Memory

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/pmem20

Boosting confidence without boosting performance: item strength creates the illusion of source accuracy

D. Merika W. Sanders, Rosemary A. Cowell, Johanny Castillo & Jeffrey J. Starns

To cite this article: D. Merika W. Sanders, Rosemary A. Cowell, Johanny Castillo & Jeffrey J. Starns (2022) Boosting confidence without boosting performance: item strength creates the illusion of source accuracy, Memory, 30:9, 1172-1191, DOI: 10.1080/09658211.2022.2098338

To link to this article: https://doi.org/10.1080/09658211.2022.2098338

Published online: 14 Jul 2022.

MEMORY
2022, VOL. 30, NO. 9, 1172–1191
https://doi.org/10.1080/09658211.2022.2098338

Routledge
Taylor & Francis Group

Check for updates

Boosting confidence without boosting performance: item strength creates the illusion of source accuracy

D. Merika W. Sanders [a,b], Rosemary A. Cowell[b], Johanny Castillo[b] and Jeffrey J. Starns[b]

[a]Department of Psychology, Harvard University, Cambridge, MA, USA; [b]Department of Psychological and Brain Sciences, University of Massachusetts Amherst, Amherst, MA, USA

ABSTRACT
People often express high confidence for misremembered sources. Starns and Ksander ([2016]. Item strength influences source confidence and alters source memory zROC slopes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(3), 351-365; hereafter SK16) found that this happens more often when a person is highly confident in memory for the item itself, and that simply increasing item memory can increase high-confidence source errors. Under the decision heuristic account, this pattern emerges because strong item memories contaminate source judgments by promoting high confidence responses even when source evidence is relatively weak. Consequently, strengthening item memory is predicted to increase confidence for *both* correct and incorrect source responses; however, SK16 could not assess this key prediction because their item-strength manipulation also impaired source memory. We report two experiments with new item-strengthening manipulations designed to minimise source memory impairments. Results replicated the evidence for the decision heuristic account reported by SK16 and provided additional support by showing a boost in source confidence for both correct and error responses when item memory was strengthened without accompanying source impairments .

ARTICLE HISTORY
Received 25 February 2022
Accepted 30 June 2022

KEYWORDS
Source memory; confidence ratings; decision processes; hierarchical Bayesian modelling; receiver operating characteristic (ROC) functions

Source memory is the ability to link remembered information to the context in which it was encountered, such as remembering where, when, or how you learned a specific claim, remembering who you were with the first time you saw a movie, or remembering what you wore on a job interview (Johnson et al., 1993). Source memory is often contrasted with item memory, the ability to remember the central information that defines an event, such as the content of a claim, the plot of a movie, or what you were asked in an interview. Source confusions can have a range of deleterious consequences. For example, the spread of misinformation and false news stories is an issue made prominent by the 2016 US presidential election and further intensified during the national response to the COVID-19 pandemic (Bavel et al., 2020; Lazer et al., 2018). Rather than censoring the content of individual stories, one strategy to combat both belief in and spread of misinformation has focused on highlighting the source of the news. For example, Facebook's context button provides more information about the source of an article by linking to the publisher's Wikipedia entry (Hughes et al., 2018), and YouTube has added notices to videos published by sources that receive government or public funding (Samek, 2018). But this strategy could be undermined if someone remembers the content of a false news story but misremembers its source – a scenario that highlights the fundamental and important difference between item and source memory.

Unfortunately, people often make the mistake of attributing remembered information to incorrect sources (Schacter, 1999). Further, people often express high confidence in their source misattributions (e.g., Zaragoza & Lane, 1994; Zaragoza & Mitchell, 1996). Understanding the cognitive processes that underlie high-confidence source memories, both veridical and false, is a critical goal for memory research.

In the laboratory, one common approach to investigating source memory is a list-learning paradigm (Johnson et al., 1993). Participants first study a list of words, each simultaneously presented with one of two possible sources (e.g., either with an image of a face or a scene). On a subsequent test, participants then rate both their confidence that each word was studied (e.g., *definitely new* to *definitely old*) and their confidence that each word was paired with a given source (e.g., *definitely paired with a face* to *definitely paired with a scene*). The first rating provides a measure of recognising the word, or item memory, regardless of the image it appeared with, while the second rating reflects the specific memory for the accompanying image, or source memory. A wide range

CONTACT D. Merika W. Sanders merikasanders@fas.harvard.edu Department of Psychology, Harvard University, Cambridge, MA 02138, USA

of characteristics can be used to define different sources, including different individuals (e.g., Starns & Hicks, 2013), different locations (e.g., Starns & Hicks, 2008), and different processing tasks (e.g., Marsh & Hicks, 1998).

Studies using this approach have found that when participants have high confidence in their item memory (i.e., they are very sure that an item was studied), they are often highly confident in their source memory (e.g., very certain which image it was paired with; DeCarlo, 2003; Slotnick et al., 2000; Slotnick & Dodson, 2005). Moreover, this can even hold true when their source judgment is in fact incorrect, and even when the test word was not studied at all, but is instead a falsely recognised lure word (Starns et al., 2013; Starns & Ksander, 2016). We will use the term "confidence contamination effect" for any empirical pattern in which high item strength promotes high source confidence without an accompanying improvement in source accuracy. The idea is that strong item memories contaminate source judgments by promoting high source confidence even when memory for the specific source information is vague.

Our conceptualisation of item-memory contamination is similar to the item-memory misattribution (IMM) model outlined by Dobbins and McCarthy (2008). This model concerns tasks in which participants are directly queried about one specific source – e.g., "Was this word studied with a scene?" – and the model basically proposes that participants sometimes respond "yes" because they remember the item even when they fail to remember specific source details. In the more common procedure of asking participants to choose between possible sources – "Was this word studied with a scene or a face?" – item memory contamination should not necessarily increase responding for one particular source, but high item strength might contribute to a high level of confidence in whichever source the participant selects. We did not directly apply the IMM model, as it is not designed to be applied to a joint recognition and source task. However, the IMM model and the models considered herein are certainly conceptually linked, and we discuss their relationship further in the General Discussion.

Signal detection models of joint recognition and source memory can accommodate item-strength effects on source confidence with a decision heuristic called the "converging criteria" account, as detailed in the *Implications for Signal Detection Models of Memory* section below (Starns et al., 2013, 2014; Starns & Ksander, 2016; see Onyper et al., 2010 and Klauer & Kellen, 2010 for similar approaches). At a psychological level, this account holds that people are more willing to make highly confident source responses when they have strong item memories. For example, for items with a moderate level of source memory, this account predicts that people report low source confidence when they have a weak memory for the item itself, whereas they report high confidence if they have a strong memory for the item itself. In other words, high item strength promotes high source confidence even when the strong item memory is not accompanied by a clear source memory. The "criteria" referred to in the account are participants' standards for how well they need to remember source information to justify making a high confidence source response. The statement that these criteria "converge" with high item strength means that participants become more lax with their standards; that is, they become increasingly likely to report high source confidence even for relatively vague source memories. We will sometimes discuss this general psychological pattern without tying it to a particular model implementation, in which case we will refer to it as the "decision heuristic account".

Prior findings from Starns and Ksander (2016)

One key prediction of the decision heuristic account is that strengthening only item memory (i.e., without strengthening source memory) should increase source confidence. This is because the boost to source confidence is produced by relaxing the standards for making a high confidence response, so the heuristic should affect responding even without any accompanying increase in source accuracy. Starns and Ksander (2016; hereafter SK16) tested this prediction in two experiments. In Experiment 1, participants studied words that were either presented only once with one of two faces (male or female; the *No repetition* condition) or presented once with a face (male or female) and twice with one of two animal images (bird or fish; the *Different-source repetition* condition). The goal of the different-source repetition manipulation was to strengthen item memory for the repeated words without improving source memory for whether the word was studied with a male or female face (see Dobbins & McCarthy, 2008, Experiment 3, for a similar manipulation). Experiment 2 was similar to Experiment 1 except that in addition to the *No repetition* and *Different-source repetition* conditions, a subset of items was presented three times with the same face (i.e., a *Same-source repetition* condition). Both Experiments 1 and 2 found that strengthening item memory without strengthening source memory resulted in greater overall source confidence, as predicted by the decision heuristic account. Unsurprisingly, strengthening both item and source memory (i.e., *Same-source repetition*) also increased source confidence.

In addition to an overall increase in source confidence, the decision heuristic account predicts that, all else being equal, strengthening item memory should increase source confidence for both correct source responses and source errors. Unfortunately, SK16's strengthening manipulation had a side effect that prevented them from assessing this prediction; namely, different-source repetition impaired source memory. The source memory drop from the *No repetition* to the *Different-source repetition* condition was fairly large, with source discriminability measured by d' dropping from 1.55 to 1.08 in Experiment 1 and 1.45 to 1.06 in Experiment 2. This drop in source memory should decrease source confidence for correct judgments, counteracting the boost in source confidence predicted by the decision heuristic account. In fact, SK16 observed that, relative to no repetition, different-source repetition increased the proportion of high-confidence source responses for source errors (.25 vs. .37 in Exp. 1 and .26 vs. .39 in Exp. 2) but *not* correct responses (.65 vs. .63 in Exp. 1 and .59 vs. .62 in Exp. 2). SK16 argued that this pattern was produced by the joint influence of a confidence contamination effect and the source memory decrement, with these two factors cancelling out for correct source judgments. This explanation implies that reducing the source memory decrement would show that item strengthening increases source confidence for both correct and incorrect source judgments.

Current aims

The purpose of the current experiments was to replicate and extend SK16, with the primary goals of further testing the decision heuristic account and developing item-strengthening manipulations that minimise collateral effects on source memory. Our experiments provide the opportunity to replicate the evidence for the decision heuristic account reported by SK16; specifically, that strengthening item memory increases source confidence for errors. Our experiments will also attempt to expand the evidence for the decision heuristic account by testing whether item strengthening can boost source confidence for *both* correct and incorrect source responses. SK16 could not cleanly assess source confidence effects for correct source responses, because their item-strength manipulation produced substantial source memory decrements. We explored alternative item-repetition manipulations in an attempt to attenuate the source memory decrement observed by SK16. Specifically, across two experiments, we progressively eliminated two characteristics of the SK16 manipulation that theoretically contributed to the source memory decrease.

First, the different-source repetition in SK16 used encoding trials that were very similar to the target event, which might have contributed to a high level of interference (Postman, 1971; Underwood, 1949). SK16 asked participants to remember the event of studying a word with a picture of a face, and different-source repetition involved studying a word with a picture of an animal. These events are easily confusable. To reduce potential interference, both experiments herein strengthened item memory with encoding events that were more distinct from the target source events. Specifically, item strengthening was achieved via a recall task that did not use pictures as stimuli, whereas the source task involved studying words with either a picture of a scene or a face.

Second, SK16 always presented the encoding trials designed to promote strong item memory (i.e., the different-source trials) before the target source encoding trials, so participants might have devoted less attention to the source encoding trials given that the item was already familiar from earlier in the study phase (e.g., Kim et al., 2012). In Experiment 2, we eliminated this potential factor by having all item-strengthening trials follow an initial source encoding trial for the same item. In summary, our experiments attempted to reduce source memory decrements by limiting source interference (Experiment 1) or by both limiting source interference and eliminating previous study trials with the same word (Experiment 2).

We will assess whether these measures are successful in minimising the negative impact of item strengthening on source discrimination. If we succeed in minimising the source memory decrements produced by the item-strengthening manipulation, then we should see an increase in source confidence for both correct and incorrect responses according to the decision heuristic account. More generally, having manipulations that more selectively impact item strength will be helpful for testing theories of the effect of item strength on source decision processes.

We will also evaluate whether the current experiments replicate previously-observed correlations between item and source confidence that are consistent with a decision heuristic account (SK16; Starns et al., 2013, 2014). Specifically, we will evaluate whether participants are more likely to make high confidence source responses for lures that are falsely recognised with high versus low item confidence. The decision heuristic account predicts a positive relationship between item and source confidence: even though lure items were not studied in either source, high item strength (indexed by high-confidence recognition responses) makes participants amenable to reporting high source confidence. We checked for a similar relationship between item and source confidence for targets that were recognised but attributed to the incorrect source. These correlational patterns are subject to more interpretational difficulties than our experimental manipulation of item strength, as a number of factors might vary across trials in which participants report high versus low recognition confidence. Nevertheless, observing correlations in item and source confidence would offer additional support for the decision heuristic account if results also show a confidence contamination effect produced by the experimental manipulation of item strength.

Finally, SK16 used source memory receiver operating characteristics (ROCs) to contrast memory-based and decision-based models of item-strengthening effects, and the current experiments provide an opportunity to replicate these findings and to see if they hold when the impact of item strengthening on source memory is minimised. We mostly defer discussion of these issues to the *Implications for Signal Detection Models of Memory* section that follows presentation of the empirical results from both experiments. In short, when the decision heuristic account is implemented as converging criteria in a signal detection model of recognition and source memory, the account predicts that item strengthening will change the properties of source memory ROCs. Thus, we will evaluate ROC data as a further test of the decision heuristic account.

The converging criteria account – and the decision heuristic account more generally – predict that we should observe a boost in source confidence and replicate the ROC patterns reported by SK16 even with item strengthening manipulations that minimise decrements in source memory. An alternative possibility is that the source memory decrement in SK16 is critical for producing these outcomes. The *Implications for Signal Detection Models of Memory* section discusses a potential mechanism by which the source decrement itself could produce the SK16 results. If this mechanism is driving the effect, then reducing the source decrement should eliminate the effect of item strength on source confidence and efface the ROC patterns predicted by the converging criteria model.

Experiment 1

Experiment 1 was designed to assess whether reducing source interference reduces the decrement in source memory produced by an item-strengthening manipulation and to explore the confidence contamination effect reported by SK16 with a different encoding procedure. SK16 selectively strengthened item memory by presenting words with a secondary source image in addition to presentations with the target source image. Compared to presenting word items only once, this different-source repetition condition significantly impaired source memory performance. The current design instead strengthened item memory via a recall task with words alone, an encoding task that is dissimilar to studying pictures and therefore is expected to attenuate interference with memory for the picture-based sources. If the predictions of the decision heuristic account hold, then strengthening item memory should increase source confidence. Further, if the current design succeeds in minimising deleterious effects on source memory, then the increase in confidence should be observed for both correct and incorrect source judgments.

Method

Participants

Thirty-six undergraduate students from the University of Massachusetts Amherst participated in exchange for extra credit in their psychology courses. No participants were excluded after applying an inclusion criterion of item memory accuracy numerically greater than chance (.50).

Materials

For each participant, we sampled word item stimuli from a subset of the SUBTLEX$_{US}$ database (Brysbaert & New, 2009); the pool was limited to 412 words with frequency ratings between four and six occurrences per million. Each word was presented with either a face or a scene image. The face image was a photograph of a white male with a neutral expression selected from the Chicago Face Database (Ma et al., 2015) and the scene image was a photograph of a pastoral scene selected from stimuli previously used in Ross et al. (2018); these two individual images (one face, one scene) were used repeatedly throughout the study.

In the study phase, participants completed 10 cycles, each composed of two immediately successive encoding blocks. Figure 1 provides example encoding blocks. In the first encoding block for Experiment 1 (encoding block Type A), participants learned six word items, presented alone for 2000 ms (with a 100 ms inter-trial blank screen), and then completed a self-paced mini-test block of six trials containing an initial-letter cued recall task for each of the just-learned words (Figure 1, top panel, far left). In the second encoding block (encoding block Type B), participants learned 12 word items (six words from the first encoding block and six new words), each paired with a source image (face or scene; 2000 ms presentation with a 100 ms inter-trial blank screen), and then completed a self-paced mini-test block of twelve source memory judgment trials, one for each just-learned word-picture pair (Figure 1, top panel, second from left). Word-picture pairings were randomised, while maintaining an equal distribution of items paired with scenes and faces. We also randomised the assignment of words to encoding blocks (i.e., which words were presented twice) and the within-block presentation order for each participant. In total, participants viewed 120 word-picture pairs, half of which additionally appeared as words alone.

In the test phase, participants completed two self-paced memory tests. First, participants made item memory judgments for 120 studied words and 60 unstudied words sampled from the same pool as the studied words (Figure 1, top panel, second from right). Of the 120 studied words, participants had seen an equal number of strong (presented in both encoding blocks) and weak items (presented in one encoding block) in each source (face/scene). After completing recognition judgments for all 180 word items, participants made source memory judgments for the 120 studied word items and for any of the 60 unstudied word items that they incorrectly judged as "old" from the first memory test (Figure 1, top panel, far right).

Twenty-one of the thirty-six participants completed two sessions of the experiment. In the cases where multiple sessions were completed, each session sampled a new subset of words.

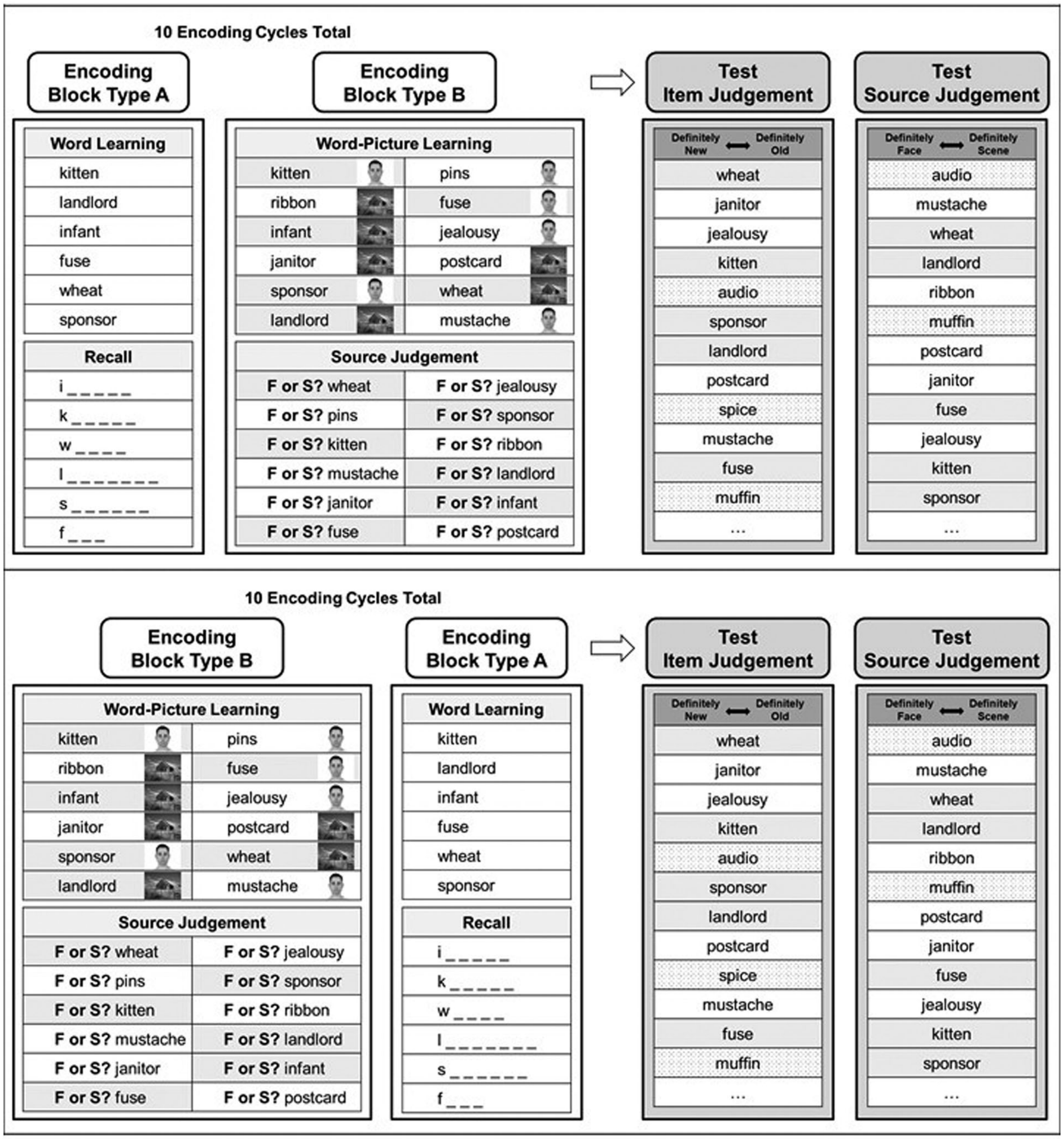

Figure 1. Experiment 1 and Experiment 2 paradigms. Experiment 1 (top panel) and Experiment 2 (bottom panel) designs were identical apart from the sequence of tasks in the study phase. During the study phase, participants completed two types of encoding blocks. In encoding block Type A (Experiment 1: far left; Experiment 2: second from left), participants learned six words presented in succession. Immediately following, participants completed an initial-letter recall task for the words just seen. In encoding block Type B (Experiment 1: second from left; Experiment 2: far left), participants learned 12 words paired with either a face or scene image. Each image type was featured equally among the word-picture pairs. Immediately after, participants were presented with each word and prompted for a source judgment (i.e., "Press 'f' if the word was seen with a face and 's' if the word was seen with a scene"). Importantly, six of the twelve words in encoding block Type B (shaded) were identical to the six words in encoding block Type A, and thus studied twice. In both experiments, the two encoding blocks repeated 10 times with unique words each time prior to the start of the test phase. The test phase was identical in each experiment and consisted of an item recognition test (second from right) of all 120 studied words and 60 unstudied words (dotted) followed by a source memory test (far right) of all 120 studied words and any of the 60 unstudied word items that were incorrectly judged as *old* from the first memory test.

Procedure

Participants completed both study and test phases in a quiet room while seated at a desk with a computer monitor and keyboard. Before the start of the experiment, participants were verbally informed that the experiment was a memory test and that they would study several short lists of words either alone or paired with an image. Written, on-screen instructions informed participants that when the list of words first appeared they should just pay close attention and no response was necessary. Instructions then told participants that after presentation of a word list, they would complete one of two possible tasks. When the words on the list had appeared alone at study (i.e., without a paired source image), they would be given the initial letter of a studied word, which would serve as a prompt to type a word from the list they had just seen, beginning with that letter (Block Type A in Figure 1). When a word list had appeared with images, they instead would be presented with a word from the list they had just seen, and they would be asked to identify which of the two images had appeared with that word (Block Type B in Figure 1). Examples of both encoding blocks followed. The study phase instructions also stated that this first phase would repeat several times and that participants would receive feedback (i.e., the presentation of the correct word item or source image for 1000 ms) after each response.

Instructions for the test phase immediately followed the instructions for the study phase. Participants read that the second phase consisted of two tests that would each serially present one long list of words, some of which had appeared during the first phase and others that would be new. The first test required old-new item decisions on a 6-point confidence scale, where the numbered keys *1*, *2*, *3*, *4*, *5*, and *6* corresponded to each response from *definitely new* to *definitely old*. The second test required face-scene source decisions on a 6-point scale, where the keys "z", "x", "c", ",", ".", and "/" corresponded to each response from *definitely face* to *definitely scene*. The instructions told participants that they would not receive any feedback following their responses in this phase.

As in SK16, participants then practiced using the full 6-point confidence scale for the item and source decisions. Each practice trial presented symbols on the screen that indicated the correct confidence level response to provide. The letter (*O*, *N* or *F*, *S*) indicated the response (*old*, *new* or *face, scene*) and the number of letters indicated the confidence level. For example, *OOO* corresponded to a *definitely old* response, *OO* to a *probably old* response, and *O* to a *maybe old* response. Participants received feedback (i.e., the presentation of the correct response and confidence level) following each practice trial. In total, participants completed one block of 24 practice trials, evenly dispersed among the six confidence levels, for each decision type before the experimental trials began.

Table 1. Source performance measures.

Experiment and Condition	Performance measure				
	p("F" | F)	p("F" | S)	d'	p(hc| C)	p(hc| E)
Experiment 1					
Weak	.65 (.02)	.31 (.01)	.98 (.04)	.57 (.01)	.34 (.02)
Strong	.64 (.02)	.34 (.01)	.84 (.04)	.55 (.01)	.40 (.01)
Lure					.38 (.02)
Experiment 2					
Weak	.63 (.02)	.33 (.02)	.88 (.06)	.55 (.02)	.29 (.02)
Strong	.65 (.02)	.35 (.02)	.81 (.06)	.59 (.02)	.40 (.02)
Lure					.33 (.03)

Note: Standard errors are in parentheses. p("F" | F) = proportion of *face* responses to face items; p("F" | S) = proportion of *face* responses to scene items; $p(hc|\ C)$ and $p(hc|\ E)$ = proportion of source attributions made at the highest confidence level for correct and error responses, respectively.

Results and discussion

We used descriptive statistics and traditional frequentist statistics to explore basic empirical patterns, and we supplement this approach with Bayesian statistics for the modelling analyses reported in the *Implications for Signal Detection Models of Memory* section below. Our primary goals for the initial analyses were to assess whether the item-strengthening manipulation affected source memory, whether item strengthening inflated source confidence for errors as in SK16, and whether item strengthening additionally inflated source confidence for correct responses. We used 1-tailed p values to test for an item-strengthening effect on source confidence because the decision heuristic account makes a strictly directional prediction for this effect. We also conducted 2 (correct vs. incorrect) × 2 (item memory strengthened vs. not strengthened) ANOVAs to see if different interpretations for correct and incorrect responses were licensed by an interaction.

Recognition memory

Repetition (i.e., having a word presented in both encoding block types) improved recognition memory. When collapsing "old" responses across the three confidence levels, the false alarm rate was .16 and the hit rates were .72 and .89 for weak and strong items, respectively. Correspondingly, the difference in d' scores between weak and strong items reflected an improvement in recognition memory following item repetition ($d'_{\text{Weak}} = 1.75$ vs. $d'_{\text{Strong}} = 2.58$).

Source memory

We explored the same source performance measures as reported in SK16 (Table 1). Strong word items that were repeated in both the recall and source memory judgment encoding blocks showed a decrease in source discrimination relative to weak items that appeared only in the source memory judgment encoding block ($d'_{\text{Strong}} = .84$ vs. $d'_{\text{Weak}} = .98$; t (35) = −2.37, $p = .02$, 95% CI [−.27, −.02]). However, this decrement in source memory was relatively small compared to the substantial drop in performance found in SK16 (Experiment 1: $d'_{\text{Strong}} = 1.08$ vs. $d'_{\text{Weak}} = 1.55$; Experiment 2: $d'_{\text{Strong}} = 1.06$ vs. $d'_{\text{Weak}} = 1.45$). Our strategy of reducing interference between the source and item-strengthening trials seems to have been partially successful in decreasing collateral effects of item strengthening on source memory, but there was still a detectable decrement to source performance.

We next analysed the effect of item strength (strong, weak) and source accuracy (correct, error) on the proportion of source responses made at the highest confidence level using a 2×2 ANOVA.[1] For correct responses, we added the number of *definitely face* responses for face items and the number of *definitely scene* responses for scene items and divided the sum by the total number of correct responses at any confidence level. Similarly, for errors, we added the number of *definitely face* responses for scene items and the number of *definitely scene* responses for face items and divided the sum by the total number of errors at any confidence level. We completed these computations separately for each strength condition. There was a main effect of accuracy, F (1, 32) = 33.39, $p < .001$, such that participants were more likely to indicate high confidence for correct responses (.56) than for errors (.37). Further, there was a significant interaction between strength and accuracy, F (1, 32) = 8.01, $p = .008$. For source memory errors, participants were more likely to express high confidence for strong items (.40) than for weak items (.34; t (32) = 3.10, $p = .002$). A similar increase in confidence was not observed for correct source responses; indeed, the proportion of high confidence responses was slightly lower for strong (.55) than weak (.57) items, (t (32) = −.94, $p = .82$).

In a final assessment of the decision heuristic account, we evaluated its prediction that high-confidence source ratings would accompany new items that participants falsely, but confidently, claimed to remember (Figure 2, left panel). When participants reported high confidence in a false alarm (i.e., an unstudied word item that was neither presented with a face nor a scene), they also made a high-confidence source judgment on 30.6% of trials. In comparison, for false alarms made with low or medium confidence, participants reported high confidence in only 19.8% of the subsequent source decisions. The greater proportion of high-confidence source responses for high-confidence false alarms was confirmed with a Chi-square test, $\chi^2(1) = 8.15$, $p = .004$. This replicates the relationship between recognition and source confidence levels reported in both SK16 and Starns et al. (2013). We tested the analogous case for source errors, rather than false alarms; that is, cases when the item was studied and correctly recognised, but the source was incorrectly identified. The same relationship held: the proportion of times participants expressed high confidence in their source errors for targets that were recognised with high item confidence (35.7%) was greater than for targets recognised with low item confidence (17.4%), Chi-square test, $\chi^2(1) = 41.70$, $p < .001$ (Figure 3, left panel).

Conclusions

Despite the altered encoding phase, Experiment 1 successfully replicated many of the results of SK16. Words presented twice – here, in two dissimilar encoding tasks (recall and source memory) – had better recognition performance relative to words presented in only one encoding task. The item-strengthening manipulation also produced a drop in source memory, so we did not succeed in eliminating the source memory "side effect" of strengthening item memory that SK16 noted. The results also replicated the confidence contamination pattern reported by SK16; namely, strengthening item information increased the proportion of high-confidence source errors without meaningfully affecting source confidence for correct source responses. In short, we replicated the support for the decision heuristic account reported by SK16, but we did not expand on this support by extending the confidence contamination effect to correct responses. As in SK16, the fact that the effect was limited to error source responses could be a consequence of the lower source memory for repeated items, as this memory decrement should decrease confidence in correct source responses and counteract the decision heuristic that inflates source confidence for strong item memories.

Results also showed a relationship between item confidence and source confidence for errors, as predicted by the decision heuristic account. Unstudied items that participants were highly confident were "old" were accompanied with high-confidence source responses at a greater rate than other recognition confidence levels. Similarly, participants expressed higher confidence in their incorrect source judgments for targets that they recognised with high versus low recognition confidence.

Experiment 2

The item-strengthening manipulation in Experiment 1 decreased source memory performance, although the size of the effect on source memory seemed to be attenuated compared to the manipulations used by SK16. Experiment 2 attempted to further reduce the impact on source memory by switching the order of the source-encoding trials and the item-strengthening trials. In the Experiment 1 study phase, the source trial for a given word always followed the recall-task (i.e., item-strengthening) trial, so the remaining source memory decrement could be driven by decreased attention to repetitions of the same item. We altered the experimental design such that each word was first encountered in the source encoding task, and then half of the words were subsequently strengthened by reappearing in a recall task (Figure 1, bottom panel).

Results will show whether having item strengthening follow source encoding helps to minimise the impact on source memory. If so, the decision heuristic account predicts that item strengthening will increase high-confidence source responses for both correct and incorrect source judgments. In other words, the confidence effect on source errors observed by SK16 and in Experiment 1 should replicate and extend to correct responses as well. Experiment 2 also provides another opportunity to replicate the finding that participants report higher confidence in source judgments when they are more confident that an

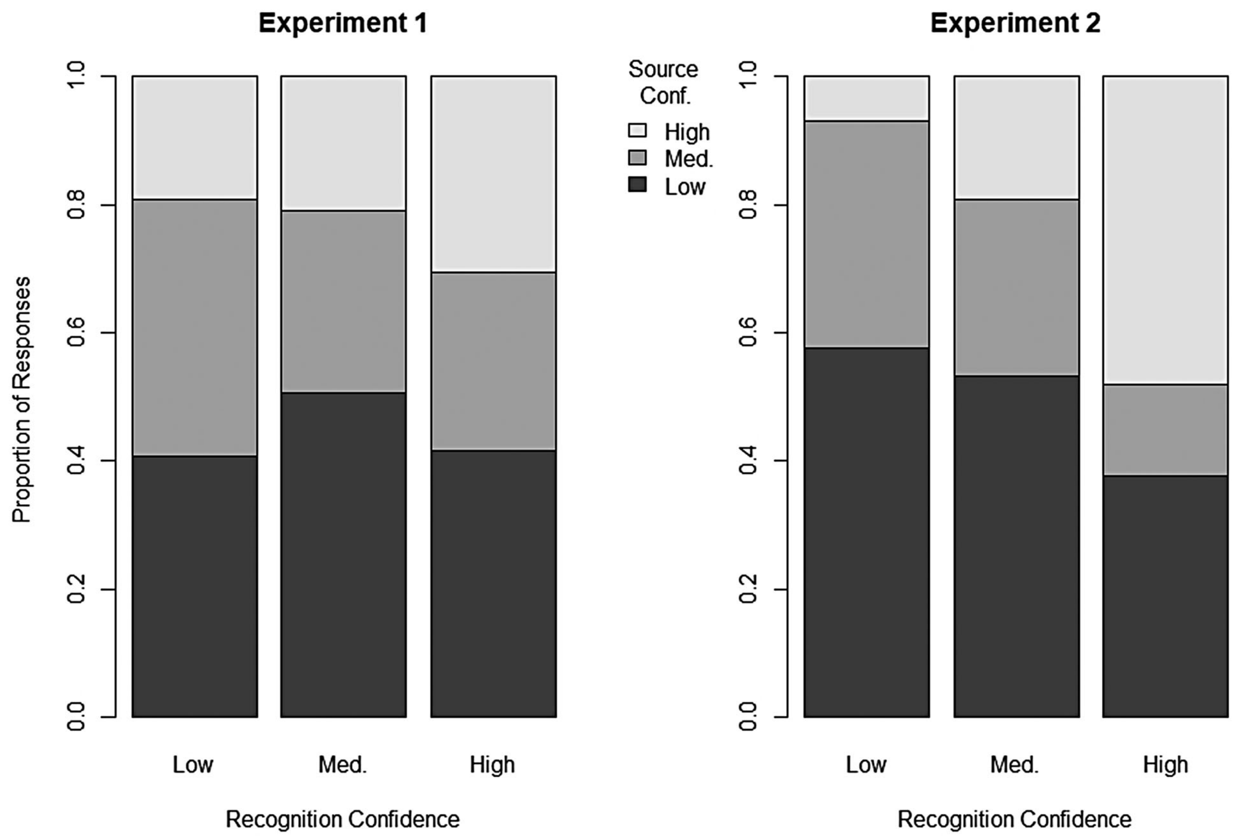

Figure 2. Source confidence for unstudied items. In Experiment 1 (left panel) and Experiment 2 (right panel), participants incorrectly identified some new word items as "old" with various levels of confidence (bars along x-axis, Low = maybe old, Med. = probably old, High = definitely old). For each level of item confidence, the proportion of subsequent source confidence responses are shown in various shades of *colour*. The darkest shade represents the lowest source confidence (maybe face or maybe scene) and the lightest shade represents the highest source confidence (definitely face or definitely scene).

item appeared in the study phase, even when the source judgment is incorrect.

Method

Participants

Thirty-five undergraduate students from the University of Massachusetts Amherst participated in exchange for extra credit in their psychology courses. The analysis included 34 subjects after applying an inclusion criterion of item memory accuracy numerically greater than .5, which excluded one subject.

Materials

The materials were identical to those of Experiment 1; however, each participant completed only one session of the experiment.

Procedure

In Experiment 2, we switched the order of the two encoding blocks during the study phase. Each of the 10 study cycles began with an encoding phase consisting of learning 12 word-picture pairs and then a source memory judgment. In the second encoding block, six of the word items from the first encoding block re-appeared alone in an initial-letter cued recall task. All other procedures were identical to that of Experiment 1.

Results and discussion

Recognition memory

As in Experiment 1, item strengthening improved recognition memory. When collapsing "old" responses across the three confidence levels, the false alarm rate was .14 and the hit rates were .75 and .93 for weak and strong items, respectively. The difference in d' scores between weak and strong items showed a large improvement in recognition memory produced by item repetition ($d'_{\text{Weak}} = 1.95$ vs. $d'_{\text{Strong}} = 2.93$).

Source memory

The results of a *t*-test revealed that strengthening item memory did not produce a substantial decrease in source discrimination ($d'_{\text{Weak}} = .88$ vs. $d'_{\text{Strong}} = .81$; t (33) $= -.80$, $p = .43$, 95% CI [−.26, .11]). Of course, obtaining a nonsignificant result does not provide compelling evidence that the effect size is exactly zero, but the confidence interval shows that the results are only consistent with small effects on source accuracy. Thus, the results show that the strengthening manipulation effectively

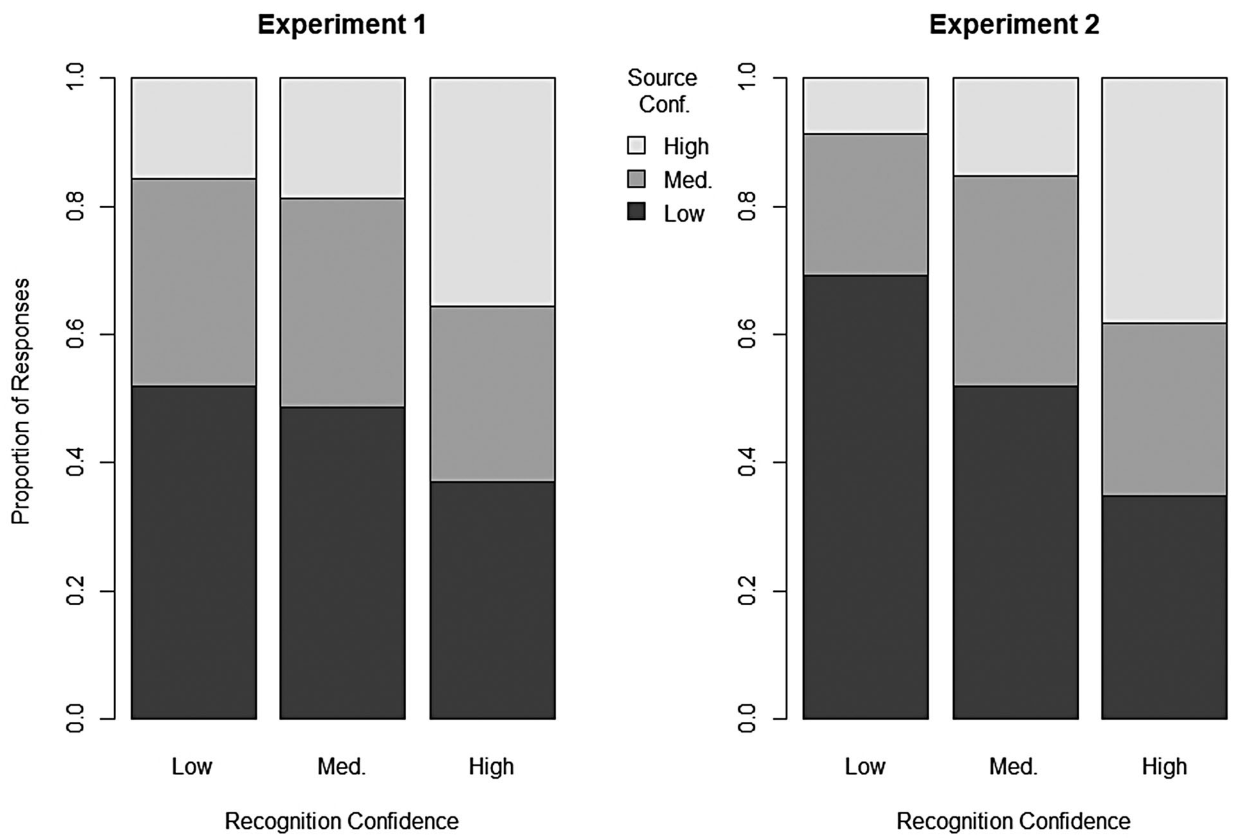

Figure 3. Source error confidence for studied items. In Experiment 1 (left panel) and Experiment 2 (right panel), participants correctly identified some studied word items as "old" with various levels of confidence (bars along x-axis, Low = maybe old, Med. = probably old, High = definitely old), but incorrectly identified the source. For each level of item confidence, the proportion of subsequent source error confidence responses are shown in various shades of colour. The darkest shade represents the lowest source error confidence (incorrect maybe face or maybe scene) and the lightest shade represents the highest source error confidence (incorrect definitely face or definitely scene).

produced a large increase in item memory with minimal impact on source memory.

We analysed the proportion of source responses made at the highest confidence level with a 2 (item strength) × 2 (source accuracy) ANOVA.[2] Participants were more likely to indicate high confidence for correct responses (.57) than for errors (.34), F (1, 28) = 66.99, $p < .001$, and for strong items (.50) than for weak items (.42), F (1, 28) = 17.45, $p < .001$. Additionally, we found a significant interaction between strength and accuracy, F (1, 28) = 4.28, $p = .048$. As in Experiment 1, this interaction reflected a larger confidence contamination effect for source errors than for correct responses. However, unlike Experiment 1, the current results showed a positive confidence contamination effect for both types of responses, as the proportion of high confidence responses increased from .29 to .40 for errors, t (28) = 4.23, $p < .001$, and increased from .55 to .59 for correct responses, t (28) = 1.84, $p = .039$.

The Experiment 2 data were also consistent with a positive relationship between item and source confidence, as predicted by the decision heuristic account. When participants reported high confidence in a false alarm, they also made a high-confidence source decision on 48.1% of trials. In comparison, for false alarms made with low or medium confidence, participants reported high confidence in only 11.2% of trials. This difference in the proportion of high-confidence source ratings was again confirmed with a Chi-square test, $\chi^2(1) = 45.59$, $p < .001$ (Figure 2, right panel). The same relationship held for source errors: participants differed in the proportion of times they expressed high confidence in their source errors for targets that were recognised with high (38.3%) versus low (12.3%) item confidence, Chi-square test, $\chi^2(1) = 48.82$, $p < .001$ (Figure 3, right panel).

Conclusions

Experiment 2 asked whether the strength-induced inflation in high-confidence source responses could be replicated while also eliminating the source memory decrement that was observed in both SK16 and Experiment 1. The methods of Experiment 1 and Experiment 2 were identical except for the ordering of the two encoding tasks. Experiment 2 asked participants to complete a source judgment task for mini-blocks of items before completing a recall task for a subset of those items, which was the reverse of the task order used in Experiment 1. This change to the strengthening manipulation successfully minimised the difference in source memory accuracy between strong and weak items, as there was no significant effect of item strength on source discriminability and the confidence interval was concentrated on very small effect sizes.

The results again replicated the confidence contamination effect produced by the item-strengthening manipulation. Participants were significantly more likely to express high confidence in their source judgments for strong versus weak word items. Moreover, the results expanded evidence for the source contamination effect by demonstrating the effect for both correct and incorrect source judgments, although the effect was still larger for incorrect than correct responses. Finally, results again showed a positive relationship between item confidence and confidence in source errors. Overall, the findings are very consistent with the claim that participants use a decision heuristic in which they are more willing to express high source confidence when they have a strong memory for the item information.

Implications for signal detection models of memory

This section further tests the decision heuristic account by assessing the effect of item strengthening on source ROC functions. We will review past work that implemented the decision heuristic account in terms of converging confidence criteria in a model of joint item and source memory, and we will outline a clear prediction of the converging criteria mechanism for the effect of item strengthening on source ROC functions. We tested this prediction by fitting a signal detection model to the source ROC data. We used a hierarchical Bayesian approach, so our models were able to accommodate variation in parameter values across participants while simultaneously using all available data to inform parameter estimates (Kruschke, 2015).

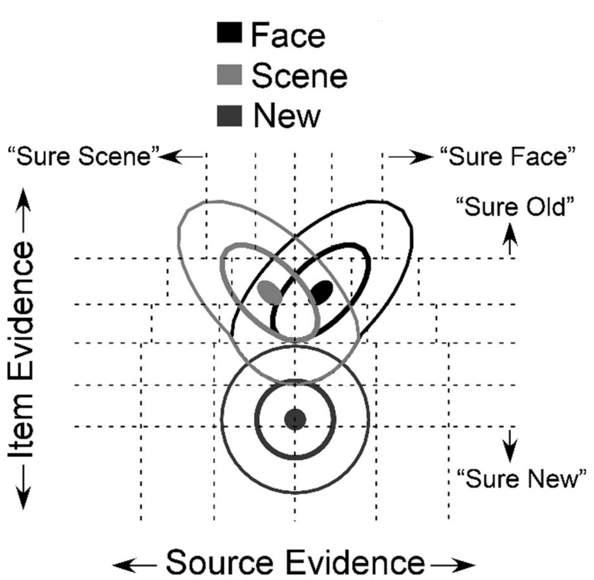

Figure 4. Decision heuristic account as a signal detection model of recognition source memory.

ROCs are a tool for data analysis that define the relationship between the proportion of correct and incorrect responses across different levels of response bias (Macmillan & Creelman, 2005). For source memory, researchers designate one source as the "target" (e.g., face) and the other source as the "lure" (e.g., scene). For each point on the ROC plot, the x-axis value is the source false alarm rate (e.g., the proportion of "face" responses for words actually paired with a scene) and the y-axis value is the source hit rate (e.g., the proportion of "face" responses for words paired with a face). For ROCs generated from confidence ratings, different points on the function come from different ranges of the confidence scale. The leftmost point includes the response rates for only the most confidently remembered "target" response (e.g., *definitely face*). Moving towards the right, the second point includes the response rates for the most confidently remembered "target" response as well as the next most confidently remembered "target" response (e.g., *definitely face* and *probably face*). This continues until the point farthest to the right includes the response rates for all confidence responses, except for the most confidently remembered "lure" response (e.g., *definitely scene*). By converting the hit and false alarm response rates to z-scores, we create zROC functions, a transformation that has been used to develop and evaluate the predictions of signal detection-based models of memory (e.g., Ratcliff et al., 1992).

ROC predictions of the decision heuristic account

Figure 4 shows a signal detection model of recognition and source memory that implements the decision heuristic account in terms of converging source criteria. The *x*-axis represents a continuum of source memory ranging from evidence in favour of the "scene" source (left) to the "face" source (right). For example, a value on the far left of this continuum means that the participant strongly remembered seeing the test word with the picture of a scene that was used on the study list. The *y*-axis represents a continuum of item memory ranging from evidence in favour of a "new" response (bottom) to an "old" response (top). For example, a value at the low end of the continuum means that the test word was a very weak match to the participant's memory of the study phase, indicating that it likely was not on the study list. Readers can think of the evidence distributions as hills rising out from the plane defined by the plot axes, with the contours interpreted as an elevation map showing regions of high and low elevation (thick and thin lines, respectively). Here, "high elevation" means that a lot of items have joint recognition and source memory strength values near that point (i.e., high probability density), and each contour shows all points on the distribution with a probability density equal to a certain value.

Words studied with a scene or face tend to have higher item strength than words that were not seen on the study list. Words studied with a scene (face) tend to have source evidence on the scene (face) end of the continuum, and lure words tend to be near the middle of the source evidence continuum, indicating ambiguous source evidence. The model incorporates a relationship between source and item memory; that is, the source distributions for scene and face items overlap completely with very low item strength and diverge as item strength increases. As a consequence, the ability to accurately discriminate sources is zero for weak item memories and increases for increasingly strong item memories.

The straight lines on the figure show criteria that map memory strength values to responses on the confidence rating scale. For example, items with source strength values on the far left get high-confidence "scene" responses and values on the far right get high-confidence "face" responses. Here, the decision heuristic account is implemented by "pulling in" the criteria on the source dimension as confidence that the item was studied increases. For example, an item recognised with high confidence would get a high-confidence "scene" response even if it was not too far to the left of the continuum; that is, source confidence is high even though source memory is relatively ambiguous.

As noted in previous studies (SK16; Starns et al., 2013, 2014), this converging criteria mechanism predicts that source memory zROC slopes should systematically deviate from a value of 1 when strong items from one source are plotted against weak items from another source. For example, SK16 plotted two mixed-strength source zROC functions by varying the strength of the "target" source (i.e., male face). One function plotted *repeated* items paired with a male face against *non-repeated* items paired with a female face (i.e., Strong-Male-WeakFemale) and the other function plotted *non-repeated* items paired with a male face against *repeated* items paired with a female face (i.e., WeakMale-StrongFemale). The converging criteria model predicts that the zROC slope should be below 1 when the "target" source is stronger than the "lure" source and above 1 when the "target" source is weaker than the "lure" source. A number of studies have observed this pattern when both item and source memory is strengthened (e.g., SK16; Starns et al., 2013; Yonelinas, 1999), and SK16 showed that the pattern also holds when item memory is strengthened without an improvement in source memory.

Dual-process signal detection account of source zROC Slope effects

zROC slopes have traditionally been interpreted in terms of memory processes, such as the relative contribution of two distinct recognition memory systems (e.g., Yonelinas, 1994) or the relative variability of continuous distributions of memory strength (e.g., Hilford et al., 2002; Slotnick et al., 2000; Slotnick & Dodson, 2005). In particular, the effect of learning strength on source memory zROC slopes has been cited as support for the dual process signal detection (DPSD) model (Yonelinas, 1994; Yonelinas & Parks, 2007), and the specific strengthening manipulation used by SK16 was designed to distinguish the dual process and decision heuristic accounts. Thus, we will also consider the implications of our results for the DPSD model.

The DPSD model assumes that memory decisions can be based on one of two independent memory systems, familiarity and recollection. Familiarity is characterised as a continuous equal-variance signal detection process. For source memory, items from either source have variable memory strengths measured on a continuous scale, which form separate probability distributions used to inform an individual's judgment. In contrast, recollection is a discrete threshold process, such that a participant either successfully recollects details from an event or fails to recollect any details at all. Importantly, recollection is associated with high-confidence responses that improve the hit rate without eliciting changes in the false alarm rate (i.e., false recollection does not occur in the Yonelinas (1999) model).

The DPSD model predicts that zROC slopes are based on the relative contribution of recollection and familiarity, and that changing the relative contributions can lead to slopes that deviate from 1. Yonelinas (1999) used a source memory task to test this prediction. Participants heard words organised in two lists, one spoken in a man's voice the other in a woman's voice. The first list was studied twice and the second only once, creating a difference in learning strength. Yonelinas assumed that familiarity-based memory strengths for both sources would be similar, owing to a trade-off between frequency (i.e., the first list source was repeated) and recency (i.e., the second list source was presented immediately prior to test). Therefore, familiarity signals should be less useful for mnemonic source judgments, forcing participants to rely on recollection alone. Results showed that additional learning produced source zROC slopes that deviated from 1, in line with the DPSD prediction (slope = .75).[3] Other studies have since replicated the effect of mixed-strength sources on source zROC slopes (e.g., Starns et al., 2013).

SK16 noted that the DPSD model and the converging criteria model predict the same direction of effect on zROC slopes when sources differ in strength, but they do so based on different mechanisms. The DPSD model attributes the effect to an increase in source recollection for stronger items, whereas the converging criteria model attributes it to an increased willingness to make high-confidence source responses for stronger items. Their different-source repetition manipulation was designed to distinguish these potential mechanisms by strengthening item memory without improving source memory, which should eliminate the slope effect if it is based on enhanced source recollection but not if it is based on lax source confidence criteria for items that are high in item strength. SK16's results showed that different-source repetition produced the slope effect, which they interpreted as support for the converging criteria model.

The zROC analyses for the current experiments allow us to determine if this result replicates and also to evaluate the zROC slope effect when item-strengthening manipulations have less influence on source memory, especially for Experiment 2. Although source memory is thought to rely primarily on recollection (Yonelinas, 1999), the DPSD model could technically produce the slope effect observed by SK16 if item strengthening selectively impaired source *familiarity* without affecting source recollection. By better equating source performance for strengthened and unstrengthened items, we can achieve a clearer test of the DPSD model. Specifically, if item strengthening has minimal effects on source performance, then the DPSD model cannot predict a zROC slope effect by assuming a selective effect on the familiarity component of source memory (sparing recollection). In other words, the DPSD model will predict no slope effect, but the converging criteria model predicts the same slope effect observed in SK16, as the latter model attributes the effect to differences in source decision making produced by a change in *item* memory strength.

Table 2. Priors on hyperparameters for across-participant distributions of SDT parameters.

Label	Description	Prior distribution
μ_{av}	Across-participant mean of *av*	Gaussian with mean of 1 and SD of 1 truncated below 0
μ_{se}	Across-participant mean of *se*	Gaussian with mean of 0 and SD of 0.5
$\mu_{log(\sigma)}$	Across-participant mean of log(σ)	Gaussian with mean of 0 and SD of 0.5
μ_c	Across-participant mean of *c*	Gaussian with mean of 0 and SD of 1
μ_{d1}	Across-participant mean of *d1*	Gaussian with mean of 0 and SD of 0.5 truncated below 0
μ_{d2}	Across-participant mean of *d2*	Gaussian with mean of 0 and SD of 0.5 truncated below 0
μ_{d3}	Across-participant mean of *d3*	Gaussian with mean of 0 and SD of 0.5 truncated above 0
μ_{d4}	Across-participant mean of *d4*	Gaussian with mean of 0 and SD of 0.5 truncated above 0
σ_{av}	Across-participant SD of *av*	Gaussian with mean of 0.25 and SD of 0.5 truncated below 0.1
σ_{se}	Across-participant SD of *se*	Gaussian with mean of 0.25 and SD of 0.5 truncated below 0.1
$\sigma_{log(\sigma)}$	Across-participant SD of log(σ)	Gaussian with mean of 0.1 and SD of 0.25 truncated below 0.05
σ_c	Across-participant SD of *c*	Gaussian with mean of 0.25 and SD of 0.5 truncated below 0.1
σ_{d1}	Across-participant SD of *d1*	Gaussian with mean of 0.25 and SD of 0.5 truncated below 0.1
σ_{d2}	Across-participant SD of *d2*	Gaussian with mean of 0.25 and SD of 0.5 truncated below 0.1
σ_{d3}	Across-participant SD of *d3*	Gaussian with mean of 0.25 and SD of 0.5 truncated below 0.1
σ_{d4}	Across-participant SD of *d4*	Gaussian with mean of 0.25 and SD of 0.5 truncated below 0.1

Note: SD = standard deviation.

Modelling methods

We used a hierarchical Bayesian approach to fit a Gaussian signal detection model to the observed source zROC functions. We fit a Gaussian signal detection model to the empirical zROC functions using hierarchical Bayesian methods. Models were fit using JAGS software (Plummer, 2003) to take Markov Chain Monte Carlo (MCMC) samples from the joint posterior distribution across all parameters.

The model assumes that source retrieval produces a continuous strength value representing how well the remembered information matches one source versus the other. A strength value of zero represented completely ambiguous information, with values farther above or below zero representing increasingly strong evidence for Source 1 or Source 2, respectively. Variation in source strength across test trials was represented by Gaussian distributions. The source distributions were symmetrical around zero, with means at μ and -μ. Thus, the distance between distributions was 2μ. The model had separate μ parameters for weak and strong items (i.e., items that were only studied in the source encoding blocks or items that were studied in both the source and recall encoding blocks). The model also had a free parameter for the standard deviation for strong items, σ, with the standard deviation for weak items fixed at a value of 1. Within a strength category, the distributions for face and scene items had equal standard deviations. Finally, the model had parameters for the position of 5 response criteria to map strength values onto the 6 confidence responses from high-confidence "face" responses to high-confidence "scene" responses. The criterion separating "face" and "scene" responses, *c*, was allowed to vary from zero, meaning that the model could accommodate biased responding (e.g., a participant with an overall preference for responding "scene"). In addition to *c*, the model had 4 criterion deviation parameters, *d1*–*d4*, for the distance between adjacent criteria. *d1* and *d2* could only take positive values, and the two confidence criteria above the "face"/"scene" criterion were placed at $c + d1$ and $c + d1 + d2$, respectively. *d3* and *d4* could only take negative values, and the two confidence criteria below the "face"/"scene" criterion were placed at $c + d3$ and $c + d3 + d4$, respectively.

The model described above applies at the individual-participant level, so predicted responding for each participant was based on 8 parameters: μ_{WEAK}, μ_{STRONG}, σ, *c*, *d1*, *d2*, *d3*, and *d4*. Each participant was allowed unique values for all parameters, but the model also used across-participant distributions to serve as prior distributions for the participant-level parameter estimates. The across-participant distributions of μ_{WEAK} and μ_{STRONG} were defined indirectly with a Gaussian distribution for the average μ value across the two strength conditions (*av*) and the strength effect (*se*) for μ, where $\mu_{WEAK} = av - .5 \times se$ and $\mu_{STRONG} = av + .5 \times se$. The model also used Gaussian distributions to represent across-participant variation in *c* and the log of σ. The model used Gaussian distributions truncated below 0 for *d1* and *d2* and Gaussian distributions truncated above zero for *d3* and *d4*. These across-participant distributions each had a mean and standard deviation, so there were 16 total hyperparameters that were estimated along with the individual-participant parameters. We used loosely informative prior distributions for these hyperparameters, as reported in Table 2.

We addressed two primary research questions with the model results: (1) How did the item-strengthening manipulation affect source memory? and (2) What was the difference in zROC slopes between the "Strong Face – Weak Scene" and "Weak Face – Strong Scene" functions? For the first question, we assessed the difference in source d' values between strong and weak items. For a given participant, the weak d' value is $2\mu_{WEAK}$ and the strong d' value is $2\mu_{STRONG}/\sigma$. To represent overall performance, we applied these equations using hyperparameters representing average parameter values across participants instead of the parameters for an individual participant. We created distributions of uncertainty by doing this for each MCMC sample. For the second question, we calculated zROC slopes using the σ parameters. For a given participant, σ is the ratio of the standard deviation in source evidence for weak and strong items, so the cross-strength functions have slopes of σ and $1/\sigma$, making the slope effect $\sigma - 1/\sigma$. To represent overall performance, we applied this formula using the hyperparameter representing the average σ value across participants.[4] We did this for every MCMC sample to create a posterior distribution of overall slope differences.

Modelling results

Throughout, we use median of the posterior distribution as a point estimate and 95% Highest Density Intervals (HDIs) to characterise the range of uncertainty. As in SK16, we evaluated both pure-strength and mixed-strength source zROCs. In reporting results, we always defined "face" as the target source and scene as the lure source, so the hit rate was the proportion of face items reported as "face" and the false alarm rate was the proportion of scene items reported as "face". For pure-strength zROCs, the functions were based on face and scene items that were either both studied once (weak) or both studied twice (strong). For the mixed-strength zROCs, the functions were based on face items from one strength condition and scene items from the other.

Experiment 1 zROC analyses

The left panel of Figure 5 shows the pure-strength zROC functions from Experiment 1. The strong function is slightly closer to the positive diagonal defining chance performance, suggesting that the item-strengthening manipulation slightly impaired participants' ability to discriminate the sources. The model estimated d' values of 1.01 and 0.83 for weak and strong item learning, respectively. Figure 6 shows the posterior distribution for the difference in the model's d' estimates in the weak and strong conditions.[5] The median of the posterior distribution was −.175 and the 95% HDI was [−.304, −.057]. Because zero is not included in the HDI, we cannot consider it a credible value and thus we can infer that strengthening item memory impaired source discrimination. Thus, the Bayesian modelling results confirmed the previous frequentist analysis.

The mixed-strength zROC functions had a lower slope when the target source had stronger learning (.93) than when the lure source had stronger learning (1.07; Figure 5, right). This is the direction of effect predicted by the converging criteria model, but the slope difference was subtle. Figure 7 shows the posterior distribution for the slope difference ("Weak Face - Strong Scene" minus "Strong Face – Weak Scene"). The median of the slope difference posterior distribution was .137 and the 95% HDI was [.017, .255]. We cannot consider zero a credible value because it is not included in the HDI. Therefore, we can infer that the mixed-strength source zROC slope effect was present and that strengthening item memory affected source zROC slopes as predicted by the converging criteria account, although the slope difference was small.

Experiment 2 zROC analyses

Figure 8 shows the zROCs from Experiment 2. The pure-strength functions (left panel) showed no evidence that the item-strengthening manipulation affected source memory, as the functions overlapped nearly completely. The model estimated d' values of .93 and .82 for weak and strong item learning, respectively. Figure 9 shows the posterior distribution for the difference in source memory d' between strengthened and unstrengthened items. The posterior distribution had median value of -.109 and a 95% HDI of [−.264, .063]. Zero is included in the HDI, and anything other than a small effect is ruled out by the interval. This implies that strengthening item memory did not produce a substantial decrease in source discrimination.

The mixed-strength zROC functions (Figure 8, right) showed the slope effect predicted by the converging criteria model, and the effect was larger than the one observed in Experiment 1. The slope estimate was 0.89 when the target source had strong item learning and 1.13 when the lure source had strong item learning.

We again assessed the slope effect with the posterior distribution for the difference in slope between the two-mixed strength zROC functions (i.e., slope $_{\text{Weak Face vs. Strong Scene}}$ – slope $_{\text{Strong Face-Weak Scene}}$). The median of the slope difference posterior distribution was .244 and the 95% HDI was [.06, .429] (Figure 10). Zero did not fall within the HDI and is not considered a credible value. Thus, results again showed the slope effect predicted by the converging criteria account.

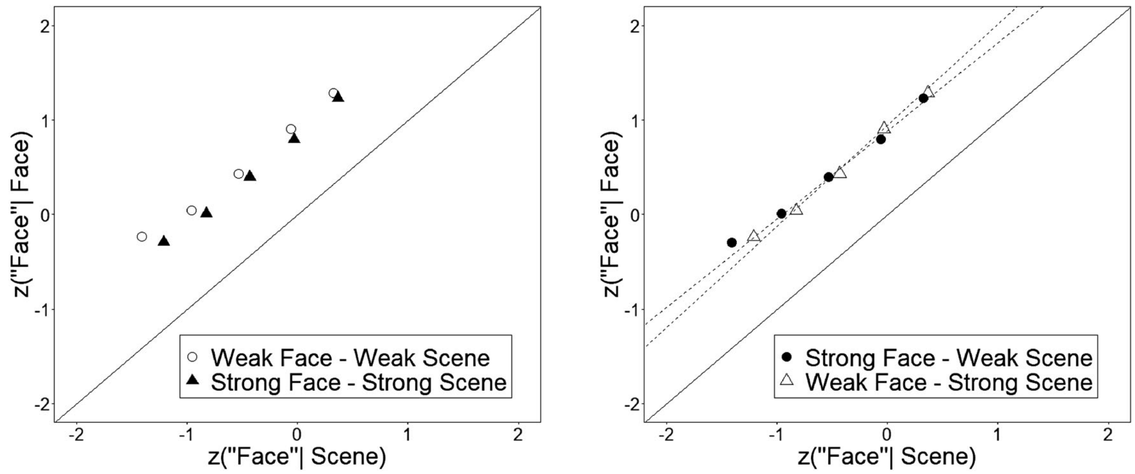

Figure 5. Source zROC functions from Experiment 1. Pure-strength functions (left) were plotted using face and scene items from the same item-strength condition (i.e., word items appearing only once in encoding block Type B [weak items; unfilled circles] or word items appearing in both encoding blocks [strong items; filled triangles]). The same data were used to plot mixed-strength functions (right) with strong items from one source plotted against weak items from the other source. In both panels, functions with filled symbols feature strong face items, functions with unfilled symbols feature weak face items, functions with triangles feature strong scenes, and functions with circles feature weak scenes. The displayed functions are not based on direct fits to the observed zROC points; rather, they are based on average parameter values from the hierarchical model that was fit to the individual-participant data.

Collectively, the zROC results replicate SK16 and provide further support for the converging criteria account. Moreover, the effect held even with no appreciable difference in source memory between strengthened and unstrengthened items (Experiment 2), which provides evidence against the idea that the slope effect is produced by a selective effect on source familiarity, as opposed to source recollection, in the DPSD model. Thus, the current results expand the support for the converging criteria model over the DPSD model first reported by SK16.

General discussion

In two experiments, we explored the effect of strengthening memory for item information on confidence in source memory judgments. The primary goal was to test a decision heuristic account whereby strong item memory makes participants more willing to report high confidence in their source responses even if they do not clearly remember source details. SK16 reported evidence for this account by showing that strengthening item memory inflated confidence in source errors, but in their dataset SK16 found no effect on confidence in correct source responses. The decision heuristic account predicts that item strength should inflate source confidence for both

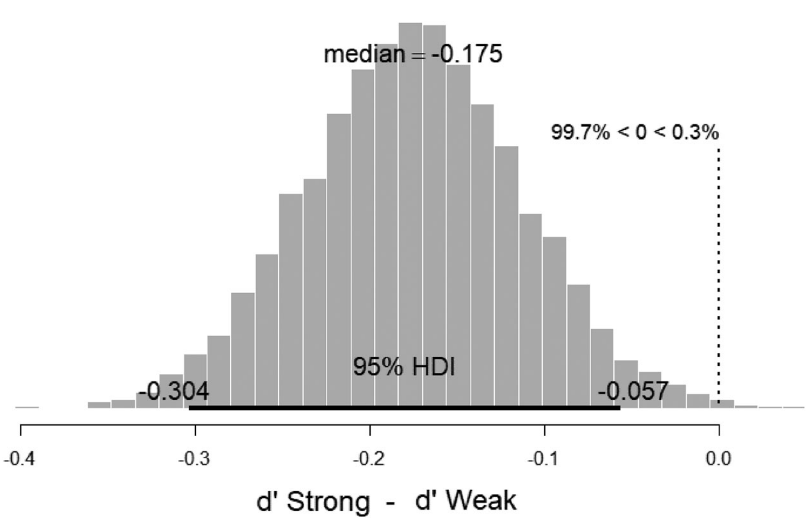

Figure 6. Posterior distribution for the difference between strong and weak mean d′ estimates in Experiment 1. The comparison value of zero (dashed vertical line) is accompanied with the percentage of the MCMC sample that falls below and above the comparison value. The bold horizontal line reflects the 95% highest density interval (HDI).

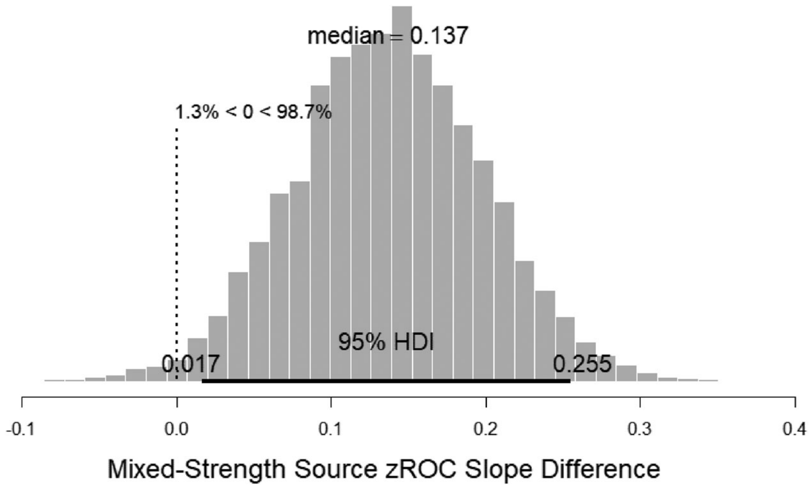

Figure 7. Posterior distribution for the slope difference between source zROC mixed-strength functions in Experiment 1. The comparison value of zero (dashed vertical line) is accompanied with the percentage of the MCMC sample that falls below and above the comparison value. The bold horizontal line reflects the 95% highest density interval (HDI).

correct and errors responses. However, the item strengthening manipulation used by SK16 also impaired source memory, which might have counteracted an increase in source confidence for correct responses. The current studies used new item-strengthening manipulations designed to minimise impacts on source discrimination by decreasing the similarity between the source-encoding and item-strengthening trials (Experiments 1 and 2) and by having the source-encoding trials precede the item-strengthening trials (Experiment 2).

The Experiment 1 results showed that the item-strengthening manipulation produced a detectable decrease in source discriminability, but the effect was smaller than the one observed in the SK16 experiments. Item strengthening increased source confidence for incorrect responses with no effect on correct responses. Thus, the results supported the decision heuristic account by showing an overall increase in source confidence, but did not show an effect for both error and correct responses. As in the SK16 experiment, the effect on correct source confidence might have been counteracted by a reduction in source discriminability. In Experiment 2, item strengthening had a negligible effect on source accuracy, and inferential tests could not rule out a null result. So we succeeded in finding a "purer" manipulation of item strength with minimal side effects on source discrimination. Under these conditions, the item strengthening manipulation inflated source confidence for both correct and erroneous source decisions. Thus, results provided strong support for the decision heuristic account.

We also assessed correlational evidence for the decision heuristic account in terms of the relationship between item and source confidence. As in previous studies (e.g., Starns et al., 2013), we found that participants were more likely to make high confidence source responses for lures that were falsely recognised with higher levels of item confidence. We also found a similar pattern for targets that were correctly recognised but attributed to the incorrect source. Participants were more likely to report high confidence in these source errors if the target was recognised with higher item confidence. Together with the results of the item strengthening manipulation, these results provide converging evidence that item strength inflates source confidence, even in the absence of accurate source retrieval (i.e., for lures not studied in any source and targets attributed to the wrong source).

Finally, we also evaluated the effect of item strengthening on source zROC slopes to further test the decision heuristic account and compare it to a dual process account. By applying a Gaussian signal detection model in a hierarchical Bayesian format, we showed that strengthening item memory affected zROC slopes as predicted by the decision heuristic account. That is, the source memory zROC slope was high when the source defined as the "target" had stronger item memory than the alternative source, and the slope was low when the source defined as the "target" had weaker item memory than the alternative source. This result is difficult to accommodate in a dual process account, which attributes changes in zROC slope to differences in the relative contribution of familiarity versus recollection on source discrimination. This is especially true for Experiment 2, for which item strengthening had no discernible impact on source discrimination but still affected source zROC slopes.

Our results suggest that multiple factors contribute to source memory impairments produced by item-strengthening manipulations, as seen in SK16 and our Experiment 1. We speculated that the source impairments in SK16 could be attributed to two factors: source interference and reduced attention to repeated presentations of the same word. Experiment 1 minimised source interference by strengthening item memory with a recall task that did not include any picture sources like the ones on source

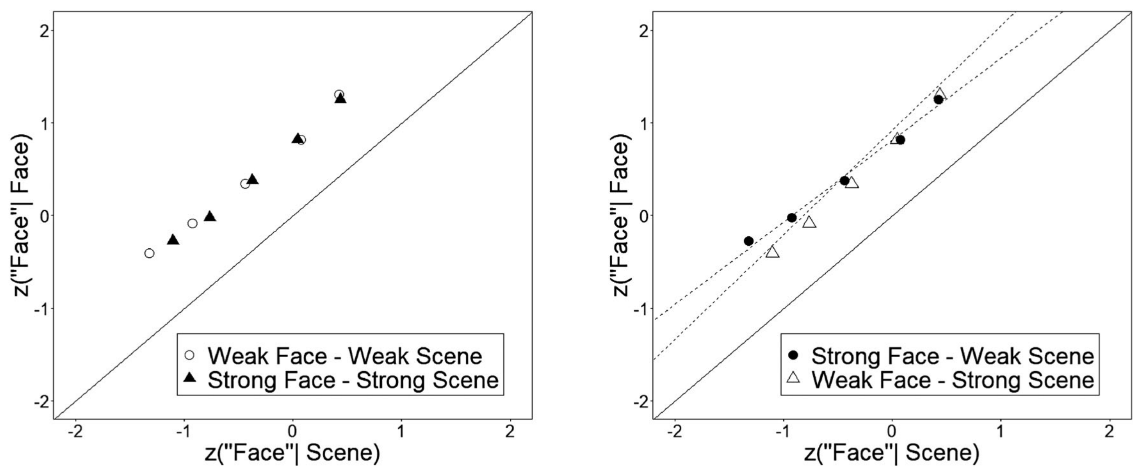

Figure 8. Source zROC functions from Experiment 2. Pure-strength functions (left) were plotted using face and scene items from the same item-strength condition (i.e., word items appearing only once in encoding block Type B [weak items; unfilled circles] or word items appearing in both encoding blocks [strong items; filled triangles]). The same data were used to plot mixed-strength functions (right) with strong items from one source plotted against weak items from the other source. In both panels, functions with filled symbols feature strong face items, functions with unfilled symbols feature weak face items, functions with triangles feature strong scenes, and functions with circles feature weak scenes. The displayed functions are not based on direct fits to the observed zROC points; rather, they are based on average parameter values from the hierarchical model that was fit to the individual-participant data.

encoding trials, but this experiment maintained the SK16 practice of always having the item-strengthening trial precede the source-encoding trial for a given word. This attenuated source decrements relative to SK16, but still produced a detectable effect on source discrimination. Experiment 2 maintained the encoding procedures of Experiment 1 but flipped the order such that the source-encoding trial for a given word always preceded the item-strengthening trial. This procedure was more successful in minimising source-impairment side effects, and results were consistent with either no decrement or a very small decrement. The potential influence of attention here highlights the importance of considering additional cognitive processes when evaluating memory findings and designing manipulations of memory strength. We recommend that future experiments investigating the effects of item strength on source memory both make source and item trials as distinct as possible and give participants the opportunity to encode source details associated with an item before it appears in the item-strengthening task.

Beyond the zROC results, the high-confidence source error findings that we report are also inconsistent with the standard DPSD model. The DPSD model associates recollection with increases in high-confidence *correct* source responses, but not source *errors*, and therefore cannot explain how strengthening item memory, which

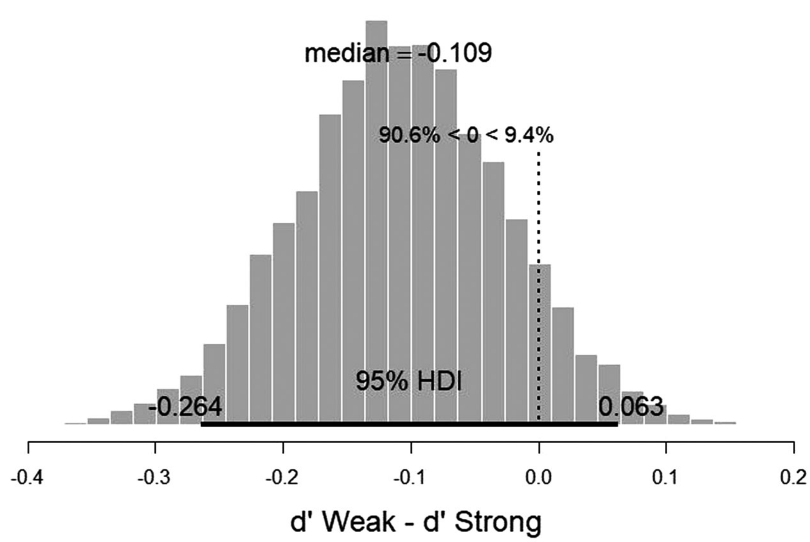

Figure 9. Posterior distribution for the difference between strong and weak mean d′ estimates in Experiment 2. The comparison value of zero (dashed vertical line) is accompanied with the percentage of the MCMC sample that falls below and above the comparison value. The bold horizontal line reflects the 95% highest density interval (HDI).

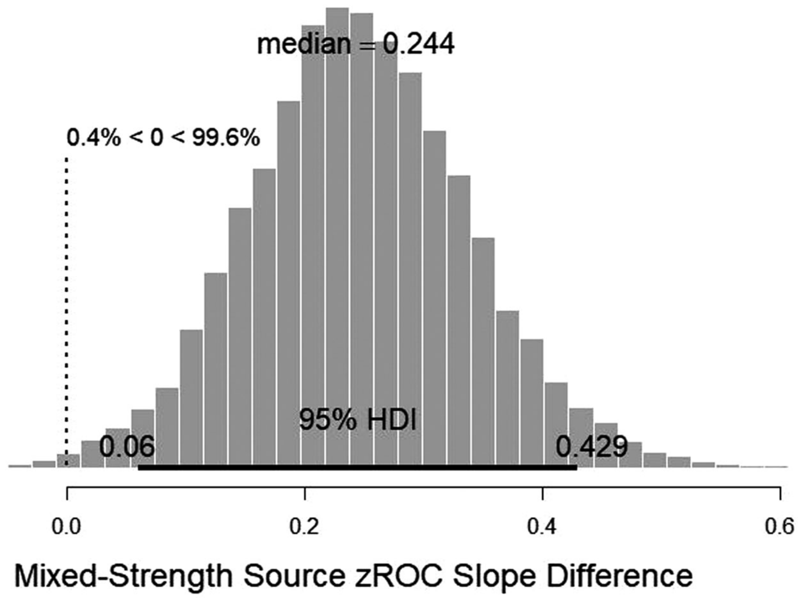

Figure 10. Posterior distribution for the slope difference between source zROC mixed-strength functions in Experiment 2. The comparison value of zero (dashed vertical line) is accompanied with the percentage of the MCMC sample that falls below and above the comparison value. The bold horizontal line reflects the 95% highest density interval (HDI).

is assumed to increase the contribution of recollection, would also increase participants' confidence in source errors. In contrast, the converging criteria model predicts that confidence would increase for source errors because high item strength is associated with lax criteria for making a high-confidence source response, and this mechanism is not dependent on any other mechanism that affects source accuracy, like recollection.

One critique of the conclusion made here may be that we have failed to consider fully the influence of a source memory decrement and that such a decrement may be required to produce the slope effect when *only* item memory is strengthened. For instance, although both frequentist and Bayesian analyses of Experiment 2 did not reveal a statistically significant impairment, there still remained a numeric decrease in source memory performance between weak and strong items as measured by d'. In combination with the impairment found in our Experiment 1 and both experiments of SK16, it may appear that we have selectively focused on the exception among these results. To explore the possible influence of source memory impairments on the slope effect, we plotted the group-level estimates of these two values across our two experiments and the two experiments of SK16 (Figure 11). Although SK16 demonstrated greater source memory deficits and slope effects, relative to our experiments, a comparison of their two experiments revealed similar slope effects (Exp1 = .30; Exp2 = .31) despite differences in the size of the source impairment (Exp1 = .47, Exp2 = .39). Further, we observed a slope effect in both of our experiments with a much smaller source impairment than SK16, and a comparison of our experiments showed that the experiment with the smaller source impairment (Exp1 = .14 vs. Exp2 = .07) actually had a larger slope effect (Exp1 = .11 vs. Exp2 = .25). Therefore, there was no indication of a relationship between these two factors that would discredit the converging criteria account of the slope effect in favour of mechanisms related to source memory deficits.

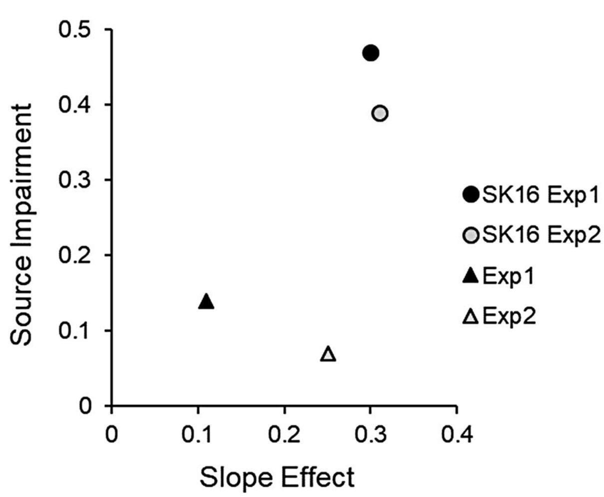

Figure 11. The relationship between source memory performance and the slope effect. Source memory impairment (y-axis), as measured by the difference in d′ between weak and strong items, was plotted against slope effect (x-axis), as measured by the difference between mixed-strength source zROC slopes, for Experiment 1 (black) and Experiment 2 (grey) of the current study (triangles) and of SK16 (circles).

As mentioned in the Introduction, the source of a memory can be defined in reference to a wide range of dimensions, such as location, presenter, or even internal cognitive and emotional states (Johnson et al., 1993). Although it is standard in laboratory studies to use two source categories, a source decision can involve any number of categories, such as when one attempts to remember which news outlet carried a story. We expect that the relationship between item strength and source confidence noted herein will generally hold across many different versions of a source task, but future studies will be required to demonstrate this empirically. Past studies suggest that many source memory phenomena are robust to changes in source dimension (e.g., Hicks & Marsh, 2001).

A number of possible mechanisms could underlie the relationship between item and source confidence, ranging from simple response tendencies (e.g., a predilection for reusing response options) to basic properties of memory retrieval. This list includes mechanisms consistent with the claim that high item strength affects source responding in itself (like the converging criteria model) and mechanisms that propose that source judgments are instead affected by the feeling of confidence conferred by high item strength, or even the overt response action promoted by high item strength. Regardless of mechanism, it is important for memory researchers to consider the decision heuristic account; for example, researchers would not want to interpret zROC patterns as evidence for threshold recollection if they are in fact produced by decision heuristics. Our design helps to rule out some of the most mundane mechanisms for the link between item and source confidence, as we had separate test lists for these two judgments. Thus, it is unlikely that the relationship reflects a simple tendency to repeat response categories, as the preceding response for a given source judgment was a source judgement for a separate item, not the item judgment for the same item (that response was made on a previous test list). Of course, it is possible that participants remember their previous item response before making a source response.

A recent study by Fox and Osth (2022) provides stronger evidence that item strength itself impacts source judgments. They used a reversed paradigm in which the source test preceded the item test. Thus, source judgments could not be affected by previous item judgments, but it is still reasonable to expect that items recognised with high confidence on the subsequent recognition test tended to have stronger item memories than items not recognised with high confidence. The goal of the original study was to assess source discrimination for unrecognised items, but the data can also be used to explore the relationship between item and source confidence. Results from this analysis show the same relationship between item strength and source confidence predicted by the decision heuristic account, and this holds even when analyses are limited to source errors. Specifically, participants reported high confidence in their source errors 11% of the time for weak item memories versus 26% of the time for strong item memories. Although not definitive, these results suggest that high item memory directly promotes high confidence source responding with no need for mediation via an explicit item judgment.[6]

As noted in the Introduction, the item-memory misattribution model (IMM; Dobbins & McCarthy, 2008) proposes that people will sometimes claim to have studied an item in a specific source based on item memory even when they fail to retrieve specific source details. For the more standard task of selecting between alternative sources, one could potentially extend this account by claiming that high item strength is misattributed to source evidence for whichever source the participant selects on a given trial. Functionally, this would increase the variability of source evidence as item strength increases. That is, slight evidence for Source 1 would be perceived as strong evidence for Source 1 when high item strength is misattributed to source, and the same is true for Source 2. Notably, compressing the confidence criteria is mathematically equivalent to increasing the variability of evidence distributions (Starns et al., 2013). Thus, misattributing item evidence as source evidence is a potential mechanism for the decision heuristic account, and indeed, some versions of a misattribution account could be mimicked by converging confidence criteria.

In sum, the current experiments support the converging criteria account over the DPSD account by upholding three specific predictions of the former account. First, when we eliminated interfering factors, we observed an increase in the proportion of high-confidence responses for both correct and error source responses. Second, we replicated a dependency between confidence for item recognition errors (i.e., false alarms) and later source confidence. Third, we observed an effect of item strength on source zROC slopes even when the strength manipulation had no effect or a slight negative effect on source performance.

At a more general level, our findings highlight the dissociation between source confidence and accuracy. Returning to the issue of misinformation, our results indicate that strong memory for the content of an article will encourage people to express high levels of confidence for their memory of the source of the article, even if they misremember the source itself. It is important to consider the factors that influence source confidence in the fight against misinformation, because high levels of confidence in memory for the source of a piece of information could influence an individual's decision to share it with others. Our findings suggest that researchers cannot develop a full understanding of source judgments unless they account for how memory for source interacts with memory for the item itself.

Notes

1. We had to remove three participants from the analyses of the proportion of high-confidence source responses: two participants never made a high-confidence source response, regardless of item strength or source accuracy, and one participant

never made a high-confidence source response for any weak items.

2. We had to remove five participants from the analyses of the proportion of high-confidence source responses: one participant never made a high-confidence source response, regardless of item strength or source accuracy; one participant never made a high-confidence source error; one participant never made a high-confidence error for weak items nor a high-confidence correct response for strong items; and two participants never made a high-confidence source error for strong items.
3. As mentioned earlier, a source memory zROC requires designating one source as the "target" and the other source as the "lure". The direction of the slope deviation (i.e., whether the slope is greater than or less than 1) depends on which source has been designated as the "target". Yonelinas (1999; Experiment 3) designated the *stronger* source (List 1) as the "target" source. Therefore, a slope *less* than 1 captured a boost in high-confidence correct responses for the stronger source. However, if instead they designated the *weaker* source (List 2) as the "target" when plotting the source zROC, a slope *greater* than 1 would reflect a boost in high-confidence correct responses for the stronger source.
4. The model had an across participant distribution on log(σ), so we exponentiated the hyperparameter for the mean of this distribution to get the overall raw sigma.
5. We plotted all posterior distributions using the *plotPost* function, which is defined in Kruschke's (2015) DBDA2E-utilities.R, a collection of utility functions for performing Bayesian analysis in R (R Core Team, 2016).
6. We do not make the strong claim that item and source judgments make reference to the same memory representations. Retrieving strong item information could impact source decision processes even if item and source representations are stored and retrieved separately.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the NSF (Award #1454868 and #1554871).

Data availability statement

All stimuli, data, and analysis code publicly available at the following OSF website: https://osf.io/g3dyf/.

ORCID

D. Merika W. Sanders http://orcid.org/0000-0001-5213-7234

References

Bavel, J. J. V., Baicker, K., Boggio, P. S., Capraro, V., Cichocka, A., Cikara, M., Crockett, M. J., Crum, A. J., Douglas, K. M., Druckman, J. N., Drury, J., Dube, O., Ellemers, N., Finkel, E. J., Fowler, J. H., Gelfand, M., Han, S., Haslam, S. A., Jetten, J., … Willer, R. (2020). Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour*, *4*(5), 460–471. https://doi.org/10.1038/s41562-020-0884-z

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

DeCarlo, L. T. (2003). Source monitoring and multivariate signal detection theory, with a model for selection. *Journal of Mathematical Psychology*, *47*(3), 292–303. https://doi.org/10.1016/S0022-2496(03)00005-1

Dobbins, I. G., & McCarthy, D. (2008). Cue-framing effects in source remembering: A memory misattribution model. *Memory & Cognition*, *36*(1), 104–118. https://doi.org/10.3758/MC.36.1.104

Fox, J., & Osth, A. F. (2022). Does source memory exist for unrecognized items? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(2), 242–271. https://doi.org/10.1037/xlm0001111

Hicks, J. L., & Marsh, R. L. (2001). False recognition occurs more frequently during source identification than during old–new recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(2), 375–383. https://doi.org/10.1037/0278-7393.27.2.375

Hilford, A., Glanzer, M., Kim, K., & DeCarlo, L. T. (2002). Regularities of source recognition: ROC analysis. *Journal of Experimental Psychology: General*, *131*(4), 494–510. https://doi.org/10.1037/0096-3445.131.4.494

Hughes, T., Smith, J., & Leavitt, A. (2018). *Helping people better assess the stories they see in news feed with the context button*. Facebook Newsroom. https://about.fb.com/news/2018/04/news-feed-fyi-more-context/

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*(1), 3–28. https://doi.org/10.1037/0033-2909.114.1.3

Kim, K., Yi, D.-J., Raye, C. L., & Johnson, M. K. (2012). Negative effects of item repetition on source memory. *Memory & Cognition*, *40*(6), 889–901. https://doi.org/10.3758/s13421-012-0196-2

Klauer, K. C., & Kellen, D. (2010). Toward a complete decision model of item and source recognition: A discrete-state approach. *Psychonomic Bulletin & Review*, *17*(4), 465–478. https://doi.org/10.3758/PBR.17.4.465

Kruschke, J. K. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd ed.). Academic Press.

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menezer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. *Science*, *359*(6380), 1094–1096. https://doi.org/10.1126/science.aao2998

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database : A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*, 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Lawrence Erlbaum Associates.

Marsh, R. L., & Hicks, J. L. (1998). Test formats change source-monitoring decision processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(5), 1137–1151. https://doi.org/10.1037/0278-7393.24.5.1137

Onyper, S. V., Zhang, Y. X., & Howard, M. W. (2010). Some-or-none recollection: Evidence from item and source memory. *Journal of Experimental Psychology: General*, *139*(2), 341–364. https://doi.org/10.1037/a0018926

Plummer, M. (2003). *Jags: A program for analysis of Bayesian graphical models using Gibbs sampling*. Proceedings of the 3rd International Workshop on Distributed Statistical Computing.

Postman, L. (1971). Transfer, interference, and forgetting. In J. W. Kling, & L. A. Riggs (Eds.), *Woodworth and schlosberg's experimental psychology* (3rd ed., pp. 1019–1132). Holt, Rinehart & Winston.

Ratcliff, R., Sheu, C.-F., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review*, *99*(3), 518–535. https://doi.org/10.1037/0033-295X.99.3.518

R Core Team. (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/.

Ross, D. A., Sadil, P., Wilson, D. M., & Cowell, R. A. (2018). Hippocampal engagement during recall depends on memory content. *Cerebral Cortex*, *28*(8), 2685–2698. https://doi.org/10.1093/cercor/bhx147

Samek, G. (2018). *Greater transparency for users around news broadcasters*. YouTube Official Blog. https://blog.youtube/news-and-events/greater-transparency-for-users-around.

Schacter, D. L. (1999). The seven sins of memory: Insights from psychology and cognitive neuroscience. *American Psychologist*, *54*(3), 182–203. https://doi.org/10.1037/0003-066X.54.3.182

Slotnick, S. D., & Dodson, C. S. (2005). Support for a continuous (single-process) model of recognition memory and source memory. *Memory & Cognition*, *33*(1), 151–170. https://doi.org/10.3758/BF03195305

Slotnick, S. D., Klein, S. A., Dodson, C. S., & Shimamura, A. P. (2000). An analysis of signal detection and threshold models of source memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1499–1517. https://doi.org/10.1037/0278-7393.26.6.1499

Starns, J. J., & Hicks, J. L. (2008). Context attributes in memory are bound to item information, but not to one another. *Psychonomic Bulletin & Review*, *15*(2), 309–314. https://doi.org/10.3758/PBR.15.2.309

Starns, J. J., & Hicks, J. L. (2013). Internal reinstatement hides cuing effects in source memory tasks. *Memory & Cognition*, *41*(7), 953–966. https://doi.org/10.3758/s13421-013-0325-6

Starns, J. J., & Ksander, J. C. (2016). Item strength influences source confidence and alters source memory zROC slopes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(3), 351–365. https://doi.org/10.1037/xlm0000177

Starns, J. J., Pazzaglia, A. M., Rotello, C. M., Hautus, M. J., & Macmillan, N. A. (2013). Unequal-strength source zROC slopes reflect criteria placement and not (necessarily) memory processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(5), 1377–1392. https://doi.org/10.1037/a0032328

Starns, J. J., Rotello, C. M., & Hautus, M. J. (2014). Recognition memory zROC slopes for items with correct versus incorrect source decisions discriminate the dual process and unequal variance signal detection models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1205–1225. https://doi.org/10.1037/a0036846

Underwood, B. J. (1949). Proactive inhibition as a function of time and degree of prior learning. *Journal of Experimental Psychology*, *39*(1), 24–34. https://doi.org/10.1037/h0059550

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. (1999). The contribution of recollection and familiarity to recognition and source-memory judgments: A formal dual-process model and an analysis of receiver operating characterstics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(6), 1415–1434. https://doi.org/10.1037/0278-7393.25.6.1415

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

Zaragoza, M. S., & Lane, S. M. (1994). Source misattributions and the suggestibility of eyewitness memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 934–945. https://doi.org/10.1037/0278-7393.20.4.934

Zaragoza, M. S., & Mitchell, K. J. (1996). Repeated exposure to suggestion and the creation of false memories. *Psychological Science*, *7*(5), 294–300. https://doi.org/10.1111/j.1467-9280.1996.tb00377.x