Richler, Mack, Gauthier, & Palmeri, 2009). Self-reports of witnesses who claim that a face just “popped out” from a lineup and who are more likely to be correct than slow responders (e.g. Brewer & Weber, 2008; Sporer, 1992) support this notion. On the other hand, witnesses deliberating and engaging in a relative decision strategy spend more time on taking a decision and are more likely to be wrong (Sauerland & Sporer, 2007). Applied to lineup decisions of choosers, it can be predicted that decision times (response latencies) for hits are faster than for false alarms (Dunning & Stern, 1994; Sporer, 1994). By now, quite a few studies have shown that decision time – a more objective indicator than confidence – is associated with identification accuracy, that is, eyewitnesses made correct identifications faster than incorrect identifications (e.g. Brewer & Weber, 2008; Dunning & Perretta, 2002; Sporer, 1992, 1993, 1994). This is also in line with the dual-process model, which proposes that accurate decisions are made rapidly and automatically, and require no cognitive effort, while inaccurate decisions are time-consuming, conscious, and deliberate (Dunning & Perretta, 2002; Dunning & Stern, 1994). Similar tendencies have been shown for self-reported estimated decision times, namely that correct identifications were estimated by witnesses to be faster than false identifications (Sauerland & Sporer, 2009).

The question arises whether or not measured decision-times as well as self-reported decision times show the same stable relationship with identification accuracy across the lifespan. In other words, do older adults encode and retrieve faces differently and do they use different metacognitive strategies than younger adults? Sporer (2001) has proposed an in-group/out-group model which postulates that in-group faces, for example, of one’s own ethnic group (“race”), gender, and age are processed relatively more holistically than faces of out-group members. Regarding age, a recent meta-analysis of face recognition studies comparing younger and older adults has found support for this prediction (Martschuk & Sporer, 2018). Considering that all studies reported above used only young targets in their research, the findings about decision times and their association with identification accuracy may be limited to young witnesses observing young targets.

In the light of the recent demographic developments which entail that more and more witnesses will be older it is an important question whether older witnesses observing young targets show the same pattern. To the extent that older adults do not engage in the type of holistic processing mentioned, or employ different metacognitive strategies than younger adults, this association may be weakened in older age. Some insight into lineup decision processes of different age groups were provided by Colloff et al. (2017, supplemental material) in their additional analyses. Particularly, older adults took longer to make a lineup decision (i.e., to identify or reject a lineup) and to render a confidence judgment than middle-aged and young adults. These analyses, however, were only conducted *across* choosers and non-choosers. It remains an open question whether these relationships will be stronger when focusing only on choosers who are particularly important in criminal cases.

Aims and Hypotheses

The aim of this study was to assess the confidence–accuracy relationship for different age groups across the lifespan. We predicted that the confidence–accuracy relationship deteriorates with increasing age. More specifically, we expected accuracy and confidence to dissociate across the lifespan, with older participants being relatively more overconfident, leading to poorer calibration (Hypothesis 1). To investigate the postdictive value of confidence as a function of age at different prior base rates of target-presence (Sauerland, Sagana, Sporer, & Wixted, 2018; Wells, Yang, & Smalarz, 2015; Wixted & Wells, 2017), we also conducted Bayesian analyses. While most lab studies employ a base rate of 50% target-present lineups and 50% target-absent lineups, presumably, base rates vary considerably across different legislations and police departments. Importantly, the postdictive value of confidence following a lineup decision varies as a factor of this prior base rate (Wixted & Wells, 2017), with higher target-presence base rates being associated with higher postdictive value compared to lower target-presence base rates (Hypothesis 2). In addition to replicating this finding, we expected the postdictive value of confidence to decrease as age increases. Finally, as proposed with the dual-process model of the decision process, we expected that correct identifications would be made faster and would be estimated to be faster than false identifications. To the extent that older participants rely less on holistic processing or differ in their metacognitive strategies, they may show lower decision time–accuracy associations than younger adults (Hypothesis 3). Considering the age-related decline of metacognitive monitoring and control, we assumed the self-reported decision time–accuracy associations to be similarly or even more impaired in older age (Hypothesis 4).

Method

Participants

This study comprised a re-analysis of data collected by Sauerland and Sporer (2009) with 720 citizens (50% female, 50% male) who agreed to participate in the experiment. Four hundred and thirty-six participants made a positive identification decision (called “choosers”), whereas 284 rejected the lineup.

In the present study only choosers (221 male, 215 female) were included, aged between 15 and 83 years ($M = 37.63$, $SD = 15.71$, $Mdn = 34.5$). Participants were categorized into six age groups: 15–20 years ($n = 44$); 21–30 ($n = 149$); 31–40 ($n = 74$); 41–50 ($n = 58$); 51–60 ($n = 64$); and 61–83 ($n = 47$). The proportions of choosers did not differ as a function of age group: 15–20 years: 60.3%; 21–30: 59.8%; 31–40: 59.2%; 41–50: 53.2%; 51–60: 66.0%; 61–83: 70.1%; $\chi^2(5, N = 720) = 5.39$, $p = .270$, Cramer’s $V = .27$.

The majority of the participants were employees in the academic (16.3%) or non-academic field (38.6%), 27.5% were students, 7.6% were retired, 5.3% attended school, and 4.9% were unemployed.

Design

The study was a 2 (target-present vs absent) x 6 (age groups 1 to 6) between-participants design. Dependent measures were identification accuracy and post-decision confidence. We also report some results on decision times (in s) and self-reported decision times.

Procedure

Ten different targets aged between 20 and 37 years were used to fulfil the requirement of stimulus sampling (Wells & Windschitl, 1999). Targets recruited participants in pairs, while one had the role of the target and the other of the research confederate. Each target was instructed to approach male and female participants from different age groups, including old-age people, so targets would be distributed approximately equally across gender and age groups of participants. Chi-square analyses confirmed that targets were equally distributed among age groups (all participants: $\chi^2(5, N = 720) = 52.58, p = .204$, Cramer's $V = .270$; choosers: $\chi^2(5, n = 436) = 50.31, p = .271$, Cramer's $V = .34$).

Following a predetermined script, the target approached a pedestrian on the street and asked for directions to a certain location nearby. After the target walked away, a research confederate approached the pedestrian and explained the true purpose of the interaction (a face recognition study). Once participants agreed, they were randomly assigned to a target-present or a target-absent photo lineup. Unbiased lineup instructions were used, and the research confederate was blind to target presence. Immediately after the identification task participants provided their confidence on an 11-point rating scale (0% to 100%, at 10% intervals) and estimated the time they spent on making their decision on a paper-and-pencil answer sheet. The confederate inconspicuously stopped the time when the participant entered her or his response to the lineup task.

Lineups

Ten target-present and ten target-absent photo lineups were prepared for the study. Each lineup contained six frontal 9 x 13 cm photographs that were arranged in two rows of three pictures. The target or target replacement was always placed on Position 3. The lineups were constructed following a pilot study with 55 mock witnesses that determined which portrait photographs fit the general target description (Wells, Rydell, & Seelau, 1993). Those who received most selections were used as foils or replacements.

Analyses

The present study used different types of calibration analyses following suggestions to apply alternative analyses, which are more diagnostic of the confidence–accuracy relationship than point-biserial correlations (e.g., Juslin, Olsson, & Winman, 1996; Olsson, 2000; Weber & Brewer, 2003). Calibration curves plot confidence level against the proportion of witnesses who made correct decisions. However, as in the present study sample sizes for the separate age groups were too small for traditional calibration curves, we relied on graphic comparisons of mean accuracy and confidence levels using “confidence–accuracy characteristic” (CAC) analysis (Mickes, 2015), along with 95% confidence intervals, and on point-biserial correlations between confidence and accuracy. To assess curvi-linear relationships as a function of age, we calculated correlations for age and age-squared (age^2).

Following the categorizations used by Wixted and colleagues in their recent publications on low, medium and high confidence, we re-categorized the confidence values of the original 0% to 100% scale into low confidence (0–50%; $n = 191$), medium confidence (60–80%; $n = 130$), and high confidence (90–100%; $n = 115$).

Further statistical calibration analyses were included as follows. *Calibration (C)* is the extent to which an eyewitness's confidence or certainty judgment (measured in %) corresponds with the probability that the decision is correct (from $0 = \text{perfect calibration}$ to $1 = \text{weak calibration}$). *Over-/underconfidence*

(*O/U*) describes participants' tendency to be over- (positive score) or underconfident (negative score) in comparison to their accuracy. The *adjusted normalized resolution index (ANRI)* indicates the discrimination of confidence judgments (from 0 = *no discrimination* to 1 = *perfect discrimination*) and can be interpreted like the effect size η^2 (*eta*-squared). Hence, ANRI values corresponding to η^2 values of .010, .059, and .138 indicate small, moderate, and large effects (Cohen, 1988). It is important to consider different measures as good calibration does not imply good discrimination (Yaniv, Yates, & Smith, 1991). For a detailed description of the formulae used, see Brewer and Wells (2006), and Yaniv et al. (1991). Finally, the sensitivity measure d' was calculated from hits and false identifications, which describes discriminability between targets and non-targets (Macmillan & Creelman, 2005). To calculate d' , the false identification rate was divided by lineup size. Note, that the results for d' were derived from the entire data set, including non-choosers.

Results

Identification Accuracy and Confidence across Age Groups

Mean accuracy and mean confidence across the age groups

To test Hypothesis 1 that the confidence–accuracy relationship would deteriorate with increasing age, we calculated how confidence and accuracy relate to each other as a function of age group. Overall, 51.6% of choosers made a correct identification, whereas 48.8% made a false identification. The mean confidence was $M = 63.07$ ($SD = 26.47$), ranging between 0% and 100% (none was 10% confident). Chi-square analyses revealed that the effect of age group on identification accuracy was non-significant, $\chi^2(5, 436) = 10.58$, $p = .060$, Cramer's $V = .16$. Notably, the performance measure d' showed a strong increase in discriminability performance from adolescence ($d' = 1.48$) to young adulthood ($d' = 1.91$), followed by a steady decrease until older adulthood ($d' = 1.35$) as shown in Table 1.

Table 1. Calibration (C), Over-/Underconfidence (O/U), Adjusted Normalized Resolution Index (ANRI), their Respective 5% Inferential Confidence Intervals (ICI), Recognition Performance (d'), and Point-biserial Correlations (r_{pb}) of Choosers as a Function of Age Group

Age group	<i>N</i>	<i>C</i> [ICI]	<i>O/U</i> [ICI]	<i>ANRI</i> [ICI]	d'	r_{pb}
15–20 years	44	.057 [.017, .098]	.211 [.125, .297]	.315 [.085, .546]	1.48	.56**
21–30 years	149	.014 [.001, .027]	.066 [.015, .117]	.231 [.148, .315]	1.91	.45**
31–40 years	74	.061 [.018, .104]	.159 [.075, .239]	.022 [.000, .088]	1.80	.27*
41–50 years	58	.072 [.023, .122]	.188 [.092, .284]	.068 [.000, .178]	1.56	.18
51–60 years	64	.065 [.018, .111]	.173 [.076, .270]	.022 [.000, .094]	1.36	.12
61–83 years	47	.093 [.031, .155]	.242 [.145, .339]	.079 [.000, .211]	1.35	.34*
All Choosers	436	.030 [.020, .044]	.147 [.115, .180]	.109 [.064, .153]	1.66	.34**

* $p < .05$, ** $p < .001$.

Note. Confidence values of the original 0% to 100% scale were re-categorized as follows: Low Confidence = 0–50 ($n = 191$); Medium Confidence = 60–80 ($n = 130$); High Confidence = 90–100 ($n = 115$). The point-biserial correlations r_{pb} are for accuracy correlated with the original 0–100% scale. Because ANRI can be interpreted as η^2 , and η^2 is directly related to Cohen's f , cutoffs for small, moderate, and large ANRI values can be calculated using the .10, .25, and .40 cutoffs for f . The respective ANRI values are .010, .059, and .138.

Correlational analyses showed that while age and confidence were not related with each other, $r(434) = -.01, p = .883$; $\text{age}^2: r(434) = -.02, p = .748$, the relationship between age and identification accuracy was curvilinear, $r_{pb}(434) = -.09, p = .071$; $\text{age}^2: r_{pb}(434) = -.10, p = .039$, showing an increase in identification accuracy from adolescence to young adulthood, followed by a decrease until older adulthood. Figure 1 displays the mean accuracies in the six age groups and the associated levels of confidence. Mean confidence varied little across the six age groups while identification accuracy showed a small increase followed by a decrease with increasing age.

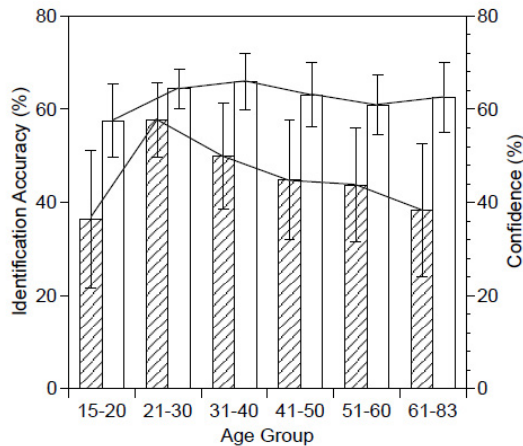


Figure 1. Mean confidence (%) and identification accuracy (%) of choosers as a function of age group. Error bars are 95% confidence intervals.

Across all age groups, confidence was significantly associated with identification accuracy, $\chi^2(2, N = 436) = 49.96, p < .001$, Cramer's $V = .33$. However, the relationships were not consistent across age groups. With increasing age, confidence was less indicative of identification accuracy: The older the participants, the smaller the difference in confidence judgments between accurate and inaccurate responses. Figure 2 shows the regression lines of correct and incorrect responses and their associated 95% confidence bands, which crossed at the age of 55 years, indicating that adults were less likely to distinguish correct and incorrect identifications at the age of 55 years or more.

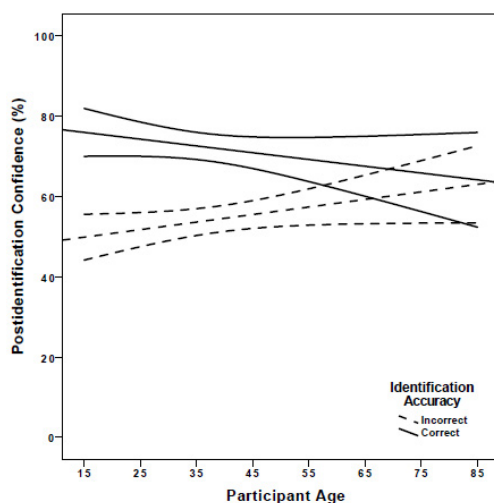


Figure 2. Confidence for correct and incorrect identification responses across the lifespan.

Confidence–accuracy characteristic analysis for each age group

To further explore the confidence–accuracy relationship across the lifespan (Hypothesis 1), we conducted a CAC analysis, a graphic presentation which resembles the calibration curve. Figure 3 displays the mean identification accuracies for each age group with their respective 95% CIs at the three levels of confidence (low, medium, high). Values along the diagonal indicate well-calibrated witnesses, values above the line – underconfidence, and values below the diagonal – overconfidence³. As shown in Figure 3, high confidence was diagnostic of accuracy for only the two youngest age groups. Although low confidence was predictive of poor identification accuracy across all groups, with increasing age medium and high confidence became less indicative of identification accuracy. In particular, high confidence was diagnostic of accuracy in the 21–30 years age group only, and to some extent in the 15–20 years age group.

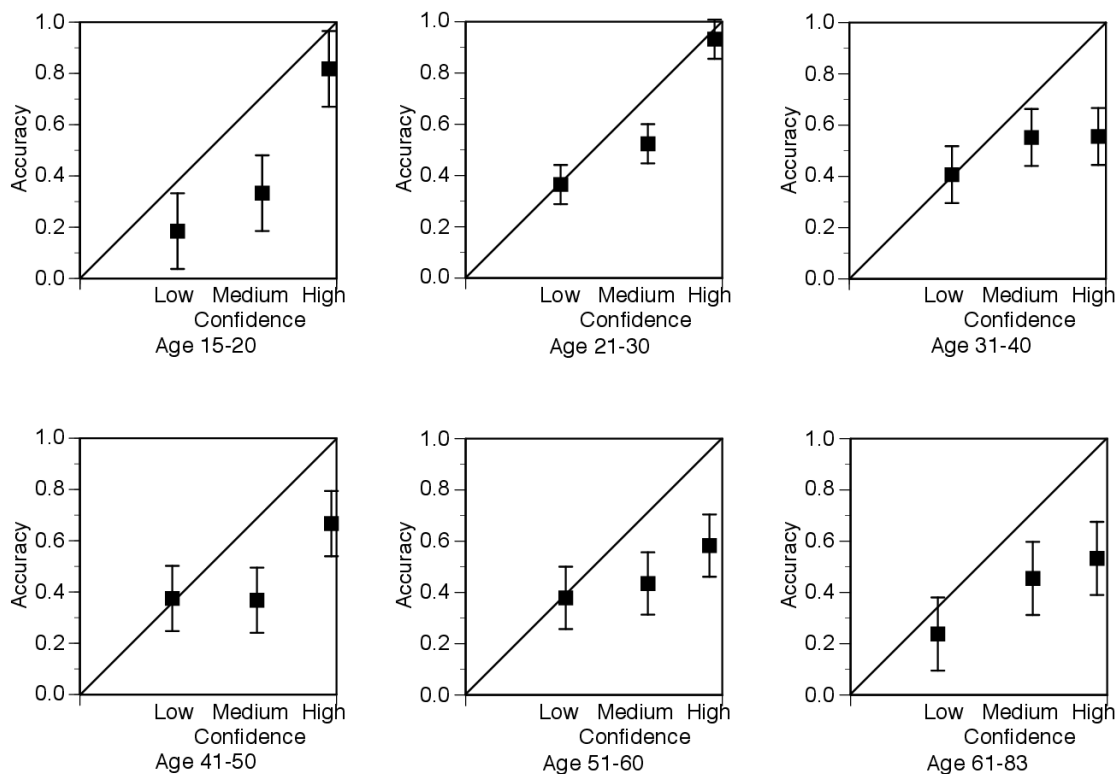


Figure 3. Confidence–accuracy characteristic analysis for choosers as a function of age group for low (0–50%), medium (60–80%), and highly (90–100%) confident witnesses. Error bars are 95% confidence intervals. Labels on the X-axis are placed under the means of the respective confidence range.

Calibration, over-/underconfidence and the discrimination of confidence judgments

To test Hypothesis 1 statistically, in addition to the graphic representation, is to assess the statistics *Calibration*, *Over-/underconfidence*, and *Adjusted normalized resolution index* (Brewer & Wells, 2006). These indices were calculated by applying a bootstrap method (Efron, 1981; Efron & Gong, 1983) with $n = 1000$ replications to correct for a potential bias and to assess standard errors (SE) for the calculation of inferential

³ Error bars are two-sided 95% confidence intervals calculated for proportions using an algorithm and spreadsheet developed by Newcombe (1998). We are grateful to Colin E. Tredoux for making us aware of this publication and sending us the spreadsheet.

confidence intervals (ICI; Tryon, 2001). The bootstrap analyses were conducted for each age group separately to calculate separate SE for them (Efron & Gong, 1983). The reduction factor E , defined as “the ratio of the SE of the difference between the two groups to the sum of the standard errors of both groups” (Tryon, 2001, p. 375), was averaged across all pairwise comparisons between the age groups. Non-overlapping ICIs showed significant differences between the age groups at the $\alpha = .05$ level. A similar approach was applied by Palmer, Brewer, and Weber (2010) who tested the confidence–accuracy relationship of witnesses following post-identification feedback.

Results for calibration, over-/underconfidence and the *ANRI*, and their respective 5% ICIs are displayed in Table 1. These indices support the aforementioned findings: Although all age groups were overconfident for low and medium confidence (see Figure 2), calibration analyses revealed differences between the groups.

Young adults (21–30 years old) were well calibrated ($C = .01$, 5% ICI [.00, .03]); they showed, on average, very little overconfidence ($O/U = .07$, 5% ICI [.02, .12]), and their discrimination was above the average of the entire sample ($ANRI = .23$, 5% ICI [.15, .32]). By contrast, with increasing age, average confidence–accuracy correlations, calibration and discrimination decreased whereas overconfidence increased. In fact, adults older than 60 years had significantly weaker calibration and significantly higher overconfidence ($C = .09$, 5% ICI [.03, .16]; $O/U = .24$, 5% ICI [.15, .34]) than adults aged between 21 and 30 years. Notably, the results for the youngest age group (15–20 years) varied across the measures: Although they showed the highest discrimination and correlation of all age groups ($ANRI = .32$, 5% ICI [.09, .55], $r_{pb} = .56$), their calibration was weaker than the calibration of 21 to 30 year olds (15–20 years: $C = .06$, 5% ICI [.12, .10]; 21–30 years: $C = .01$, 5% ICI [.00, .03]), and their degree of overconfidence was significantly higher than that of young adults, thus comparable to that of old-age participants (15–20 years: $O/U = .21$, 5% ICI [.13, .30]; 61–83 years: $O/U = .24$, 5% ICI [.15, .34]).

Interestingly, all age groups older than 31 years showed weak to no discrimination as indicated by the inferential confidence intervals that included zero (see Table 1). Furthermore, discrimination of young adults aged from 21 to 30 years was significantly higher than the discrimination of 31 to 40 year old and 51 to 60 year old adults (21–30 years: $ANRI = .23$, 5% ICI [.15, .32]; 31–40 years: $ANRI = .02$, 5% ICI [.00, .09]; 51–60 years: $ANRI = .02$, 5% ICI [.00, .09]).

Base Rates of Target-Presence

Next, we addressed the possible concern that the moderation of the confidence–accuracy relationship by age may hold only for certain base rates or prior probabilities that the suspect is the perpetrator (Hypothesis 2). Figure 4 maps the probability that a suspect identification was accurate (i.e., that the suspect is the perpetrator) across all possible base rate values from 0% (none of the lineups included a guilty suspect) to 100% (all lineups included a guilty suspect; see Wixted & Wells, 2017). One curve was created for each age group. The identity line shows where the data would fall if an identification was non-diagnostic.

The Figure 4 shows that all curves are above the identity line, demonstrating that identifications were diagnostic of guilt. Generally, as expected, the heights of the curves for confident and younger participants are far above those of the curves for less confident or older participants. Strikingly, less confident decisions made by 21–30 year olds were still more diagnostic of guilt than highly confident decisions made by 51–83 year olds. In all, the Figure suggests that the moderating effect of age on the confidence–accuracy relationship prevails across different prior base rates of target-presence.

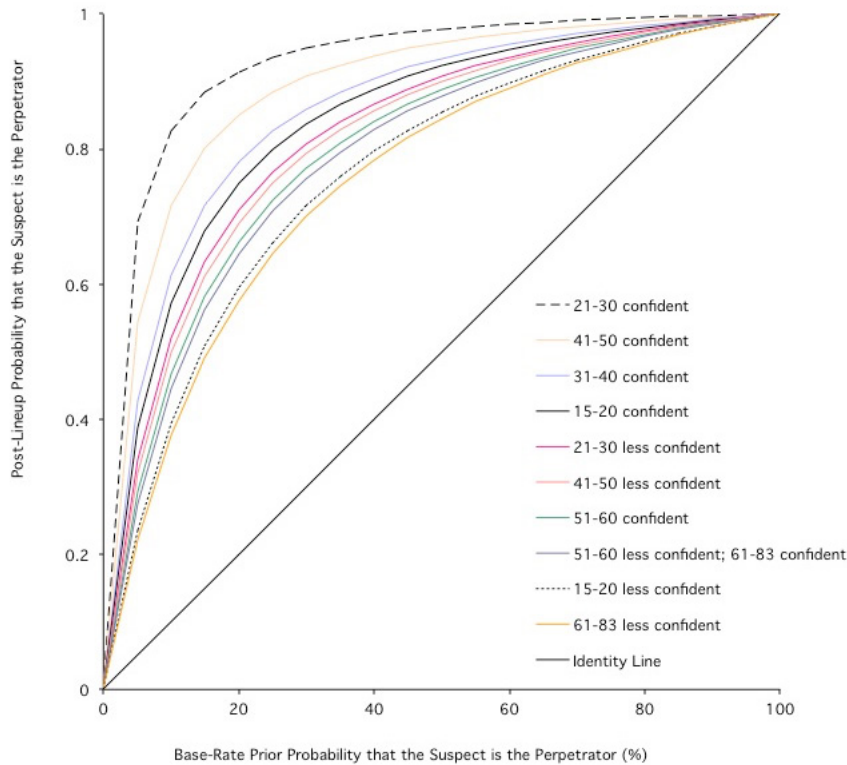


Figure 4. Post-lineup probability that the suspect is the perpetrator as a function of the base rate of target-presence for different age groups.

Observed and Self-reported Decision Time and Accuracy across the Lifespan

Observed decision times and self-estimated decision times showed positive skewness. Followingly, inferential analyses were conducted with log-transformed data (log-base 10) to approximate normal distributions. Means, SDs, and 95% CIs are reported for back-transformed values.

Analyses revealed that participants' age was not correlated with observed decision time, $r(434) = .08$, $p = .060$ (age²: $r(434) = .08$, $p = .112$). However, there was a small positive correlation between age and self-estimated decision time: The older the participants, the longer the self-estimated time to identify a person from the lineup, $r(434) = .16$, $p = .001$ (age²: $r(434) = .14$, $p = .006$).

As predicted by Hypothesis 3, the decision time was negatively correlated with the identification accuracy, $r_{pb}(434) = -.30$, $p < .001$, indicating that participants made a correct identification faster ($M = 7.05$ seconds, $SD = 2.62$, $CI [6.07; 8.17]$) than those who made a false identification ($M = 12.27$ seconds, $SD = 2.27$, $95\% CI [10.86; 13.87]$; $t(413) = 6.46$, $p < .001$). This was constant across all age groups, as shown in Figure 5a. By contrast, although the self-estimated decision time was generally shorter for correct identifications ($M = 7.35$ seconds, $SD = 3.02$, $95\% CI [6.19; 8.72]$) than for false identifications ($M = 11.59$ seconds, $SD = 2.75$, $95\% CI [9.97; 13.48]$; $t(434) = 4.48$, $p < .001$), this varied by age, supporting Hypothesis 4. Particularly adults aged between 31 and 40 years and older than 50 years overestimated their decision time when making a correct identification, and adolescents overestimated their decision time for both correct and false identifications, as shown in Figure 5b.

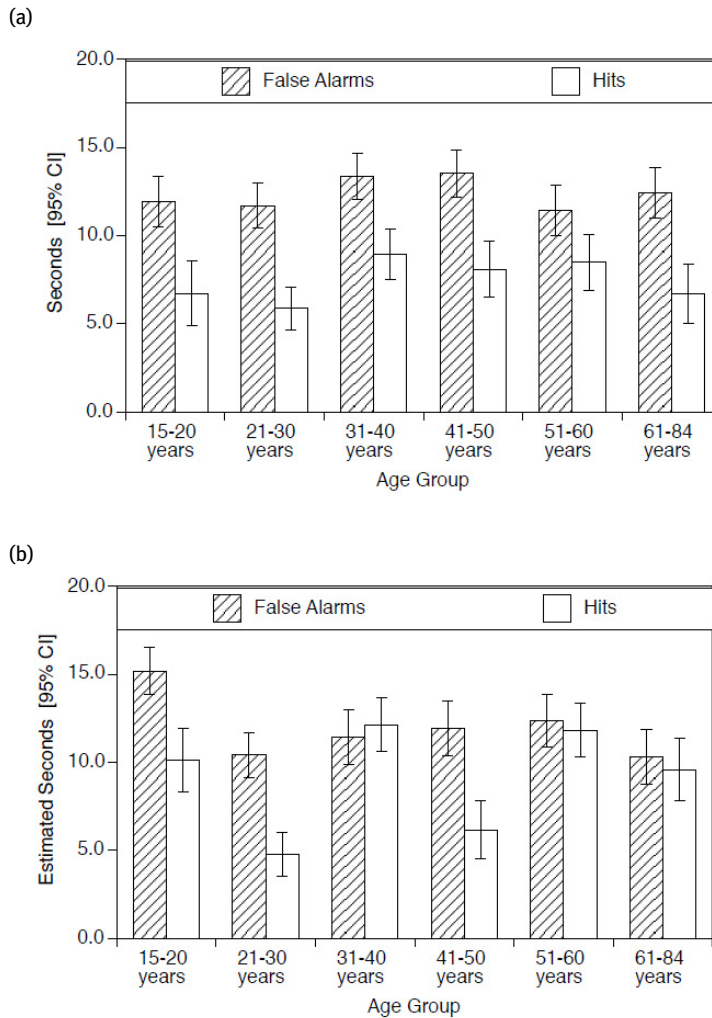


Figure 5. Decision time (a) and self-reported decision time (b) for correct and false identifications as a function of age group.

Combined Postdictive Value of Age, Confidence and Decision Time

Logistic regression analyses were conducted to combine the postdictors age, observed decision time, self-reported decision time, and postdecision confidence. The aim was to assess their respective associations with identification accuracy while controlling for the other postdictors. Adolescents were excluded from these analyses to control for curvi-linearity, which is difficult to detect in regression analyses.

First, a logistic regression analysis was conducted with the postdictors age, observed decision time, and accuracy. The full model was statistically significant, $\chi^2(3, 392) = 64.80, p < .001$, and explained 15.2% (Cox & Snell R^2) and 20.3% (Nagelkerke R^2) of the variance, respectively. Participants' age and decision time were negatively associated with identification accuracy while confidence was positively associated with it, as shown in Table 2a. The postdictive value was the strongest for decision time while controlling for age: Participants were four times more likely to correctly identify a person when the identification was fast ($OR = 0.26, 95\% CI [0.15, 0.46]$; reciprocal $OR = 3.85, 95\% CI [2.16, 6.85], p < .001$). Similarly, participants with higher confidence were almost twice as likely to correctly identify a person than those with lower confidence ($OR = 1.93, 95\% CI [1.46, 2.54], p < .001$). Age remained a significant postdictor of identification accuracy showing a small but significant effect: The older the participants, the less likely they were to correctly identify the target person ($OR = 0.98, 95\% CI [0.97, 0.99], p = .018$).

Table 2. Logistic Regression of (a) Age, Observed Decision Time and Confidence on Identification Accuracy, and (b) Age, Self-reported Decision Time and Confidence on Identification Accuracy

	<i>B</i>	<i>SE</i>	Wald's χ^2 (1)	<i>p</i>	<i>OR</i>	<i>95% CI OR</i>
(a)						
Constant	0.789	0.499	2.497	.114	2.201	
Age	-0.018	0.007	5.572	.018	0.983	[0.968, 0.997]
Observed decision time	-1.346	0.294	20.952	<.001	0.260	[0.146, 0.463]
Confidence	0.656	0.140	21.898	<.001	1.928	[1.464, 2.538]
(b)						
Constant	-0.124	0.459	0.073	.787	0.883	
Age	-0.017	0.007	5.572	.018	0.983	[0.969, 0.997]
Self-reported decision time	-0.490	0.240	4.171	.041	0.613	[0.383, 0.980]
Confidence	0.688	0.140	24.089	<.001	1.990	[1.512, 2.620]

Separate analyses were conducted to assess the postdictive value of self-reported decision time on identification accuracy while controlling for age and confidence. The model was statistically significant, $\chi^2(3, 392) = 46.39, p < .001$, and explained 11.2% (Cox & Snell R^2) and 14.9% (Nagelkerke R^2) of the variance, respectively. Self-reported decision time was negatively associated with identification confidence, showing a small but significant effect ($OR = 0.61, 95\% CI [0.38, 0.98]$; reciprocal $OR = 1.63, 95\% CI [1.02, 2.61], p = .041$). Age and confidence remained significant postdictors with little variability from the model specified above. The full model details are in Table 2b.

Notably, when observed and self-reported decision times were combined in the same logistic regression model analyses, self-reported decision time did not contribute to the improvement of the model in Table 2a, that is, it did not explain any additional variance in the model (adding 0.0% to the Cox and Snell R^2 , and 0.0% to Nagelke R^2 , respectively).

Discussion

For more than three decades, studies have assessed the postdictive value of eyewitness confidence following a suspect identification. While earlier meta-analyses concluded that confidence is an unreliable indicator of identification accuracy (Bothwell et al., 1987), a focus on choosers led to more optimistic conclusions (Sporer et al., 1995). More recent re-analyses of previously published data further emphasized the postdictive value of confidence, particularly for high confidence levels and under “pristine” identification conditions (Wixted et al., 2015; Wixted & Wells, 2017). These analyses, however, did not consider some potentially important moderators like race of target and witness (Wright, Boyd, & Tredoux, 2001), or old age of a witness. The present study aimed to fill this gap by analyzing the confidence–accuracy relationship as a function of witness age.

Our re-analyses of a large-scale field study (Sauerland & Sporer, 2009) revealed that age was a potentially important moderator of the confidence–accuracy relationship. As expected, the confidence–accuracy relationship was strong for younger participants (15–20 years) and young adults (21–30 years) and declined as age increased, confirming Hypothesis 1. This difference prevailed across different prior base rates of target-presence (Hypothesis 2), although the differences between age groups were strongest for small to moderate prior base rates of target-presence. Further, correct identifications were observed to be faster regardless of participants’ age. However, the relationship between self-reported decision time and accuracy varied by age, showing little discrimination between correct and false identifications in older age (Hypothesis 3).

Similar to the studies re-analyzed in Wixted et al. (2015), the present study was an experimentally controlled field study. The identification instructions were fair (i.e. the participants were informed that the person might or might not be present in the lineup, and the interviewer was blind to target-presence and position in the lineup). Thus, our study fulfilled what Wixted and Wells (2017) described as pristine conditions. Moreover, participants did not know that they were part of a study until after the target walked away (unlike e.g. in Palmer et al., 2013). The latter factor is important for the legal system, thus enhances the ecological validity of our findings as in many criminal cases eyewitnesses do not anticipate that they will have to identify the person while they witness an event unless it is obvious from the beginning that they witness a crime. In support of this point, prior research showed that identification performance was poorer under incidental rather than intentional encoding conditions (Read, Lindsay, & Nicholls, 1998; Sporer, 1991).

Figure 3 shows that only the youngest two age groups were reasonably well calibrated. Particularly the results of young adults (21–30 years old) are in line with the findings by Wixted et al. (2015), namely that high confidence is indicative of identification accuracy (93.2% correct identification at high confidence level). This age group performed well and above the average of the entire sample on measures of calibration, over-/underconfidence, and the discrimination indices ANRI and the SDT performance measure d' (see Table 1). By contrast, with increasing age, confidence–accuracy correlations and calibration decreased whereas overconfidence increased. Moreover, ANRI discrimination was nearly zero for participants who were older than 30 years suggesting that the age-related decline may begin much earlier than at old age. Discrimination between targets and non-targets (d') decreased with increasing age as has been shown across many studies (see the recent meta-analysis by Martschuk & Sporer, 2018). Taken together, age had a robust effect on every measure of identification performance while confidence remained unchanged across the lifespan.

The analyses in the present study differed from previous studies in the following aspects. Instead of averaging across large data sets we separated the age groups and showed a curvi-linear relationship between age and the postdictive value of the confidence–accuracy relationship. Further, the present study focused on choosers – those who would end up testifying in court. Although Key et al. (2015) analyzed the confidence–accuracy relationship between young, middle-aged and older adults, their results were not directly comparable to ours because the point-biserial correlations they reported were not separately calculated for choosers and non-choosers. As already pointed out in the introduction, the study by Colloff et al. (2017) differed in many ways from the current work. In particular the online methodology may have led to a self-selection bias in the older participant groups. Further, many analyses were conducted across choosers and non-choosers in Colloff et al. (2017) thus not making them directly comparable.

Decision time has often been found to be negatively associated with confidence and identification accuracy among choosers (e.g., Brewer & Weber, 2008; Sporer, 1992; Sauerland & Sporer, 2007, 2009). As postulated by the dual-process model, correct decisions are automatic and require little cognitive effort whereas incorrect decisions follow a conscious and time-consuming process (Dunning & Perretta, 2002; Dunning & Stern, 1994). The present findings confirmed that shorter decision times were associated with correct identifications, regardless of age. Moreover, the objectively measured decision time was a much stronger postdictor of identification accuracy (reciprocal $OR = 3.85$) than confidence ($OR = 1.93$). By contrast, self-reported decision time did not improve the postdictive value of observed decision time. It was, however, a reliable but less strong postdictor (reciprocal $OR = 1.63$) when it was substituted for measured decision time. In real cases, where decision times may not have been measured, self-reported decision time may be informative (but perhaps not with older witnesses over 50).

This further supports our theoretical assumptions that metacognitive processes vary by age. These findings demonstrate the importance of recording the entire identification procedure on video to not only capture the expressed confidence but also to be able to assess decision-time at the time of the original identification (Sporer, 1992, 1994). Confidence and self-reported decision times assessed retrospectively at later interviews or even in court may not only be affected by metacognitive deficits as a function of delay but, more troublesome, distorted by feedback effects (Stebly, Wells, & Douglass, 2014).

Limitations and Implications

The present study faced the following limitations. The participants in this study were not witnesses of a crime; instead, they were instructed to identify a person who approached them to ask for directions. Hence, the participants did not experience the level of stress that people might experience when they witness a crime. Nevertheless, the present findings are comparable to many real-life criminal cases, as the participants did not know that the person who approached them was a target until after the person walked away. Witnesses of crimes frequently do not anticipate that they will have to identify a person. Further, the scenario in the present study resembles a police call-out for people who saw or spoke to a person at a particular time and place, with the difference of a much shorter retention interval than in most criminal cases.

A further limitation of our study was the relatively small sample size per age group. Calibration curves require at least 200 participants or multiple judgments by each participant per condition to produce reliable results (Weber & Brewer, 2003). However, averaging across large data sets can obscure the results when moderators are not considered. To overcome this problem and approximate the calibration curve, we plotted the relationship between confidence and accuracy at fewer levels of confidence (low, medium, high). In addition, we applied different types of statistical analyses (as proposed Brewer & Wells, 2006). To obtain more reliable estimates of inferential confidence intervals using bootstrap analyses were conducted. This allowed the examination of the confidence–accuracy relationship from different perspectives. Our results (particularly the youngest age group) demonstrated the necessity of applying different types of analyses that yielded different coefficients that are not redundant. The calibration graph of adolescents showed robust calibration and a discrimination index that was the highest of all age groups. Nevertheless, their identification accuracy was as low (see Figure 3 above) and their overconfidence score was as high as that of people aged 61 years and over, attributable to the high confidence level despite low identification accuracy.

Further research is needed to assess the postdictive value of confidence of other samples of older people as well as of children, teenagers, or vulnerable witnesses. In addition, procedures should be explored that can improve the metacognitive processes at retrieval. For instance, Thomas, Gordon, and Bulevich (2014) suggested that metacognitive processes could be improved with supportive instructions that guide conscious processing of retrieved information, which results in more stringent metamemorial processes.

Finally, the targets in the present study were young to middle-aged adults. Previous research repeatedly demonstrated a robust effect of an own-age bias in eyewitness identification and face recognition (Kocab et al., 2018; Martschuk & Sporer, 2018; Rhodes & Anastasi, 2012). Thus it is possible that the age effect decreased if older targets were likewise included. Future research should assess whether this effect holds for older targets.

Conclusion

The results demonstrated that objective decision time had a greater postdictive value of identification accuracy than confidence and that the confidence–accuracy relationship depended on witness age. Whereas the high confidence expressed by young adults was indeed a reliable indicator of identification accuracy, the increased overconfidence of adolescents, middle-aged and older people, implies that we should take age and perhaps also the difference in age between the target and witness into account when postdicting accuracy from confidence.

Using age as a moderator, this study showed that factors which negatively influence identification accuracy do not necessarily affect post-decision confidence in the same way, leading to a decrease in the confidence–accuracy relationship. Thus, when evaluating eyewitness identification and post-decision confidence, decision-makers should be aware of factors that moderate identification accuracy, such as the cross-race or the old-age effect. Other factors might yield similar patterns and should be explored further. Finally, to better assess witness responses and how their postdictive value is evaluated by factfinders

(Kaminski & Sporer, 2017), a lineup decision should be videotaped, which also provides information on decision times (response latencies), as suggested by Sporer (1992; Sporer et al., 1995) and many other researchers thereafter (e.g., Wixted & Wells, 2017). If, however, objective decision time is not available, the self-reported decision time together with confidence can assist in the evaluation of the likelihood of a correct identification, at least for young adults.

References

- Bartlett, J. C., & Memon, A. (2007). Eyewitness memory in young and older eyewitnesses. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *The handbook of eyewitness psychology, Vol. 2. Memory for people* (pp. 309–338). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Bothwell, R. K., Deffenbacher, K. A., & Brigham, J. C. (1987). An exploratory study of personality differences in eyewitness memory. *Journal of Applied Psychology, 72*, 691–695. doi:0021-9010/87
- Brewer, N., & Weber, N. (2008). Eyewitness confidence and latency: Indices of memory processes not just markers of accuracy. *Applied Cognitive Psychology, 22*, 827–840. doi:10.1002/acp.1486
- Brewer, N., & Wells, G. (2006). The confidence–accuracy relation 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/a0031602
- Buckhout, R. (1974). Eyewitness testimony. *Scientific American, 231*, 23–31. doi:10.1038/scientificamerican1274-23
- Busey, T. A., Tunnicliff, J., Loftus, G. R., & Loftus, E. F. (2000). Accounts of the confidence–accuracy relation in recognition memory. *Psychonomic Bulletin & Review, 7*, 26–48. doi:10.3758/BF03210724
- Castel, A. D., McGillivray, S., & Friedman, M. C. (2012). Meta-memory and memory efficiency in older adults: Learning about the benefits of priority processing and value-directed remembering. In M. Naveh-Benjamin & N. Ohta (Eds.), *Memory and aging: Current issues and future directions* (pp. 245–270). New York: Psychology Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Routledge Academic.
- Colloff, M. F., Wade, K. A., Wixted, J. T., & Maylor, E. A. (2017). A signal-detection analysis of eyewitness identification across the adult lifespan. *Psychology and Aging, 32*, 243–258. doi:10.1037/pag0000168
- Cutler, B. L., & Penrod, S. T. (1989). Forensically relevant moderators of the relation between eyewitness identification accuracy and confidence. *Journal of Applied Psychology, 74*, 650–652. doi:10.1037/0021-9010.74.4.650
- DeSoto, K. A., & Roediger, H. L. (2014). Positive and negative correlations between confidence and accuracy for the same events in recognition of categorized lists. *Psychological Science, 25*, 781–788. doi:10.1177/0956797613516149
- Dodson, C. S., Bawa, S., & Krueger, L. E. (2007). Aging, metamemory, and high-confidence errors: A misrecollection account. *Psychology & Aging, 22*, 122–133. doi:10.1037/0882-7974.22.1.122
- Dodson, C. S., & Krueger, L. E. (2006). I misremember it well: Why older adults are unreliable eyewitnesses. *Psychological Bulletin & Review, 13*, 770–775. doi:10.3758/BF03193995
- Dunlosky, J., & Tauber, S. K. (2014). Understanding people’s metacognitive judgments: An isomechanism framework and its implications for applied and theoretical research. In T. Perfect & D. S. Lindsay (Eds.), *Sage handbook of applied memory* (pp. 444–464). London, England: Sage Publications.
- Dunning, D., & Perretta, S. (2002). Automaticity and eyewitness accuracy: A 10-to 12-second rule for distinguishing accurate from inaccurate positive identification. *Journal of Applied Psychology, 87*, 951–962. doi:10.1037/0021-9010.87.5.951
- Dunning, D., & Stern, L. B. (1994). Distinguishing accurate from inaccurate eyewitness identifications via inquiries about decision processes. *Journal of Personality and Social Psychology, 67*(5), 818–835. doi:10.1037/0022-3514.67.5.818
- Efron, B. (1981). Nonparametric estimates of standard error: The jackknife, the bootstrap, and other methods. *Biometrika, 68*, 589–599. doi:10.1093/biomet/68.3.589
- Efron, B., & Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician, 37*, 36–48. doi:10.1080/00031305.1983.10483087
- Ellis, H. D., Davies, G. M., Shepherd, J. W. (1978). Remembering pictures of real and “unreal” faces. *British Journal of Psychology, 69*, 467–474.
- Fitzgerald, R. J., & Price, H. L. (2015). Eyewitness identification across the life span: A meta-analysis of age differences. *Psychological Bulletin, 141*, 1228–1265. doi:10.1037/bul0000013
- Garrett, B. L. (2011). *Convicting the innocent: Where criminal prosecutions go wrong*. Cambridge: Harvard University Press.
- Goldsmith, M., Pansky, A., & Koriat, A. (2014). Metacognitive control of memory reporting. In T. Perfect & D. S. Lindsay (Eds.), *Sage handbook of applied memory* (pp. 481–500). London, England: Sage Publications.
- Hancock, P. J. B., 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
- Hertzog, C., & Hulstsch, D. F. (2000). Metacognition in adulthood and old age. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (2nd ed., pp. 417–466). Mahwah, NJ: Lawrence Erlbaum.

- Jenkins, L., Myerson, J., Hale, S., & Fry, A. F. (1999). Individual and developmental differences in working memory across the life span. *Psychonomic Bulletin & Review*, 6, 28–40. doi:10.3758/BF03210810
- Juslin, P., Olsson, N., & Winman, A. (1996). Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from the low confidence–accuracy correlation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1304–1316. doi:10.1037/0278-7393.22
- Kaminski, K. S., & Sporer, S. L. (2017). Discriminating between correct and incorrect eyewitness identifications: The use of appropriate cues. *Journal of Experimental Psychology: Applied*, 23, 59–70. doi:10.1037/xap0000110
- Keast, A., Brewer, N., & Wells, G. L. (2007). Children’s metacognitive judgments in an eyewitness identification task. *Journal of Experimental Child Psychology*, 97, 286–314. doi:10.1016/j.jecp.2007.01.007
- Key, K. N., Cash, D. K., Neuschatz, J. S., Price, J., Wetmore, S. A., & Gronlund, S. D. (2015). Age differences (or lack thereof) in discriminability for lineups and showups. *Psychology, Crime & Law*, 21, 871–889. doi:10.1080/1068316X.2015.1054387
- Kocab, K., Martschuk, N., & Sporer, S. L. (2018). *Identification evidence by elderly witnesses: A meta-analysis*. Unpublished manuscript.
- Koriat, A. (1997). Monitoring one’s own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, 126, 349–370. doi:10.1037/0096-3445.126.4.349
- Koriat, A., & Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: Distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology: General*, 123, 297–315. doi:10.1037/0096-3445.123.3.297
- 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
- Lindsay, D. S., Nilsen, E., & Read, J. D. (2000). Witnessing-condition heterogeneity and witnesses’ versus investigators’ confidence in the accuracy of witnesses’ identification decisions. *Law & Human Behavior*, 24, 685–697. doi:10.1023/A:1005504320565
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user’s guide* (2nd ed.). Mahwah, NJ: Erlbaum.
- Martschuk, N., & Sporer, S. L. (2018). Memory of faces in old age: A meta-analysis. *Psychology & Aging*, 33, 904–923. doi:10.1037/pag0000282
- Meissner, C. A., Sporer, S. L., & Schooler, J. W. (2007). Person descriptions as eyewitness evidence. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *The handbook of eyewitness psychology, Vol. 2. Memory for people* (pp. 3–34). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Mickes, L. (2015). Receiver operating characteristic analysis and confidence–accuracy characteristic analysis in investigations of system variables and estimator variables that affect eyewitness memory. *Journal of Applied Research in Memory and Cognition*, 4, 93–102. doi:10.1016/j.jarmac.2015.01.003
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 125–173). New York: Academic Press.
- Newcombe, R. G. (1998). Two-sided confidence intervals for the single proportion: Comparison of seven methods. *Statistics in Medicine*, 17, 857–872. doi:10.1002/(SICI)1097-0258(19980430)
- Olsson, N. (2000). A comparison of correlation, calibration and diagnosticity as measures of the confidence–accuracy relationship in witness identification. *Journal of Applied Psychology*, 85, 504–511. doi:10.1037/0021-9010.85.4.504
- Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart, and Winston.
- Palmer, M. A., Brewer, N., & Weber, N. (2010). Postidentification feedback affects subsequent eyewitness identification performance. *Journal of Experimental Psychology: Applied*, 16, 387–398. doi:10.1037/a0021034
- 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
- Palmer, E. C., David, A. S., & Fleming, S. M. (2014). Effects of age on metacognitive efficiency. *Consciousness and Cognition*, 28, 151–160. doi:10.1016/j.concog.2014.06.007
- Penrod, S., & Cutler, B. (1995). Witness confidence and witness accuracy: Assessing their forensic relation. *Psychology, Public Policy, & Law*, 1, 817–865. doi:10.1037/1076-8971.1.4.817
- Perfect, T. J., & Weber, N. (2012). How should witnesses regulate the accuracy of their identification decisions: One step forward, two steps back? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38, 1810–1818. doi:10.1037/a0028461
- Price, J., Mueller, M. L., Wetmore, S., & Neuschatz, J. (2014). Eyewitness memory and metamemory in older adults. In M. Toglia, D. Ross, J. Pozzulo, & E. Pica (Eds.), *The elderly witness in court* (pp. 167–191). New York: Psychology Press.
- Read, J. D. (1995). The availability heuristic in person identification: The sometimes misleading consequences of enhanced contextual information. *Applied Cognitive Psychology*, 9, 91–121. doi:10.1002/acp.2350090202
- Read, J. D., Lindsay, D. S., & Nicholls, T. (1998). The relationship between accuracy and confidence in eyewitness identification studies: Is the conclusion changing? In C. P. Thompson, D. Herrmann, J. D. Read, D. Bruce, D. Payne, & M. P. Toglia (Eds.), *Eyewitness memory: Theoretical and applied perspectives* (pp. 107–130). Mahwah, NJ: Erlbaum.
- Rhodes, M. G., & Anastasi, J. S. (2012). The own-age bias in face recognition: A meta-analytic and theoretical review. *Psychological Bulletin*, 138, 146–174. doi:10.1037/a0025750

- Richler, J. J., Cheung, O. S., & Gauthier, I. (2011). Holistic processing predicts face recognition. *Psychological Science*, 22, 464–471. doi:10.1177/0956797611401753
- Richler, J. J., Mack, M. L., Gauthier, I., & Palmeri, T. J. (2009). Holistic processing of faces happens at a glance. *Vision Research*, 49, 2856–2861. doi:10.1016/j.visres.2009.08.025
- Roediger, H. L., & DeSoto, K. A. (2014). Confidence in memory: Assessing positive and negative correlations. *Memory*, 22, 76–91. doi:10.1080/09658211.2013.795974
- Roediger, H. L., & DeSoto, K. A. (2015). Understanding the relation between confidence and accuracy in reports from memory. In D. S. Lindsay, C. M. Kelley, A. P. Yonelinas, & H. L. Roediger (Eds.), *Remembering: Attributions, processes, and control in human memory* (pp. 347–367). New York, NY: Psychology Press.
- Sauer, J., Brewer, N., Zweck, T., & Weber, N. (2010). The effect of retention interval on the confidence–accuracy relationship for eyewitness identification. *Law & Human Behavior*, 34, 337–347. doi:10.1007/s10979-009-9192-x
- Sauerland, M., Sagana, A., Sporer, S. L. & Wixted, J. T. (2018). Decision time and confidence predict choosers' identification performance in photographic showups. *PLoS ONE*, 13, e019041. doi:10.1371/journal.pone.0190416
- 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, 611–625. doi:10.1080/10683160701264561
- Sauerland, M., & Sporer, S. L. (2009). Fast and confident: Postdicting eyewitness identification accuracy in a field study. *Journal of Experimental Psychology: Applied*, 15, 46–62. doi:10.1037/a0014560
- Searcy, J. H., Bartlett, J. C., & Memon, A. (1999). Age differences in accuracy and choosing in eyewitness identification and race recognition. *Memory & Cognition*, 27, 538–552. doi:10.3758/BF03211547
- Searcy, J. H., Bartlett, J. C., Memon, A., & Swanson, K. (2001). Aging and lineup performance at long retention intervals: Effects of metamemory and context reinstatement. *Journal of Applied Psychology*, 86, 207–214. doi:10.1037/0021-9010.86.2.207
- Shing, Y. L., Werkle-Bergner, M., Li, S.-C., & Lindenberger, U. (2009). Committing memory errors with high confidence: Older adults do but children don't. *Memory*, 17, 168–179. doi:10.1080/09658210802190596
- Sporer, S. L. (1991). Deep–deeper–deepest? Encoding strategies and the recognition of human faces. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 17, 323–333. doi:10.1037/0278-7393.17.2.323
- Sporer, S. L. (1992). Post-dicting eyewitness identification accuracy: Confidence, decision-times and person descriptions among choosers and non-choosers. *European Journal of Social Psychology*, 20, 157–180. doi:10.1002/ejsp.2420220205
- Sporer, S. L. (1993). Eyewitness identification accuracy, confidence, and decision times in simultaneous and sequential lineups. *Journal of Applied Psychology*, 78(1), 22–33. doi:10.1037/0021-9010.78.1.22
- Sporer, S. L. (1994). Decision times and eyewitness identification accuracy in simultaneous and sequential lineups. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 300–327). New York, NY, US: Cambridge University Press. doi:10.1017/CBO9780511759192.015
- Sporer, S. L. (2001). The cross-race bias: Beyond recognition of faces in the laboratory. *Psychology, Public Policy, and Law*, 7, 170–200. doi:10.1037/1076-8971.7.1.170
- Sporer S. L., Penrod S., Read D., & Cutler B. L. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence–accuracy relation in eyewitness identification studies. *Psychological Bulletin*, 118, 315–327. doi:10.1037/0033-2909.118.3.315
- Stebly, N. K., Wells, G. L., & Douglass, A. B. (2014). The eyewitness post identification feedback effect 15 years later: Theoretical and policy implications. *Psychology, Public Policy, and Law*, 20, 1–18. doi:10.1037/law0000001
- Thomas, A. K., Gordon, L. T., & Bulevich, J. B. (2014). Uniting theory to empirical evidence: How to understand memory of the elderly witness. In M. Toglia, D. Ross, J. Pozzulo, & E. Pica (Eds.), *The elderly witness in court* (pp. 308–335). New York: Psychology Press.
- Tryon, W. W. (2001). Evaluating statistical difference, equivalence, and indeterminacy using inferential confidence intervals: An integrated alternative method of conducting null hypothesis statistical tests. *Psychological Methods*, 6, 371–386. doi:10.1037/1082-989X.6.4.371
- Tulving, E. (1985). How many memory systems are there? *American Psychologist*, 40, 385–398. doi:10.1037/0003-066X.40.4.385
- Watier, N. N., & Collin, C. A. (2011). Metamemory for faces, names, and common nouns. *Acta Psychologica*, 138, 143–154. doi:10.1016/j.actpsy.2011.05.017
- Weber, N., & Brewer, N. (2003). The effect of judgment type and confidence scale on confidence–accuracy calibration in face recognition. *Journal of Applied Psychology*, 88, 490–499. doi:10.1037/0021-9010.88.3.490
- 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
- Weil, L. G., Fleming, S. M., Dumontheil, I., Kilford, E. J., Weil, R. S., Rees, G., ... Blakemore, S.-J. (2013). The development of metacognitive ability in adolescence. *Consciousness and Cognition*, 22, 264–271. doi:10.1016/j.concog.2013.01.004
- Wells, G. L., & Murray, D. M. (1983). What can psychology say about the Neil v. Biggers Criteria for judging eyewitness accuracy? *Journal of Applied Psychology*, 68, 347–362. doi:10.1037/0021-9010.68.3.347
- Wells, G. L., Rydell, S. M., & Seelau, E. P. (1993). The selection of distractors for eyewitness lineups. *Journal of Applied Psychology*, 78, 835–844. doi:10.1037/0021-9010.78.5.835

- Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin*, 25, 1115–1125. doi:10.1177/01461672992512005
- Wells, G. L., Yang, Y., & Smalarz, L. (2015). Eyewitness identification: Bayesian information gain, base-rate effect equivalency curves, and reasonable suspicion. *Law and Human Behavior*, 39, 99–122. doi:10.1037/lhb0000125
- Wixted, J. T., Mickes, L., Clark, S. E., Gronlund, S. D., & Roediger, H. L. (2015). [Initial eyewitness confidence reliably predicts eyewitness identification accuracy](#). *American Psychologist*, 70, 515–526. doi:10.1037/a0039510
- Wixted, J. T., Read, J. D., & Lindsay, D. S. (2016). The effect of retention interval on the eyewitness identification confidence–accuracy relationship. *Journal of Applied Research in Memory and Cognition*, 4, 329–334. doi:10.1016/j.jarmac.2016.04.006
- Wixted, J. T., & Wells, G. R. (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
- Wright, D. B., Boyd, C. E., & Tredoux, C. G. (2001). [A field study of own-race bias in South Africa and England](#). *Psychology, Public Policy, & Law*, 7, 119–133. doi:10.1037/1076-8971.7.1.119
- Wright, D. B., & Stroud, J. (2002). Age differences in lineup identification accuracy: People are better with their own age. *Law & Human Behavior*, 26, 641–654. doi:10.1023/A:1020981501383
- Yaniv, I., Yates, J. F., & Smith, J. E. K. (1991). Measures of discrimination skill in probabilistic judgment. *Psychological Bulletin*, 110, 611–617. doi:10.1037/0033-2909.110