



Partitioning the sources of recognition confidence: The role of individual differences

Justin Kantner¹ · Ian G. Dobbins²

© The Psychonomic Society, Inc. 2019

Abstract

Confidence ratings during old–new recognition memory tasks are thought to index the strength of memory evidence elicited by test probes. However, various subject-specific factors may also influence reported confidence, including perceived self-efficacy and idiosyncratic interpretation of the confidence scale. To measure the contribution of subject-specific variables to confidence ratings, we performed regression analyses on extant data from three recognition experiments encompassing procedural variations in encoding and stimuli, testing the degree to which the person making the judgment (the “subject” factor), versus whether or not the judgment is accurate, influences reported confidence. Overall, confidence was less linked to changes in accuracy, for “new” than for “old” judgments. Critically, the subject factor was at least as predictive of rated confidence as accuracy for “old” judgments, whereas for “new” judgments the subject factor was substantially more predictive of confidence than accuracy. These results suggest that measured confidence is largely a function of who is making the rating, especially when items are identified as “new.” This suggests that the utility of confidence in predicting memory accuracy will be limited when stable estimates of subject contributions are unavailable, such as when each subject provides one or a few responses.

Keywords Recognition memory · Confidence · Individual differences

In a recognition memory test, observers are presented with stimuli (e.g., words, faces, pictures), some of which were also presented during an earlier study phase (“old” items), and some of which were not (“new” items). The task is to judge whether each item is old or new. These judgments are often characterized as a signal-detection decision process (e.g., Parks, 1966) in which test probes generate strength of evidence “signals” indicating the relative likelihood that the item was studied. If strength exceeds a criterial level established by

the observer, the item will be called “old”; otherwise, the item will be called “new.” A recognition judgment, then, is taken to indicate whether an item’s evidence of prior occurrence was sufficient to exceed a criterion.

Beyond measuring whether items are classified as “old” versus “new,” researchers are often interested in estimating *how much* memory evidence they elicit. A common method of estimating relative memory strength is to ask observers to rate their confidence using a graded scale; for example, from 1 (*definitely new*) to 6 (*definitely old*). Using this 1–6 scale assumes that evidence of prior study increases monotonically with observed confidence rating. Thus, a rating of 2 indicates more evidence of oldness than a rating of 1, although both are considered “new” judgments, while a rating of 5 indicates greater evidence of oldness than a rating of 4, although both are considered “old” judgments. This assumption is central to the use of confidence judgments in building receiver operating characteristic (ROC) curves, which plot hit and false alarm rates at a series of theoretical response criteria that are inferred from the confidence ratings and have played a major role in testing theories of recognition (Macmillan & Creelman, 2005; Yonelinas & Parks, 2007). Likewise, models that use reaction

Electronic supplementary material The online version of this article (<https://doi.org/10.3758/s13423-019-01586-w>) contains supplementary material, which is available to authorized users.

✉ Justin Kantner
justin.kantner@csun.edu

✉ Ian G. Dobbins
idobbins@wustl.edu

¹ Department of Psychology, California State University, Northridge, Northridge, CA 91330, USA

² Department of Psychological & Brain Sciences, Washington University in Saint Louis, Saint Louis, MO 63130, USA

time data to infer evidence accumulation/random walk processes underlying confidence ratings assume that these processes, and thus the resulting confidence ratings, are initiated and driven by the match between a test probe and the contents of memory (e.g., Pleskac & Busemeyer, 2010; Ratcliff & Starns, 2009).

Although recognition confidence ratings differ for accurate and inaccurate responding, and thus are systematically related to signal strength, the potential influence of subject-specific factors has been overlooked. Indeed, subjects' general metacognitive confidence appears to be intraindividually stable. For example, Kelemen, Frost, and Weaver (2000) found individual differences in confidence that were stable across tasks (e.g., paired associate learning, text comprehension) and time (1 week). Kantner and Lindsay (2014) tested recognition in the same group of participants across two ostensibly unconnected experiments an average of 2.5 weeks apart. Recognition confidence was significantly correlated across experiments ($r = .38$) even after controlling for recognition sensitivity and response bias, and despite differences in memoranda and confidence reporting method (at the same time as versus following the old–new judgment). Thus, some participants will tend to express more confidence than others, for reasons that may be idiosyncratic and unrelated to their accuracy. One such reason may be variation in self-efficacy across participants, some of whom come to a memory task lacking belief in their own memory abilities (and may express generally lower confidence in their judgments) while others believe memory to be their strong suit and give higher confidence ratings. Alternatively (or in addition), individual differences in confidence ratings may reflect differences in confidence scale interpretation. Confidence ratings are often solicited using labels such as “low/medium/high” or “maybe/probably/definitely,” but the mapping of internal feelings of certainty onto these descriptive labels is inherently subjective and likely varies across participants.

Models characterizing or predicting confidence ratings solely based on assumed memory evidence may be misleading to the extent that nonmemory factors also drive reported confidence. In addition, individual differences in confidence would carry particular implications for settings in which confidence is used to infer accuracy, such as eyewitness memory. However, the relative magnitude of subject-level factors' influence on confidence has not been established. We conducted regression analyses on confidence rating data from three extant recognition data sets including experiments varying in encoding and stimuli in order to test the degree to which the accuracy of the report (correct or incorrect) versus the origin of the report (which subject it came from) influenced the level of confidence. We refer to these as the accuracy and subject factors, respectively.

Analyses were performed separately on “old” and “new” responses to test the possibility that the subject factor would

play a greater role in “new” compared with “old” report confidence. This possibility is motivated by dual process theories of recognition (e.g., Yonelinas, 1994), which assume that a highly diagnostic source of information, recollection of context, is only possible for “old” conclusions, leading to a generally higher level of confidence for “old” versus “new” conclusions, and a greater resistance to extraneous influences (e.g., Jacoby, 1991; Jaeger, Cox, & Dobbins, 2012). This prediction is also consistent with research demonstrating that the confidence–accuracy relationship is generally better calibrated for “old” than for “new” judgments (Weber & Brewer, 2004).

To the extent that “old” conclusions rest on a firmer basis of evidence than do “new” conclusions, one might also expect that the subject factor will play a lesser role. If so, then a greater proportion of confidence variation in “new” judgments may be linked to factors other than memory evidence. Together, these analyses were designed to determine whether the source of recognition judgments makes a greater or lesser contribution to stated confidence than their accuracy.

Extant Data Set 1: Kantner and Lindsay (2012), Experiments 1, 2, and 4

To establish the relative influences of the accuracy and subject factors in a baseline recognition task (i.e., one without any variables beyond old/new item status or procedural additions such as feedback), we first analyzed the combined data from four baseline recognition experiments (Experiments 1, 2, and 4 in Kantner & Lindsay, 2012, and a fourth experiment with an identical recognition procedure; total $N = 186$). These experiments tested item recognition of words in a typical study–test design, contained 48-item study lists and 96-item test lists (50% old, 50% new) drawn from the same pool of common English nouns, and varied only in the number of study–test cycles (two experiments contained two cycles and the other two contained one). Recognition judgments and confidence ratings were collected in “one shot” (rather than sequentially) on a scale ranging from 1 (*definitely new*) to 6 (*definitely old*).

Figure 1 shows the subjects' mean confidence in erroneous and correct “new” and “old” judgments. The slope of each line reflects the gain in confidence achieved with correct versus incorrect responding. The thick lines indicate the average gain across the subjects for the two types of response, and the dots indicate that subject's average confidence, regardless of accuracy.

The plot reveals several important patterns. First, confidence was higher following correct responses than following errors, but this difference (most easily seen in the slopes of the thick lines) was greater for old responses than for new responses. That is, the increase in confidence from error to correct report for old responses (.75 on a 3-point scale) was greater than the increase in confidence from error to correct report

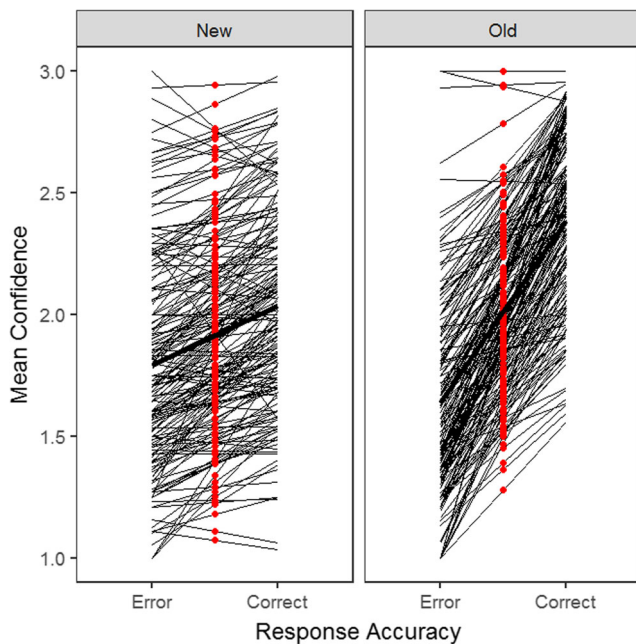


Fig. 1 Mean confidence for each participant following both errors and correct responses in the Kantner and Lindsay (2012) data. Each thin line corresponds to one participant, and dots represent the average confidence for that participant, presented separately for “new” responses (left) and “old” responses (right). Thick lines represent the average across all participants

for new responses (.24; see right and left panels of Fig. 1, respectively). A paired-samples *t* test applied to the slope coefficients from simple regressions predicting confidence from accuracy confirmed that this difference was significant, $t(184) = 17.57, p < .001$, Cohen’s $d = 1.29$. This finding suggests that subjects have more metacognitive information distinguishing correct and incorrect “old” reports than correct and incorrect “new” reports, on the assumption that the variation across subjects in mean confidence is inversely related to the amount of information they have access to for informing their confidence. In the extreme, if subjects had perfect access to the diagnostic value of memorial information, all errors would receive the lowest confidence and all correct reports the highest confidence, which would completely eliminate across-subject differences in average confidence. As metacognitive information decreases from this extreme, variability in confidence should increase. Because variability is higher for “new” than “old” mean confidence in these data, one could infer that less information was available for “new” responses, which would also explain why the difference in confidence between incorrect and correct responses was relatively small for “new” responses compared to “old” responses. As noted above, if there are stable individual differences in the tendency to express high versus low levels of confidence (Kantner & Lindsay, 2014), then these would presumably be maximized in the absence of memorial information, and should thus be a stronger determinant of confidence ratings for “new” than for “old” responses.

The above data suggest that both accuracy and subject variables determine the level of reported confidence. To determine the relative contribution of each factor, we ran fixed effects, simple regression models¹ in which either subject or accuracy was used to predict report confidence. Four models were run: two using the subject factor, and two using response accuracy to predict confidence, separated for “old” versus “new” responses. Figure 2 shows the outcome of these four models in terms of the proportion of variance in confidence accounted for (R^2).

Figure 2 illustrates that the subject factor accounts for approximately 15% of the variation in confidence for “old” responses. Because subject is dummy coded, the first subject serves as the reference, and the regression coefficients reflect whether each remaining subject demonstrates higher or lower average confidence than the reference subject. The actual accuracy of the report also accounted for 15% of the variance in confidence. Turning to “new” responses, the subject factor explains 26% of variance in the confidence data, while the accuracy factor explains only 2%. The results of this analysis support the conclusion that subject-level factors are as predictive as accuracy for “old” confidence judgments, and are much more predictive than accuracy for “new” confidence judgments. This pattern is consistent with the finding that the gain in confidence from incorrect to correct judgments was much larger for “old” judgments than for “new” judgments (see Fig. 1): When participants made a “new” judgment (relative to when they made an “old” judgment), their confidence in that judgment had comparatively little to do with their accuracy, leaving more to be explained by factors other than accuracy, such as subject-level factors. Overall, these regression analyses indicate that for “old” responses, knowing who the response comes from is just as informative as knowing whether or not the response was accurate when predicting observed confidence, and that both reliably predict confidence (indeed, both reliably predict confidence in every analysis of each data set reported here). For “new” responses, by contrast, knowing who provided a report is vastly more important than knowing whether it is accurate when predicting observed confidence.

The reanalysis of the Kantner and Lindsay (2012) data suggest that subjects play a substantial role in driving confidence rating data. These data strongly suggest that treating confidence levels as equivalent across observers is problematic, and that for “new” reports, the dominant contributor to

¹ The regressions treat the subject factor as a fixed effect, because the question is not whether a given effect is reliably expressed within individuals (which would require the treatment of subject as a random effect), but the size of the effect present across individuals within an experiment. To test whether the trends reported here would be different, were subject treated as a random effect, we conducted hierarchical linear modeling analysis on this data set (for details and results, see the [Supplementary Materials](#)). These analysis supported conclusions drawn from the regressions below.

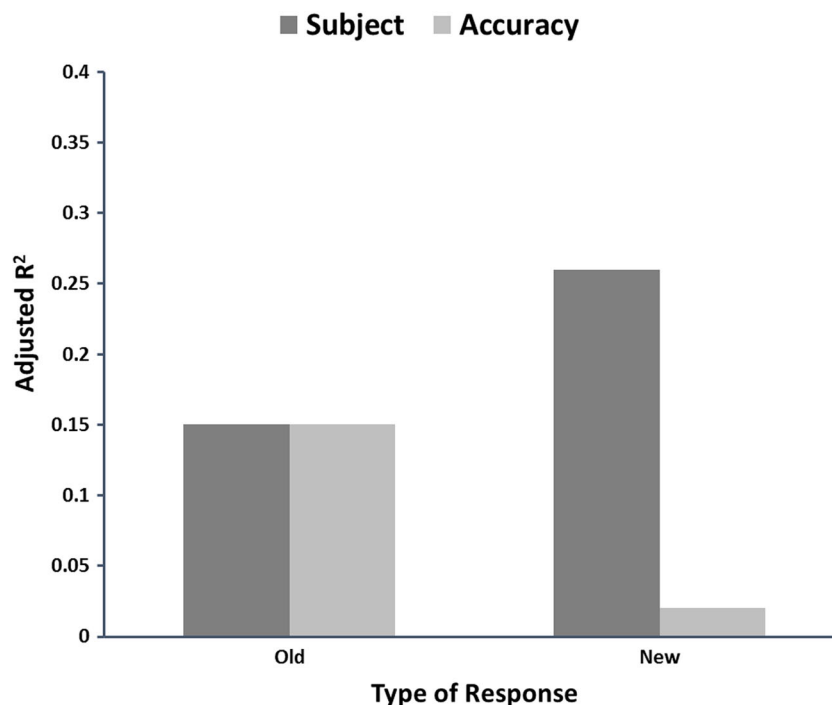


Fig. 2 Proportion of variance in confidence ratings accounted for by the subject factor (darker bars) and response accuracy (lighter bars), separately for “old” and “new” responses, in the Kantner and Lindsay (2012) data

reported confidence is who the report came from, not whether it was accurate. In subsequent reanalyses, we sought to determine whether the sizable influence of individual differences on confidence is robust to variations in experimental design, the method of reporting confidence, and the stimulus materials.

Extant Data Set 2: Selmecky and Dobbins (2013), Experiment 1

Here we examine a recognition memory cuing paradigm, during which some recognition probes are preceded by external cues (or hints) to their study status (“likely old” or “likely new”), while others are uncued (“????”). The primary goal of Selmecky and Dobbins (2013) was to examine whether individual differences in metacognitive awareness affected the degree to which subjects profited from the provision of environmental cues. The cues heavily influenced response bias at the trial level; for example, the probability of an old response was reliably higher following the *likely old* cue than the *likely new* cue or no cue, demonstrating that subjects were biasing judgments based on the cues. Because recognition judgments were heavily influenced by an outside source of information (the cues), the data provide a strong test of whether subject factors still robustly drive confidence reports.

Thirty-seven subjects completed four recognition study–test cycles, two using shallow encoding (an alphabetic order judgment) and two using deep (an abstract/concrete

judgment). For brevity, we focus only on the shallow tests, although deep tests yielded similar results. In addition, we focus only on “cued” trials (those trials beginning with a predictive cue), as these trials constitute the test of interest regarding the durability of subject-level contributions to confidence. For brevity, we do not discuss the results from uncued trials, but they were consistent with those of the cued trials. Each test contained 100 studied items and 100 new items. Following old–new judgments, subjects reported confidence on a 6-point (50% to 100%) confidence scale.

As Fig. 3 illustrates, the patterns of confidence across correct and incorrect “old” and “new” judgments on cued trials were similar to those in the Kantner and Lindsay (2012) data. The increase in confidence from errors to correct responses was greater for “old” judgments (8.93 points on a 50-point scale) than for “new” judgments (5.55 points), which was reliable, $t(33) = 5.34$, $p < .001$, Cohen’s $d = 0.92$. Thus, these results suggest that confidence ratings in the memory cuing paradigm are built upon a stronger metacognitive foundation when they follow “old” responses, leaving more of “new” response confidence to be driven by individual differences.

The results of the fixed effect regression analysis performed on these again reveal the strong contribution of individual differences (see Fig. 4). The subject factor accounts for approximately 25% of the variation in “old” response confidence, while the accuracy factor accounts for 5%. This result is a departure from the “old” confidence analyses of Kantner and Lindsay (2012), in which subject and accuracy accounted for a similar amount of the variance. Here, subject is the

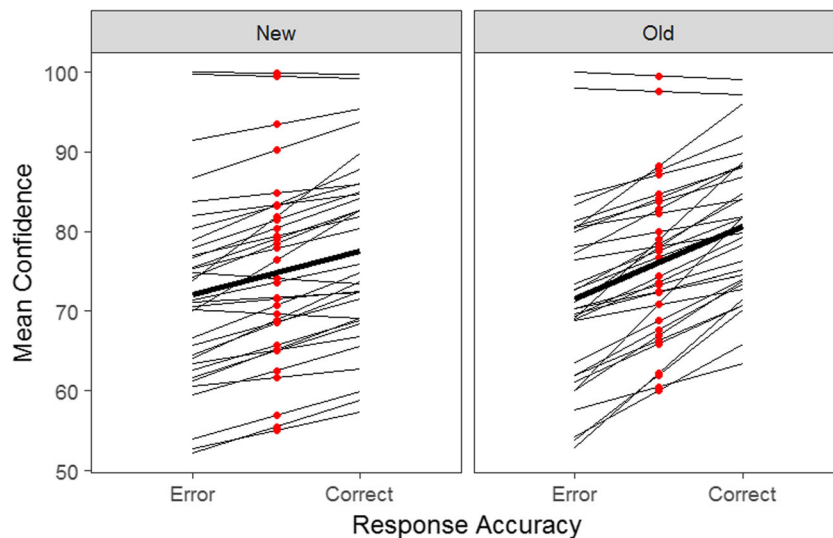


Fig. 3 Mean confidence for each participant following both errors and correct responses during cued trials in Selmezy and Dobbins (2013), Experiment 1. Each thin line corresponds to one participant, and dots

represent the average confidence for that participant, presented separately for “new” responses (left) and “old” responses (right). Thick lines represent the average across all participants

dominant factor in explaining “old” response confidence. The results for “new” responses mirror those reported above: The subject factor explains considerably more variance in confidence (40%) than does accuracy (2%).

Therefore, analyses of the Selmezy and Dobbins (2013) data largely converge with those of the Kantner and Lindsay (2012) data, despite the very different manipulations in place. The current paradigm used an explicit cueing procedure that greatly altered the response probabilities of the subjects on a

trial-wise basis, and clearly also altered their reported confidence during cued trials. Nonetheless, individual differences across participants, captured by the subject factor, remained highly predictive of confidence ratings for both “old” and “new” judgments, and were much more predictive of confidence on a given trial than accuracy on that trial. These results suggest that even when the nature of recognition judgments is heavily moderated by an external source of information, the confidence of those judgments is driven largely by individual

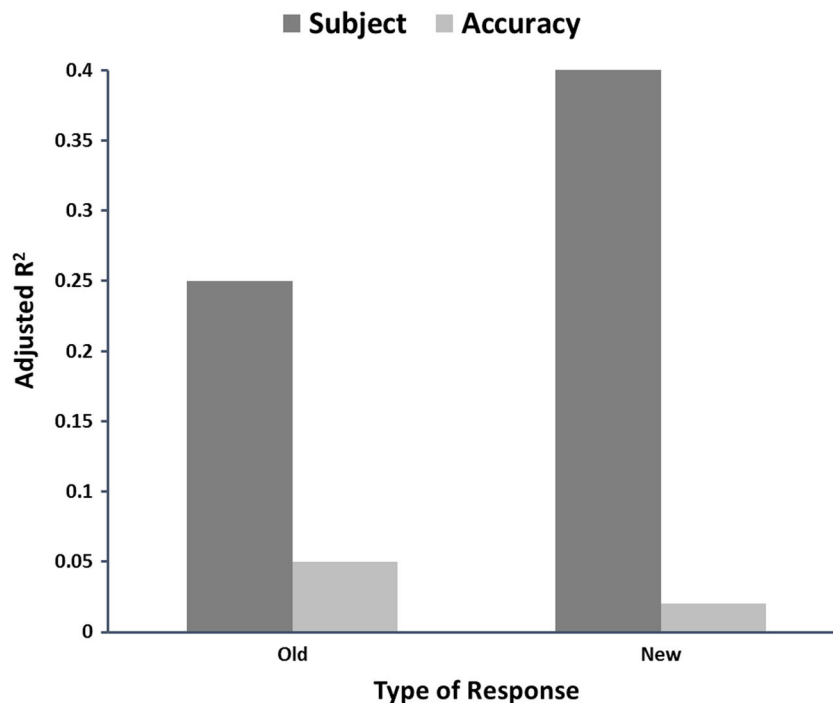


Fig. 4 Proportion of variance in confidence ratings accounted for by the subject factor (darker bars) and response accuracy (lighter bars), separately for “old” and “new” responses, in Selmezy and Dobbins (2013), Experiment 1, cued trials

differences. Indeed, perhaps the impact of this information on subjects' confidence was itself a function of individual differences (e.g., the cues made some subjects more confident in memory decisions and others less confident), such that the provision of external cues amplified the role of subject in confidence judgments. Regardless, it clearly did not diminish the influence of intersubject variability.

Extant Data Set 3: Kantner and Lindsay (2012), Experiment 3

Analysis of the above two extant data sets revealed a substantial role for individual differences in producing participants' reports of their recognition judgment confidence. To further test the generality of this phenomenon, we next analyzed confidence data from an experiment in which the memoranda were pictorial (instead of verbal). Participants made old–new recognition and confidence judgments simultaneously, using the same scale as in [Extant Data Set 1](#). Each participant ($N = 44$) engaged in two study–test cycles, where study lists contained 48 images of paintings, and each test block contained 96 paintings, half previously studied, and half new. Test responses of 1, 2, 3 and 4, 5, 6 were converted to “new” and “old” judgments, respectively, and 1/6, 2/5, and 3/4 were converted to high, medium, and low confidence ratings.

Figure 5 depicts the relationship between confidence following errors and correct responses, and Fig. 6 depicts the variance accounted for by the subject and accuracy factors. The results were consistent with those of the above analyses. Confidence was higher following correct than following incorrect responses, a trend that was greater for “old” (.72-point

increase on a 3-point scale) than for “new” responses (.24-point increase), $t(43) = 8.47$, $p < .001$, Cohen's $d = 1.28$. The subject factor accounted for slightly more variance in confidence ratings than the accuracy factor for “old” responses (16% vs. 14%), and accounted for far more variance than accuracy for “new” confidence (26% vs. 2%). Thus, the trends observed above were robust to a change to pictorial memoranda: The individual subject is at least as predictive of confidence on a given trial as their accuracy on that trial for “old” responses, and is dramatically more predictive than accuracy for “new” judgments.

General discussion

Confidence in recognition (and other tasks amenable to signal-detection analysis) is usually assumed to arise largely from the stimulus evidence. However, the current analyses demonstrate that, compared with the accuracy of the judgment, knowing who generated a confidence rating is at least as predictive of confidence during “old” responses, and decidedly more predictive for “new” responses. Across the four conditions considered here, the subject factor explained 13×, 13×, 13×, and 20× more variance in “new” judgment confidence than the accuracy factor. Whether individual differences in self-efficacy, scale use, or other factors explain intersubject differences in confidence ratings is an open question, but these data highlight that any given single confidence report is determined largely if not predominantly by individual differences, broadly defined.

The presence of individual differences in confidence judgments raises the broader point that confidence as reported in a

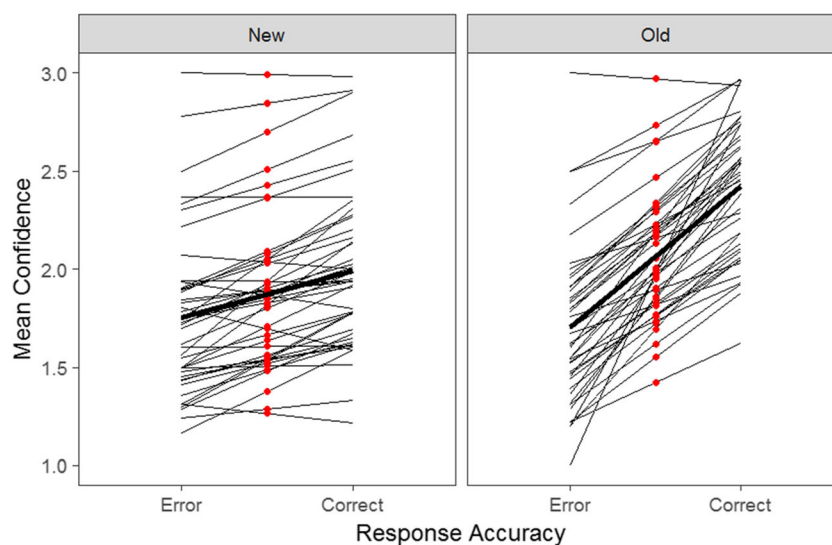


Fig. 5 Mean confidence for each participant following both errors and correct responses in Kantner and Lindsay (2012), Experiment 3. Each thin line corresponds to one participant, and dots represent the average

confidence for that participant, presented separately for “new” responses (left) and “old” responses (right). Thick lines represent the average across all participants

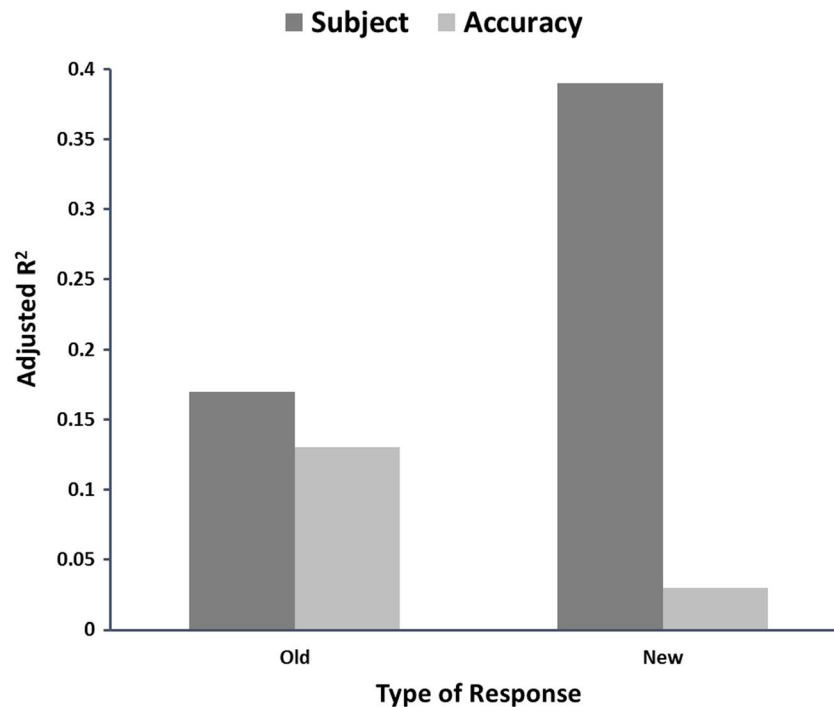


Fig. 6 Proportion of variance in confidence ratings accounted for by the subject factor (darker bars) and response accuracy (lighter bars), separately for “old” and “new” responses, in Kantner and Lindsay (2012), Experiment 3

cognitive task is not a pure expression of evidence strength, but arises from a combination of influences, only some of which are evidence based. Despite the intuitive link between a judgment’s confidence and its accuracy, factors unrelated to the outcome (especially subject factors) govern confidence. This finding has important implications for understanding and interpreting confidence ratings, perhaps particularly in areas such as eyewitness recognition, in which the confidence–accuracy relationship is highly consequential and long debated (e.g., Wixted, Read, & Lindsay, 2016). The current results suggest that when a witness rejects a lineup (essentially judging all lineup members to be “new”), the confidence of the report may essentially reflect an individual difference untethered to report accuracy. In the case of a selection of a perpetrator (essentially equivalent to an “old” decision), the current data raise the possibility that the confidence of the report is as likely to reflect individual differences in mean certainty as it is to reflect the degree of accuracy, as the two effects are similar in magnitude. In addition, the typical eyewitness situation presents a unique problem in that an observer typically provides only a single judgment, making it difficult to know the degree to which a given subject’s single confidence value rests on idiosyncratic subject effects versus the underlying evidence signal. This underscores the need to develop methods for estimating subject-level contributions to cognitive confidence even in paradigms which only gather single reports. For example, in the case of eyewitness reports, perhaps a general knowledge battery, given after the identification, would be useful in estimating a subject’s general

certainty in cognitive judgments. This information could then be used to improve the diagnostic value of his or her prior eyewitness confidence.

More generally, the present analyses indicate that a substantial amount of information is lost when confidence analyses collapse across individuals. Ideally, ROC curves should be fitted to individual subject data rather than group data; due to large intersubject differences in mean confidence, group ROCs are likely to be noisy (and, for this reason, require very large sample sizes). How noisy they are likely varies across experimental paradigms. Thus, one direction for future research is to establish how different experimental manipulations affect the relative importance of subjects in driving confidence ratings. A deep orienting task at study, for example, should engender recollective experiences at test, thereby increasing the correspondence between accuracy and confidence and decreasing the impact of subjects. Other factors, such as corrective feedback or predictive cues, may reinforce or even increase the influence of individual differences, perhaps because subjects’ reactions to these elements of a test are themselves idiosyncratic.

Finally, given that individual differences appear to weigh heavily on confidence ratings (especially for “new” judgments), another direction for research is to develop ways of mitigating their influence. Tekin and Roediger (2017) recently demonstrated that confidence–accuracy calibration is similar across different confidence scale formats. In the context of the current study, this raises the question of whether some types of confidence scales increase or decrease intersubject variability

in mean confidence. If confidence variability is primarily driven by variability in the interpretation of the confidence scale, then the development of less ambiguous scales might be expected to reduce individual differences in mean confidence. By contrast, if confidence variability is a function of differences in global or preexperimental confidence across subjects, the type of scale should have little effect. Thus, experiments attempting to reduce the impact of subject-level factors can serve a broader goal of helping to unpack these factors and identify which produce the significant effects on confidence observed here.

References

- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, 30(5), 513–541. [https://doi.org/10.1016/0749-596X\(91\)90025-F](https://doi.org/10.1016/0749-596X(91)90025-F)
- Jaeger, A., Cox, J. C., & Dobbins, I. G. (2012). Recognition confidence under violated and confirmed memory expectations. *Journal of Experimental Psychology: General*, 141(2), 282–301. <https://doi.org/10.1037/a0025687>
- Kantner, J., & Lindsay, D. S. (2012). Response bias in recognition memory as a cognitive trait. *Memory & Cognition*, 40(8), 1163–1177. <https://doi.org/10.3758/s13421-012-0226-0>
- Kantner, J., & Lindsay, D. S. (2014). Cross-situational consistency in recognition memory response bias. *Psychonomic Bulletin & Review*, 21(5), 1272–1280. <https://doi.org/10.3758/s13423-014-0608-3>
- Kelemen, W. L., Frost, P. J., & Weaver, C. I. (2000). Individual differences in metacognition: Evidence against a general metacognitive ability. *Memory & Cognition*, 28(1), 92–107. <https://doi.org/10.3758/BF03211579>
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.
- Parks, T. E. (1966). Signal-detectability theory of recognition-memory performance. *Psychological Review*, 73(1), 44–58. <https://doi.org/10.1037/h0022662>
- Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review*, 117(3), 864–901. <https://doi.org/10.1037/a0019737>
- Ratcliff, R., & Starns, J. J. (2009). Modeling confidence and response time in recognition memory. *Psychological Review*, 116(1), 59–83. <https://doi.org/10.1037/a0014086>
- Selmecey, D., & Dobbins, I. G. (2013). Metacognitive awareness and adaptive recognition biases. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(3), 678–690. <https://doi.org/10.1037/a0029469>
- Tekin, E., & Roediger, H. L. (2017). The range of confidence scales does not affect the relationship between confidence and accuracy in recognition memory. *Cognitive Research: Principles and Implications*, 2(1), 49.
- Weber, N., & Brewer, N. (2004). Confidence-accuracy calibration in absolute and relative face recognition judgments. *Journal of Experimental Psychology: Applied*, 10(3), 156–172. <https://doi.org/10.1037/1076-898X.10.3.156>
- 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*, 5(2), 192–203. <https://doi.org/10.1016/j.jarmac.2016.04.006>
- Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(6), 1341–1354. <https://doi.org/10.1037/0278-7393.20.6.1341>
- Yonelinas, A. P., & Parks, C. M. (2007). Receiver operating characteristics (ROCs) in recognition memory: A review. *Psychological Bulletin*, 133(5), 800–832. <https://doi.org/10.1037/0033-2909.133.5.800>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.