CrossMark

Confidence carryover during interleaved memory and perception judgments

Justin Kantner [1] · Lisa A. Solinger [2] · David Grybinas [2] · Ian G. Dobbins [2]

Abstract

Recognition memory tests typically consist of randomly intermixed studied and nonstudied items that subjects classify as *old* or *new*, often while indicating their confidence in these classifications. Under most decision theories, confidence ratings index an item's memory strength—the extent to which it elicits evidence of prior occurrence. Because the test probes are randomly ordered, these theories predict that confidence judgments should be sequentially independent: confidence on trial n should not predict confidence on $n + 1$. However, analysis of two extant data sets demonstrated reliable serial correlations in recognition memory confidence (confidence carryover). In a new experiment, we examined the domain specificity of confidence carryover by serially interleaving recognition and perceptual classification judgments. Analysis revealed domain-general and domain-specific confidence carryover effects: The confidence of a current recognition judgment was shown to reflect both the confidence of an immediately preceding perceptual gender judgment (domain-general carryover at Lag 1) and also the confidence of the recognition judgment prior to that (domain-specific carryover at Lag 2). Moreover, the domain-specific effect was sensitive to response consistency: Confidence carryover was highest when old–new classifications repeated across trials. Whereas the domain-general effect may reflect metacognitive monitoring of internal factors such as alertness, the domain-specific effect was easily simulated by assuming that evidence within domains is "sticky," such that current memory or perceptual evidence is pulled toward prior evidence representations.

Keywords Recognition memory · Confidence · Sequential dependencies

In a typical recognition memory test, observers classify each of a series of items as previously studied (*old*) or unstudied (*new*) and often also rate the confidence of these classifications (e.g., *high*, *medium*, or *low*). Under signal detection models of recognition, each test probe is associated with some amount of memory evidence (or memory strength) that is evaluated in determining whether the probe is old or new. By virtue of their prior appearance within the context of the experiment, old items are associated with a higher average memory strength than are new items; accordingly, the *hit rate* (proportion of old items called old) substantially exceeds the *false-alarm rate* (proportion of new items called old) in most

recognition experiments. However, the evidence favoring old or new judgments is assumed to be continuous and noisy, and imperfectly diagnostic of the correct decision. Observers must therefore use a central decision criterion such that recognition probes that evoke memory evidence higher than the criterion yield old classifications, whereas those whose evidence falls below this criterion yield new classifications. Confidence in the old–new classification is a function of the distance between the criterial evidence value and the evidence evoked by the probe. Evidence values falling increasingly to the left of the criterion are associated with increasing confidence that the probe is new, while evidence values falling increasingly to the right of the criterion are associated with increasing confidence that the probe is old (e.g., Parks, 1966). Specific confidence ratings are assumed to be governed by individual criteria that observers hold for making each available confidence rating. The internal resolution or acuity of the observer (d') is assumed to be independent of the placement of classification and confidence criteria (Macmillan & Creelman, 2005).

✉ Justin Kantner
justin.kantner@csun.edu

✉ Ian G. Dobbins
idobbins@wustl.edu

[1] California State University, Northridge, USA

[2] Washington University in Saint Louis, Saint Louis, USA

Because the presentation order of recognition probes is randomized, signal detection and other decision models assume that evidence values are sequentially independent. This widespread assumption leads to the strong prediction that neither the classification nor confidence on trial n of a recognition test should predict the classification or confidence on trial $n + 1$. Put simply, judgments and confidence ratings on successive trials should be independent of one another because the memoranda (and associated evidence values) themselves have been sequentially randomized.

Despite its importance to recognition memory theory and measurement, the assumed serial independence of recognition classifications appears to be violated in simple item recognition tasks. Malmberg and Annis (2012) conducted a detailed investigation of sequential dependencies in recognition judgments, comparing these with established dependencies in perceptual judgment tasks such as absolute stimulus identification. In their Experiment 1, they found that a hit was approximately 10% more likely following a hit than following a miss. This sequential dependency of classifications was replicated in a host of other recognition studies and is potentially consistent with the idea that studied probes trigger recollection of other temporally associated probes, engendering a dependency. Critically, however, false alarms displayed analogous contingencies: A false alarm was considerably more likely following a hit than following a miss, and was also more likely following a false alarm than a correct rejection, demonstrating that old judgments spur subsequent old judgments regardless of the accuracy of the reports. This finding weighs against the idea that sequential dependencies are driven by veridical contextual retrieval.

Malmberg and Annis (2012) focused on the dependency of recognition classifications across trials, but did not address whether confidence judgments also display dependencies across trials. More recently, however, Rahnev, Koizumi, McCurdy, D'Esposito, and Lau (2015) demonstrated serial dependencies in subjective perceptual confidence and proposed a predictive decision model to explain them. In their Experiment 1, participants viewed briefly flashed arrays of colored and intermixed Xs and Os and judged, in an alternating fashion, the particular letter (X vs. O, each in two different colors) and the particular color they believed was more numerous for each display. Each judgment was followed by a confidence report. Critically, the two confidence judgments were reliably correlated ($r = .23$), even though the perceptual evidence (letter-based vs. color-based numerosity differences) was orthogonal. We refer to this effect as confidence carryover. To explain it, Rahnev et al. (2015) proposed that the level of confidence on trial $n - 1$ is used to predict the discriminability of evidence (and thus the appropriate level of confidence) on trial n. They posited that subjects use this predictive strategy because they have learned to expect continuity in the perceptual conditions of the visual environment.

For example, having a bright clear view at Time 1 often forecasts a bright clear view at Time 2. Thus, if numerosity differences based on one visual feature are easy to resolve, then they should also be easy if one shifts to another visual feature in close temporal proximity.

Mechanistically, the Rahnev et al. (2015) model assumes that following unusually high confidence in a judgment, the subject contracts his or her confidence criteria, whereas following unusually low confidence the subject expands them (see Stretch & Wixted, 1998, for a similar model). This expansion and contraction causes confidence dependencies across the trials. To help appreciate the model's predictions, an extreme version of this criterion movement is illustrated in Fig. 1. As is clear from the figure, the expansion of criteria following a low confidence response causes the proportion of evidence falling into the highest confidence bins to shrink, whereas the proportion of evidence falling in the low and medium confidence bins increases. In contrast, when the criteria contract following a high confidence response, the proportion of evidence falling into the high confidence bins increases and the proportion falling into the low and medium confidence bins decreases. The model thus predicts that the likelihood of a high confidence rating increases as prior-trial confidence increases and decreases as prior-trial confidence decreases.

The findings of Rahnev et al. (2015) raise the question of whether sequential dependencies in confidence are also present in recognition judgments, and, if so, whether they manifest for the same reasons. One reason they may not stems from the rationale motivating the Rahnev et al. model; namely, that the
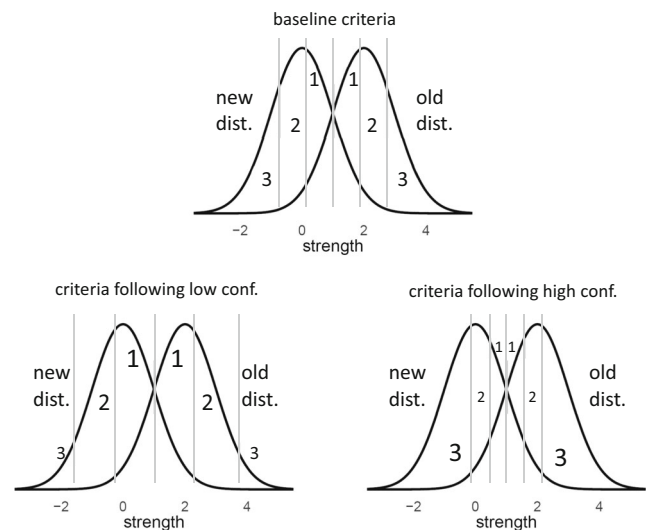

Fig. 1 Criterion behavior under the Rahnev et al. (2015) model. Relative to their baseline positions (top panel), the criteria contract following high confidence judgments (lower right panel) and expand following low confidence judgments (lower left panel). The numbers indicate confidence ratings, and the size of the font illustrates how the proportion of each confidence rating changes in response to the criterion movement

serial dependency results from having learned to predict continuity in perceptual conditions. For example, a current clear view anticipates a subsequent clear view a brief time later, and hence easy discrimination based on one visual attribute (e.g., shape) anticipates easy discrimination later using another visual attribute (e.g., color). However, this type of continuity is not expected for item recognition evidence because signals of oldness and newness are unlikely to predict one another. For example, when entering a foreign airport, one encounters many faces evoking strong signals of novelty. While these signals forecast continued novelty, they do not forecast that any encountered signals of familiarity should also be strong; indeed they may forecast that experienced familiarity in that environment will be weak. In this scenario, then, high confidence in "new" judgments would not predict high confidence in subsequent "old" judgments; thus, confidence would not carry over across the two response classes. The same within-class predictive bias holds for familiar environments such as one's workplace; a strong signal of familiarity anticipates continued familiarity, but it does not anticipate strong signals of novelty. A predictive model tailored for serial item recognition judgments should therefore predict positive within-class confidence dependencies (i.e., confidence in novelty predicting confidence in novelty or confidence in familiarity predicting confidence in familiarity), but perhaps negative or null between-class dependencies.

An alternative possibility is that the sequential dependencies observed in Rahnev et al. (2015) may not be tailored for particular domains of experience, such as perception versus recognition, and might instead manifest across a host of judgment domains in a similar manner. For example, if confidence carryover reflected self-assessments of current vigilance or fatigue levels or general task ease, then serial confidence correlations might occur across tasks from entirely different psychological domains because they represent a domain-general form of metacognitive monitoring or reasoning. A related question is whether confidence carryover is best understood as the manifestation of trial-by-trial criterion shifts, as in the Rahnev et al. (2015) model (see Fig. 1) or as an inappropriate carryover of evidence-linked information across trials. We address this question in the Model Simulations section below.

To begin to consider these possibilities, we first reanalyzed two existing recognition data sets, and then conducted a new experiment investigating the domain-generality of confidence carryover using interleaved verbal recognition and facial gender discrimination judgments at test. The rationale behind this experiment was to interleave judgment tasks for two largely independent psychological domains (recognition and perception). We can think of no environmental relationship outside of the laboratory where the ease of a gender discrimination decision predicts the subsequent ease of judging a printed word as previously seen or not. Thus, if confidence carries over across these tasks, then it cannot be because of learned environmental regularities outside of the laboratory. Instead, domain-general explanations for carryover, such as general metacognitive monitoring of fatigue or vigilance, might instead be required.

Extant data and confidence carryover

Extant Data 1: Han and Dobbins (2008), Experiment 1

These data were collected in the control condition of a study that manipulated recognition feedback. Because the aim of the study was to demonstrate that subtle imbalances in feedback accuracy could markedly alter recognition decision biases, it was important to establish whether the provision of feedback per se altered recognition biases. Thus, control participants ($N = 16$) received two blocks of recognition testing, with or without veridical feedback, to determine whether performance was altered with the provision of veridical feedback. There were four study/test cycles, with each test containing 60 studied and 60 new items. The stimuli were typical verbal materials (i.e., English nouns). Recognition status (old or new) and then confidence in the judgment (low, medium, or high) were entered separately, and following the confidence judgment either veridical feedback was delivered (two blocks), or no feedback was given (two blocks). The provision of feedback was pseudorandomized across the four blocks.

No-feedback blocks We begin by considering the blocks without feedback. Figure 2 shows the probability of each level of current report confidence (1 = low, 2 = medium, 3 = high) conditioned on the immediately preceding confidence. The prior level of confidence clearly alters the relative probability of each level of current confidence. For example, the probability of high confidence reports drops as the prior confidence level goes from high (right panel) to low (left panel). The reverse happens for low confidence reports, which are more likely following low confidence than high confidence.

To confirm that these dependencies were reliable and unrelated to the accuracy of judgments, we used a hierarchical linear model (HLM) with the subject's prior level of confidence and current accuracy as predictors of the current level of confidence (i.e., current confidence ~ current accuracy + prior confidence). Intercepts and slopes were modeled as independent random effects across the subjects. HLM was performed in R (R Core Team, 2016) using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015). The model confirms that prior confidence is a robust predictor of current recognition confidence above and beyond the accuracy of the current response (see Table 1).
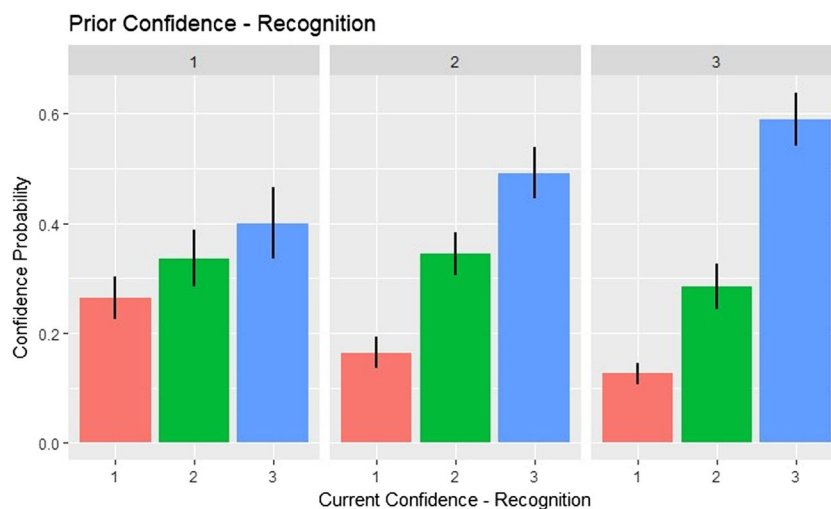
Fig. 2 The probability of a low (1), medium (2), and high (3) confidence judgment (bottom axis) following a low (column 1), medium (column 2), and high (column 3) confidence judgment on the previous trial. Error bars indicate ±1 *SEM*

Feedback blocks The same pattern of results emerged in feedback blocks, with conditional probabilities demonstrating dependence (see Fig. 3) and the HLM confirming reliability at the trial level (see Table 2).

Discussion of Extant Data 1 reanalysis The analysis of the feedback and the no feedback blocks yielded remarkably similar findings, with confidence carryover illustrated in both, and to a similar degree. The fact that confidence carryover occurs during blocks in which subjects receive fully veridical feedback carries implications for any account of the carryover phenomenon. Rahnev et al. (2015) explained carryover during perceptual discrimination judgments by

assuming that it was the result of statistical environmental learning outside of the lab, as environmental conditions are correlated across time, such that the quality of perception at time *n* anticipates the quality at time *n* + 1. Such a strategy is invalid for randomized trials in the laboratory, but in the absence of explicit feedback, Rahnev et al. suggested that participants may not realize this. They suggested that feedback might eliminate the phenomenon, at least for perceptual judgments. The current reanalysis of the Han and Dobbins (2008) data weighs against this idea, unless unlearning these ostensible regularities requires considerably more feedback-based learning. However, it should be noted that subtle manipulations of the veridicality of feedback during recognition cause robust changes in recognition decision biases (Han & Dobbins, 2008, 2009), demonstrating that feedback can easily alter recognition decision criteria in this type of testing situation. Together, these findings would require concluding that confidence-based predictions of future accuracy are insensitive to the presence of feedback, whereas classification biases are highly sensitive to the presence of feedback.

Table 1 Hierarchical linear model for the effect of prior recognition confidence on current recognition confidence, Extant Data 1 (no feedback blocks)

	Current Memory Confidence		
	B	CI	*p*
Fixed parts			
(Intercept)	1.64	1.44, 1.83	**.0000**
Current accuracy	0.46	0.38, 0.53	**.0000**
Prior confidence	0.16	0.09, 0.22	**.0001**
Random parts			
σ^2	0.395		
$\tau_{00,\ subj}$	0.124		
N_{subj}	16		
ICC_{subj}	0.230		
Observations	3,772		
R^2	.325		

CI = 95% confidence interval on parameter. R^2 are pseudo-r-square values. ICC = proportion of variance accounted for by subject factors. $\tau_{00,\ subj}$ reflects the variance of the subject intercept values

Extant Data 2: Kantner and Lindsay (2012), Experiment 3

The Han and Dobbins (2008) study used verbal materials and a two-part reporting procedure in which the recognition classification was followed by the confidence rating. Here, we consider recognition of pictorial information (scans of lesser-known paintings from masterwork artists) in which the recognition and confidence judgments are indicated simultaneously using a 6-point scale (*N* = 37). For comparison with the prior studies, recognition was dichotomized (e.g., 1, 2, and 3 = *new* and 4, 5, and 6 = *old*), and confidence was rescaled such that both 3, 2, and 1 and 4, 5, and 6 corresponded to low, medium,
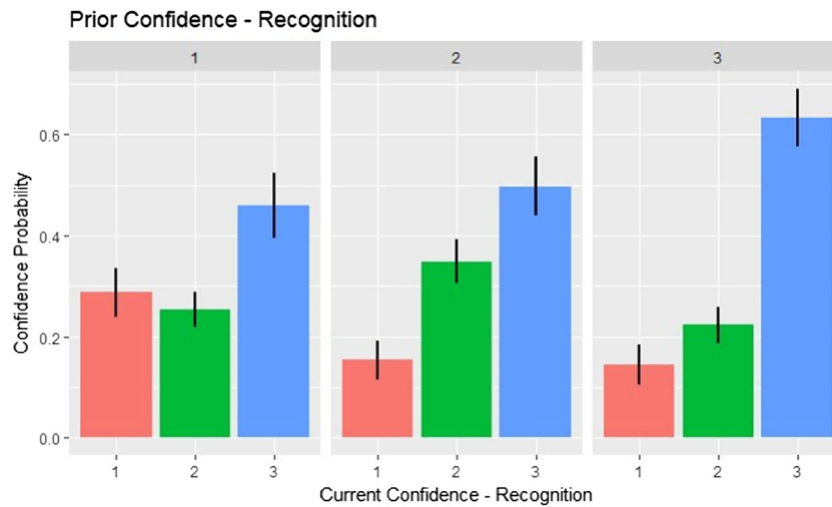
Fig. 3 The probability of a low (1), medium (2), and high (3) confidence judgment (bottom axis) following a low (column 1), medium (column 2), and high (column 3) confidence judgment on the previous trial. Error bars indicate ±1 *SEM*

and high confidence, respectively. There were two study/test cycles, and each test contained 48 studied and 48 novel pictures.

Although somewhat more modest, the confidence carry-over phenomenon is still clearly visible in the conditional probabilities (see Fig. 4) and was confirmed in the HLM (Table 3). Thus, confidence carryover also occurs during the use of simultaneous judgment scales and during the recognition of pictorial information.

Finally, we jointly examined both extant data sets to see whether the category of pairwise sequential recognition judgments (old–old, new–new, old–new, new–old) moderates the confidence carryover effect. For example, is confidence as

Table 2 Hierarchical linear model for the effect of prior recognition confidence on current recognition confidence, Extant Data 1 (feedback blocks)

	Current Memory Confidence		
	B	CI	*p*
Fixed parts			
(Intercept)	1.59	1.44, 1.75	**.0000**
Current accuracy	0.51	0.45, 0.57	**.0000**
Prior confidence	0.15	0.10, 0.20	**.0000**
Random parts			
σ^2	0.437		
$\tau_{00,\ subj}$	0.067		
N_{subj}	16		
ICC_{subj}	0.130		
Observations	3,775		
R^2	.228		

CI = 95% confidence interval on parameter. R^2 are pseudo-r-square values. ICC = proportion of variance accounted for by subject factors. $\tau_{00,\ subj}$ reflects the variance of the subject intercept values

likely to carry over from a new judgment into an old judgment as it is to carry over from a new judgment into another new judgment? As noted earlier, outside the laboratory it appears that novelty predicts novelty, whereas familiarity predicts familiarity. If carryover reflects predictive environmental learning, then one might expect a drop in carryover when judgments change. To test this possibility, we fit four linear models for each possible pairwise sequence of recognition conclusions. In each model current confidence was modeled as a function of current accuracy plus the immediately preceding confidence. Figure 5 shows boxplots reflecting the trends of the carryover coefficients (i.e., the regression coefficient for the lagged confidence predictor) across the subjects for the two data sets. The results suggest that carryover is lessened for classifications that switch (old–new or new–old) across pairs of trials versus classifications that remain the same (old–old or new–new).

The data also suggest that carryover is most prominent for sequential new judgments (new_new) and smallest when the judgments change across the pair of trials (old_new and new_old). We confirmed this pattern with a mixed-design ANOVA with data set as a between-subjects factor (Han & Dobbins, 2008, or Kantner & Lindsay, 2012) and response sequence as a within-subjects factor (new_new, new_old, old_new, or old_old). The analysis yielded main effects of data set, $F(1, 51) = 13.26$, partial-eta^2 = .21, $p < .001$, and response sequence, $F(3, 153) = 9.033$, partial-eta^2 = .15, $p < .001$, with no interaction between the two ($F < 1$). The main effect of data set simply reflects that carryover is more prominent for the verbal than for the pictorial recognition studies. Follow-up, Bonferroni-corrected comparisons for the main effect of response sequence demonstrated that confidence carryover was larger for the new_new sequence than the remaining three ($ps < .023$). There were no other reliable pairwise differences.
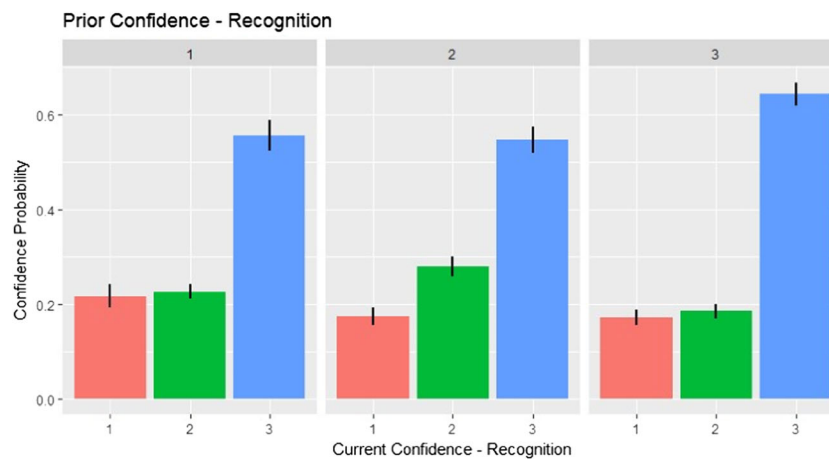
Fig. 4 The probability of a low (1), medium (2), and high (3) confidence judgment (bottom axis) following a low (column 1), medium (column 2), and high (column 3) confidence judgment on the previous trial. Error bars indicate ±1 *SEM*

Summary of extant data findings

The extant data establish that confidence carryover reliably occurs during recognition memory. It is most prominent for sequences of new judgments (new_new) and is reliably lower when responses change across recognition judgments (old_new or new_old) or when both responses are old (old_old). Whether the old_old sequence demonstrates reliable carryover depended on the data set. For the verbal recognition data of Han and Dobbins, the carryover effects were reliably positive for all sequences (see Fig. 5). In contrast, for the picture recognition data of Kantner and Lindsay (2012), only the new_new sequence reliably differed from zero. However, there are numerous differences across these two experiments aside from the materials, leaving the locus of

Table 3 Hierarchical linear model for the effect of prior recognition confidence on current recognition confidence, Extant Data 2

	Current Memory Confidence		
	B	CI	*p*
Fixed parts			
(Intercept)	1.81	1.72, 1.91	**.0000**
Current accuracy	0.53	0.46, 0.60	**.0000**
Prior confidence	0.08	0.05, 0.11	**.0000**
Random parts			
σ^2	0.481		
$\tau_{00, \text{subj}}$	0.051		
N_{subj}	37		
ICC_{subj}	0.090		
Observations	6,956		
R^2	.219		

CI = 95% confidence interval on parameter. R^2 are pseudo-r-square values. ICC = proportion of variance accounted for by subject factors. τ00, subj reflects the variance of the subject intercept values

the difference in outcomes (i.e., stimulus type vs. confidence judgment format or another factor) an open question.

To further test the specificity of confidence carryover effects, we designed an experiment in which high-level perceptual gender judgments were interleaved with verbal recognition judgments. The goal was to see whether confidence carries over across tasks from fundamentally different domains, or whether carryover is instead restricted within the domains of perception and recognition memory.

Experiment 1: Interleaved perception and verbal recognition

In Experiment 1, we interleaved a high-level perceptual judgment (face gender discrimination) and a verbal episodic recognition judgment, collecting report confidence for both. Given that there seems to be no environmental scenario in which one's confidence in recognition of isolated verbal materials should be predictively linked with confidence in discriminating the gender of faces, we viewed this as a particularly strong test of an environmental learning account of confidence carryover effects (at least in recognition memory). Ideally, if environmental learning of external regularities were the only cause of confidence carryover, we should see no carryover from adjacent perceptual discrimination and verbal recognition memory trials.

Method

Participants

Thirty undergraduates (11 male; mean age = 19.0 years; range: 18–21 years) enrolled in psychology courses at Washington University in Saint Louis participated in exchange for course credit. All participants provided informed consent in accordance with the university's Institutional Review Board.
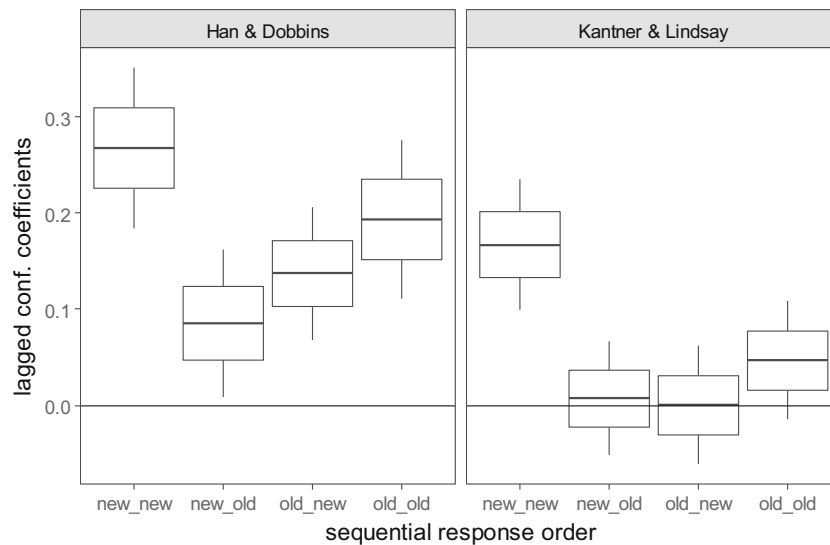
Fig. 5 Sequential response effects on the strength of confidence carryover. *X*-axis indicates the four possible sequences of responses for two consecutive recognition trials. *Y*-axis shows the size of the coefficient for the lagged confidence predictor for each fitted subject. Right and left panels isolate the two extant data sets. Box reflects one standard error of the mean, whereas box plus whisker reflects two standard errors of the mean

Materials

Stimuli consisted of common words for the recognition task and grayscale images of faces for the perceptual classification task. For each participant, 400 words were randomly drawn from a 1,216-item pool with an average of 7.09 letters, 2.34 syllables, and 8.85 per million printed word frequency (Kucera & Francis, 1967). Using the randomly drawn words, four lists of 100 items (50 old and 50 new) were created for each of four test blocks.

Likewise, 400 images were randomly selected from a database of 640 photographs of young Caucasian adults with neutral expressions (Endl et al., 1998). Each face was rounded-cropped to remove all peripheral features (ears, hair, etc.), sized to 190 pixels in width, and placed upon a 200 × 200 pixel uniform white background (see Fig. 6). The 400 randomly selected images were used to create four lists of 100 faces (50 male and 50 female) for each test block.

Procedure

Study and test materials were presented on Windows-based PCs. PsychoPy software was used to control presentation and timing (Version 1.83; www.psychopy.org; Peirce, 2007).

Data were collected in four blocks, each of which consisted of a study phase and a test phase. Of the 30 participants in our
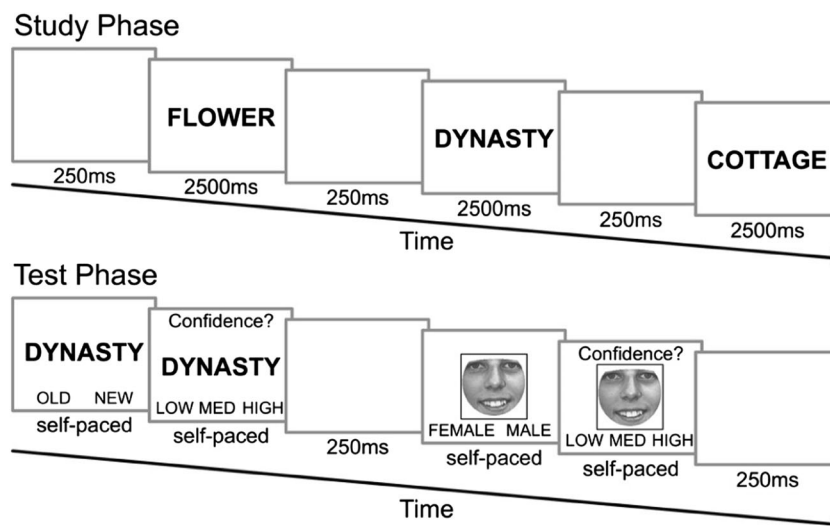

Fig. 6 Design schematic. Example of three study trials and one test trial. Each test trial consisted of one recognition memory judgment immediately followed by a confidence judgment and one gender classification judgment immediately followed by a confidence judgment

analyses, 29 completed all four blocks and one participant completed three blocks because they accidentally unplugged the computer. During each study phase, participants were presented with 50 words one at a time in the center of the screen. Each word appeared for 2,500 ms, followed by a blank screen for 250 ms. Participants were instructed to remember the words for an upcoming memory test. Each study block was immediately followed by the corresponding test block, which consisted of alternating trials of recognition memory and gender classification (100 trials of each; see Fig. 6).

During recognition, participants indicated study status (*old* or *new*) and then rated confidence on a 3-point scale (*low, medium, high*). Recognition confidence judgments were immediately followed by gender identification trials, wherein a centrally presented face was rated as male or female, followed by a confidence rating. In order to increase the range of judgment confidence for gender identification, half of the faces were inverted during testing and the other half were presented upright. This also allowed a manipulation check of the confidence carryover phenomenon: Because inverted faces are generally harder to identify than upright faces, the former should yield lower subsequent recognition memory confidence than the latter if confidence carries over across perceptual and recognition judgments. Testing was self-paced, with responses via keyboard. To control for motor priming, left-hand key presses were used for one task and right-hand key presses for the other. Key assignment was counterbalanced such that half of the participants used the *s, d,* and *f* keys during the recognition memory task and the *j, k,* and *l* keys during the gender classification task. The classification judgment (gender or recognition) was made using the inner two keys (*s, d; j, k*) and the subsequent confidence judgment using all three keys. Thus, the keys used for one task were different than the keys used for the other task for all participants.

Results

In order to match the format of the prior analyses, we focus on predicting current verbal recognition memory confidence using the immediately preceding perceptual judgment confidence, or confidence in the previous recognition judgment one step further back. The top panel of Fig. 7 depicts the conditional confidence probabilities. As is clear from the figure, perceptual confidence carries over into recognition confidence, demonstrating confidence carryover across judgment domains that are wholly unrelated to one another. The bottom panel of Fig. 7 demonstrates that recognition confidence two steps back (i.e., Rec_{prior} … $Percep_{prior}$ … $Rec_{current}$) also appears to carry over into the current recognition trial. One possibility, however, is that this carryover between recognition trials is largely or fully mediated by the intervening perceptual trial.

To address this possibility, we constructed three HLM models, shown in Table 4. In all three models, accuracy is statistically controlled, and the intercepts and main effects are modeled as random and independent across the subjects. Model 1 demonstrates that recognition confidence two steps back (i.e., skipping the intermediate perceptual trial) predicts the current recognition confidence. Model 2 demonstrates that the immediately preceding perceptual trial also predicts the current recognition confidence. Critically, Model 3 demonstrates that when both prior trial types are entered, they uniquely predict current recognition confidence. Indeed, the coefficient for prior recognition is not appreciably lowered when prior perception is also entered into the model. Thus, the intervening perceptual confidence is not merely serving as a mediator. Instead, there appear to be unique contributions of both domain-specific (memory to memory) and domain-general (perception to memory) confidence carryover effects.

Finally, we again considered whether the carryover effect for recognition depended upon the pairwise sequence of recognition decisions by fitting separate regressions for each subject, for each possible pairwise sequence of responses (i.e., new_new, new_old, old_new, and old_old). The model contained predictors of current recognition accuracy and preceding recognition confidence, which was the predictor of interest. Figure 8 shows the coefficients for the prior recognition confidence predictor and displays a pattern remarkably similar to that in the extant data (see Fig. 5), with the strongest carryover occurring for serial "new" conclusions, $F(3, 87) = 4.51$, parital-eta^2 = .13, p = .005. Because the analysis constitutes a planned follow-up of the previous findings demonstrated in Fig. 5 (as opposed to an exploratory analysis) and was contingent upon the significant omnibus, Bonferroni correction was not applied, constituting Fisher's Least Significant Difference (LSD) approach. Post hoc tests demonstrated that the new-new sequences yielded stronger dependency than new-old and old-new (ps < .03). Additionally, unlike the extant data, the old_old sequence also yielded carryover effects larger than the new_old sequence (p = .038). Jointly, the data support the conclusion that carryover effects are strongest for serial "new" recognition judgments and are consistently reduced when judgments change across the trials (new–old and old–new). The magnitude of the old_old carryover effect appears intermediate between new_new effects and the effects observed when responses change.

Because the memory confidence carryover effect demonstrated sensitivity to whether sequential recognition judgments changed, we also examined whether perceptual confidence carryover effects were likewise response dependent. To do so, we analyzed the face recognition trials in an analogous manner, separately measuring carryover effects for the four possible pairwise gender judgments (female_female, female_male, male_female, male_male). Within these
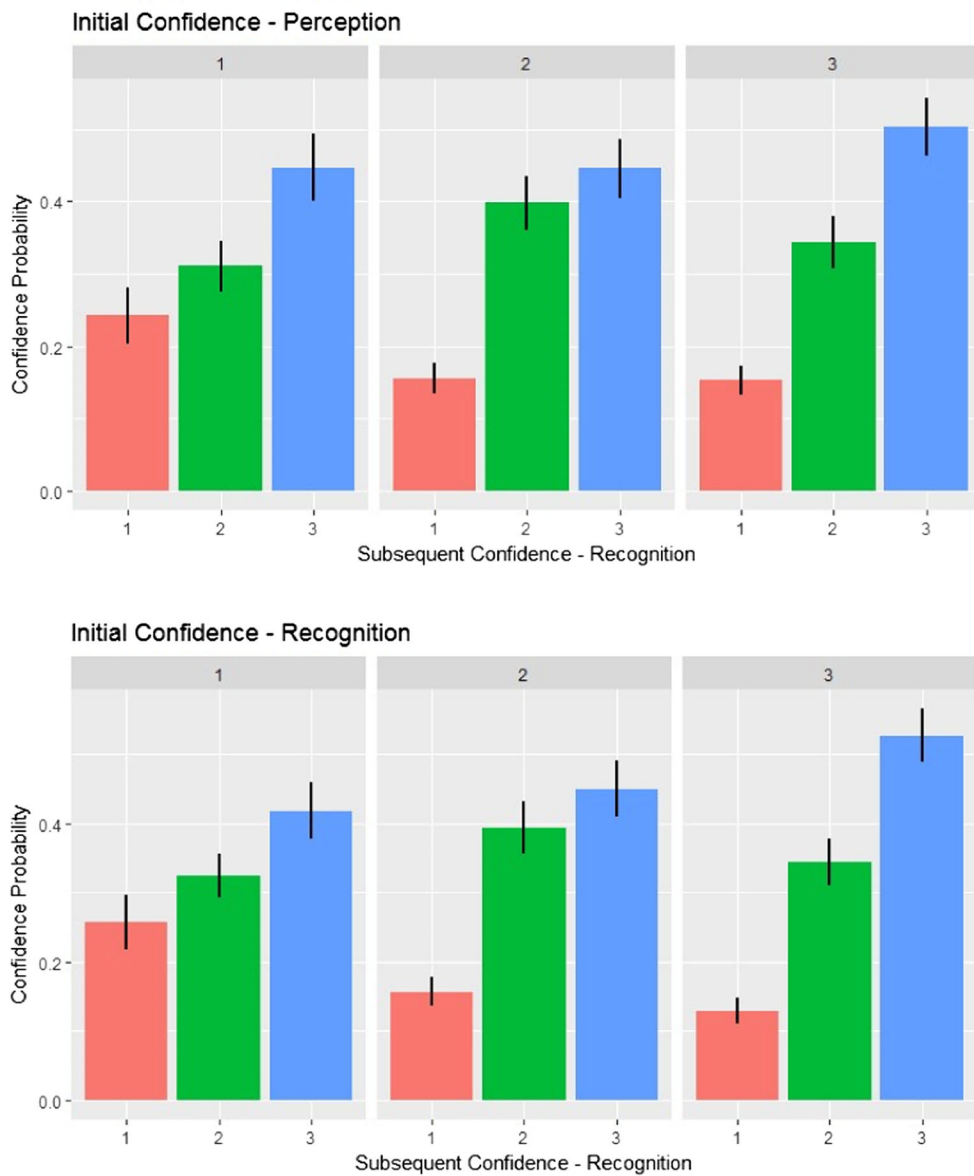
Fig. 7 Probability of a low (1), medium (2), and high (3) confidence judgment (bottom axis) following a low (column 1), medium (column 2), and high (column 3) confidence judgment on the previous trial. Error bars indicate ±1 *SEM*. Upper panel demonstrates the effect of a prior perceptual judgment on a current recognition judgment. Lower panel demonstrates the effect of a prior recognition judgment (two steps back) on a current recognition judgment

pairings, a model was fit for each subject in which current confidence was predicted by current accuracy plus the confidence of the prior gender classification (skipping the intermediate recognition trial). Figure 9 demonstrates that, as with recognition, confidence carryover during perception is strongest for judgments that repeat versus those that change, $F(1, 87) = 3.18$, partial-eta^2 = .10, $p = .028$. Pairwise LSD tests revealed that male–male carryover was reliably larger than both of the response change conditions (male_female, $p = .045$; female_male, $p = .028$), which did not differ from one another. The female_female condition tended toward greater carryover than the female_male condition ($p = .068$). No remaining comparisons were reliable (ps > .19).

Model simulations

The analysis of extant data and the new empirical findings demonstrate two new confidence carryover phenomena. The fact that carryover occurs from high level visual perception to verbal recognition (see Table 4, Models 2 and 3) demonstrates a domain general carryover phenomenon that cannot result from predictive learning about external environmental contingencies. We consider whether the Rahnev et al. (2015) perceptual carryover model can be expanded to account for this domain general carryover phenomenon in the General Discussion. However, there also appears to be a separate carryover phenomenon within the domain of recognition

Table 4 Hierarchical linear models for the effect of prior recognition confidence (Model 1), prior gender judgment confidence (Model 2), and both prior judgments combined (Model 3) on current recognition confidence, Experiment 1

	Model 1			Model 2			Model 3		
	B	CI	p	B	CI	p	B	CI	p
Fixed parts									
(Intercept)	1.68	1.51, 1.85	**.0000**	1.73	1.57, 1.88	**.0000**	1.57	1.40, 1.74	**.0000**
Accuracy	0.48	0.37, 0.59	**.0000**	0.50	0.39, 0.61	**.0000**	0.48	0.37, 0.59	**.0000**
Prior memory confidence	0.11	0.05, 0.17	**.0007**				0.10	0.05, 0.15	**.0006**
Prior gender confidence				0.08	0.03, 0.13	**.0031**	0.06	0.03, 0.08	**.0012**
Random parts									
σ^2	0.389			0.397			0.388		
$\tau_{00,\,sub}$	0.211			0.171			0.197		
N_{sub}	30			30			30		
ICC_{sub}	0.299			0.255			0.286		
Observations	11,782			11,782			11,782		
R^2	.320			.307			.323		

CI = 95% confidence interval on parameter. R^2 are pseudo-r-square values. ICC = proportion of variance accounted for by subject factors. $\tau00$, subj reflects the variance of the subject intercept values

memory that spans an intermediate perceptual judgment (see Table 4, Models 1 and 3), and an analogous phenomenon within the domain of gender judgment that spans an intermediate recognition judgment. These effects are domain specific in the sense that their magnitudes depend upon whether the current classification matches the previous one in that judgment domain, and because they are not mediated by the intervening task (see Table 4). For example, the size of the memory carryover effect depends upon the response sequence, such that new_new sequences yield the strongest confidence carryover (see Figs. 5 and 8), and carryover is larger for gender judgments when the classification repeats than when it changes (see Fig. 9). The Rahnev et al. (2015) model does not apply to these domain-specific carryover effects because it is purposely

crafted at a higher level of judgment abstraction; the specific responses are not represented in the model. For example, if the observer concludes that Xs are more numerous than Os on the current trial, this judgment cannot favor carryover to one versus the other stimulus color on the next trial. Instead, it is the confidence of numerosity discrimination on the current trial (regardless of which attribute is attended) that carries over to the confidence of the numerosity discrimination on the next.

The results of Experiment 1 raise the question of how one might craft a decision model that anticipates the sequential dependencies shown in serial recognition and serial perceptual judgments. Most formal theories of sequential dependencies have been applied to perceptual judgments, particularly absolute identification (e.g., Brown, Marley, Donkin, & Heathcote,
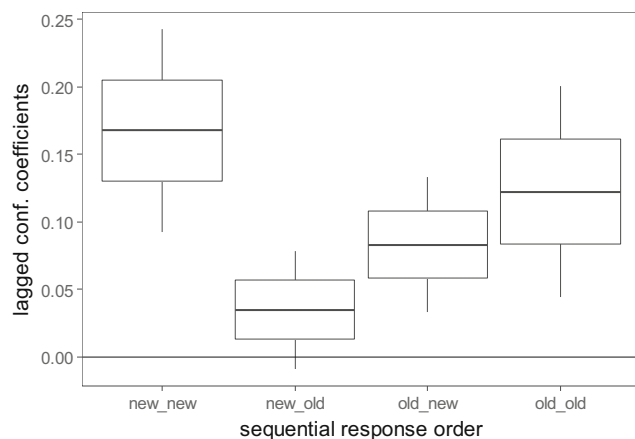

Fig. 8 Coefficients from individual fits using prior memory confidence (two trials back) to predict current memory confidence for each of four possible pairwise sequences of recognition responses. Box indicates ±1 SEM. Box plus whiskers indicates ±2 *SEM*
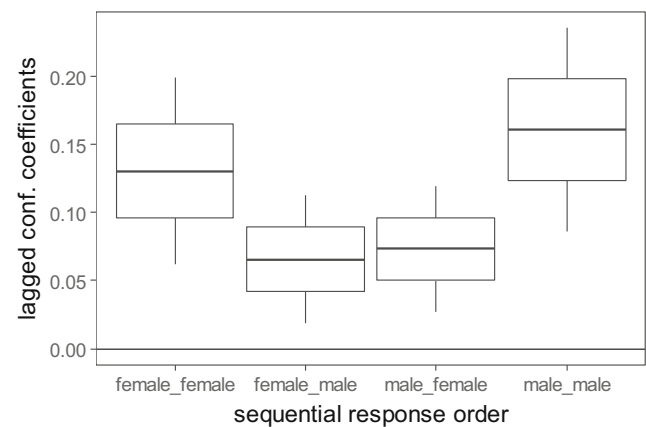

Fig. 9 Coefficients from individual fits using prior gender classification confidence (two trials back) to predict current gender classification confidence for each of four possible pairwise sequences of gender classifications. Box indicates ±1 *SEM*. Box plus whiskers indicates ±2 *SEM*

2008; Stewart, Brown, & Chater, 2005). Absolute identification and recognition are very different tasks, however, and given differences in patterns of sequential effects between the former and the latter (particularly at increasing lags), Malmberg and Annis (2012) concluded that mechanisms designed to account for dependencies in a particular perception task are not likely to apply to memory judgment dependencies. Treisman and Williams (1984) posited that sequential effects are the result of criterion shifts, a mechanism applied to explain dependencies in recognition memory data as well (e.g., Benjamin, Diaz, & Wee, 2009; Mueller & Weidemann, 2008; Ratcliff & Starns, 2009). Some have characterized these criterion shifts (and the resulting sequential dependencies) as strategic responses to the short- and long-term goals of the observer and/or as adaptations to learned environmental regularities (Rahnev et al., 2015; Treisman & Williams, 1984), while others characterize them as noise that arises from the inability of the cognitive system to maintain perfectly stable criteria (Benjamin et al., 2009). In either case, it is not clear that criterion shifts are necessary to explain sequential dependencies, nor that a criterion-based explanation of dependencies could account for the range of findings reported here. An alternative explanation for sequential effects in memory judgment tasks is that judgments on trial n correlate with those on trial $n-1$ not because of trial-by-trial criterion movement, but because some portion of the memory evidence elicited on trial $n-1$ transfers or carries over to trial n. Annis and Malmberg (2013) incorporated this assumption into a model of judgments of frequency (a memory task analogous to absolute identification), and proposed that such carryover occurs due to participant fatigue or inattention on a subset of trials. Research highlighting the effects of output interference in recognition memory (Criss, Malmberg, & Shiffrin, 2011) is also broadly consistent with the notion that decision evidence from other test trials can infiltrate the current trial. We sought to test an account of domain-specific sequential dependencies that did not rely on criterion shifts, that was psychologically plausible and applicable to confidence ratings, and that captured the systematic variability in carryover according to response sequence for both recognition and perception judgments.

Instead of assuming that observers continually adjust decision criteria on every trial, we consider a framework in which the evidence itself is "sticky": that is, evidence from trial n alters the perception or registration of evidence on trial $n+1$. There are several psychological motivations for this assumption. First, the sequential dependency in recognition classifications (as opposed to confidence) has been interpreted in terms of the carryover of actual recognition cue information from trial to trial (Annis & Malmberg, 2013), which would induce serial correlations in the actual evidence that is recovered. Second, neuromodulatory models of hippocampal processing are consistent with the idea that the processing occurring on trial n alters the memory processing on trial $n+1$,

biasing the observers towards the encoding of new information or the retrieval of episodic information (e.g., Duncan, Sadanand, & Davachi 2012). Thus, these frameworks also offer the potential for serial correlations in actual evidence values across trials as opposed to standards of evidence (i.e., decision criteria). Finally, in the case of perception, Fischer and Whitney (2014) demonstrated that sequential perceptions are altered in an assimilative manner, at least for sinusoidal grating judgments. That is, the perception on the current trial is pulled toward that of the prior trial, and this effect is strongest when the orientation of the gratings is more similar across the pair of trials (Fischer & Whitney, 2014). They proposed that this serial dependence in perception preserves the continuity of visual experience, given the general constancy of objects in the physical world. Directly applying these ideas to a recognition evidence decision model would mean that the perceived familiarity (or novelty) of a stimulus on the current trial is altered by the strength of evidence of the stimulus on the preceding trial, a process that would be beneficial if there is a general continuity of familiarity or novelty in the environment. In the context of recognition evidence, the evidence on trial $n+1$ would be pulled toward the evidence recovered on trial n, but only to the extent that they are close on the evidence axis. Critically, as we demonstrate through simulation, this would naturally lead to a reduction of carryover when responses change across pairs of trials because response changes are associated with larger average differences in evidence across the trials, and hence less possibility of assimilative effects.

To implement a "sticky evidence" model, we assume that the perceived evidence on the current trial is a weighted function of the current evidence and the immediately preceding evidence. Critically, this weighting is sensitive to the distance between the two evidence values such that when they are close, the prior evidence plays a larger role than when they are distant. This assumption was implemented in the model using a simple exponential weighting function of $1/(1 + \exp(\text{distance})^{\text{eta}})$, where eta is a scaling parameter which governs how quickly the prior trial's influence drops off with evidentiary distance. Figure 10 depicts that as the distance between current and former evidence decreases, the weight on the prior evidence increases towards one and the weight on the current evidence decreases towards zero. Thus, the recognition evidence is increasingly perceived as equivalent to that of the last trial. In contrast, as the distance increases, the current trial's evidence becomes dominant and the prior trial's evidence is increasingly disregarded. The scale parameter of the function controls how rapidly this transition occurs. We chose exponential weighting out of convenience; however, any function with this general property captures the basic idea of sticky evidence. This type of weighting will naturally lead to response specificity in the carryover phenomenon, because changes of response are typically accompanied by bigger differences in evidence across the trials. Hence, the effects of the
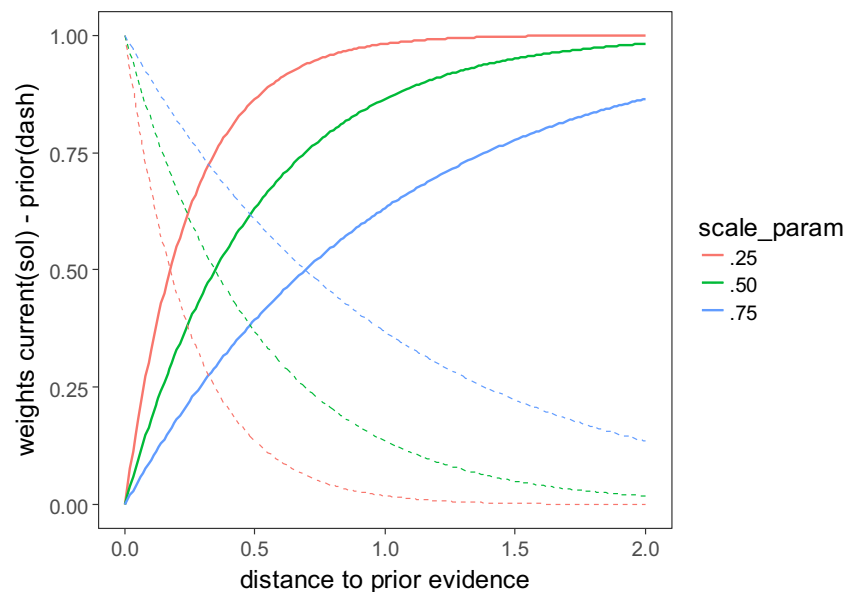
Fig. 10 Exponential weighting functions used in the sticky evidence model simulation. Solid lines represent weight given to current evidence as a function of its distance from prior evidence. Dashed lines represent weight given to prior evidence as a function of its distance from current evidence. Functions are shown for three different scaling parameters that affect the distance at which current versus prior evidence dominates the current decision. (Color figure online)

prior trial are mitigated more for response changes than for response repetitions.

For the simulation we created 100 fictive subjects with 600 trials per subject (300 targets and 300 lures). Each subject's discrimination ability (d') was randomly sampled from a uniform distribution from .5 to 1.5. Using this value, targets and lures were randomly sampled from the appropriate normal distributions and then randomly ordered. The subject's five decision criteria (old–new plus three levels of confidence) were also randomly determined, subject only to the constraint that each of the six created bins contained at least 5% of the total evidence. These criteria remain fixed throughout the trials for each fictive subject. Finally, on each trial following the first, the evidence on the current trial was adjusted to reflect the weighted average of the current and prior trial's evidence as a function of distance. Thus, the adjusted evidence values represent a series of sticky values, each of which tends to pull the next trial's evidence toward it as a function of their relative distance along the evidence continuum. The fictive subject's classifications and confidence were recorded for each trial using the sampled (static) criteria. We calculated these values for both the adjusted (sticky) and unadjusted evidence for comparison.

Figure 11 shows the results for three scaling parameters (.25, .50, and .75). The left panels plot the behavior of the lagged confidence coefficients analogous to those of the behavioral data; that is, they show the coefficients for the lagged confidence predictor when predicting current confidence across the fictive subjects (with accuracy statistically controlled). The right panels show the aggregate receiver operating characteristic (ROC) curves collapsed across subjects for both the unadjusted and adjusted confidence data. The left panels demonstrate a pattern analogous to the empirical data (see Figs. 5, 8, and 9). Confidence carryover is reliable, but it is diminished when responses change. The right panels collapse across the fictive subjects and show the overall ROCs for the sticky versus unadjusted evidence values. They illustrate that there is a very slight cost in terms of area under the curve (AUC; a measure of recognition sensitivity) from confidence carryover. This cost occurs because confidence carryover is, in the context of a randomized list, a type of random noise added to the evidence variable. That is, because the sampled prior evidence is in fact unrelated to the current evidence (before adjustment), the adjustment process itself constitutes noise. However, as the figure shows, the costs under this particular decision model are very slight in terms of AUCs for these scaling parameters. One reason that the costs are so small is that the effect is increased as the pair of evidence values become more similar. Hence, for a large portion of the trials in which the two sequential evidence values are dissimilar, there is little to no effect to be observed. Finally, from Fig. 11 it is clear that the sticky evidence model is capable of producing similar levels of confidence carryover as the empirical data (e.g., Figs. 8 and 9).

The sticky evidence model also produces carryover in recognition classifications themselves, which is important since recognition classifications have been shown to be positively serially correlated (Malmberg & Annis, 2012). In the current simulations, the classifications were serially correlated (old = 1, new = 0) on average .022, .075, and .135 across the fictive subjects for the three scale parameter values. Thus, for the highest scale parameter value, responding "old" on the prior
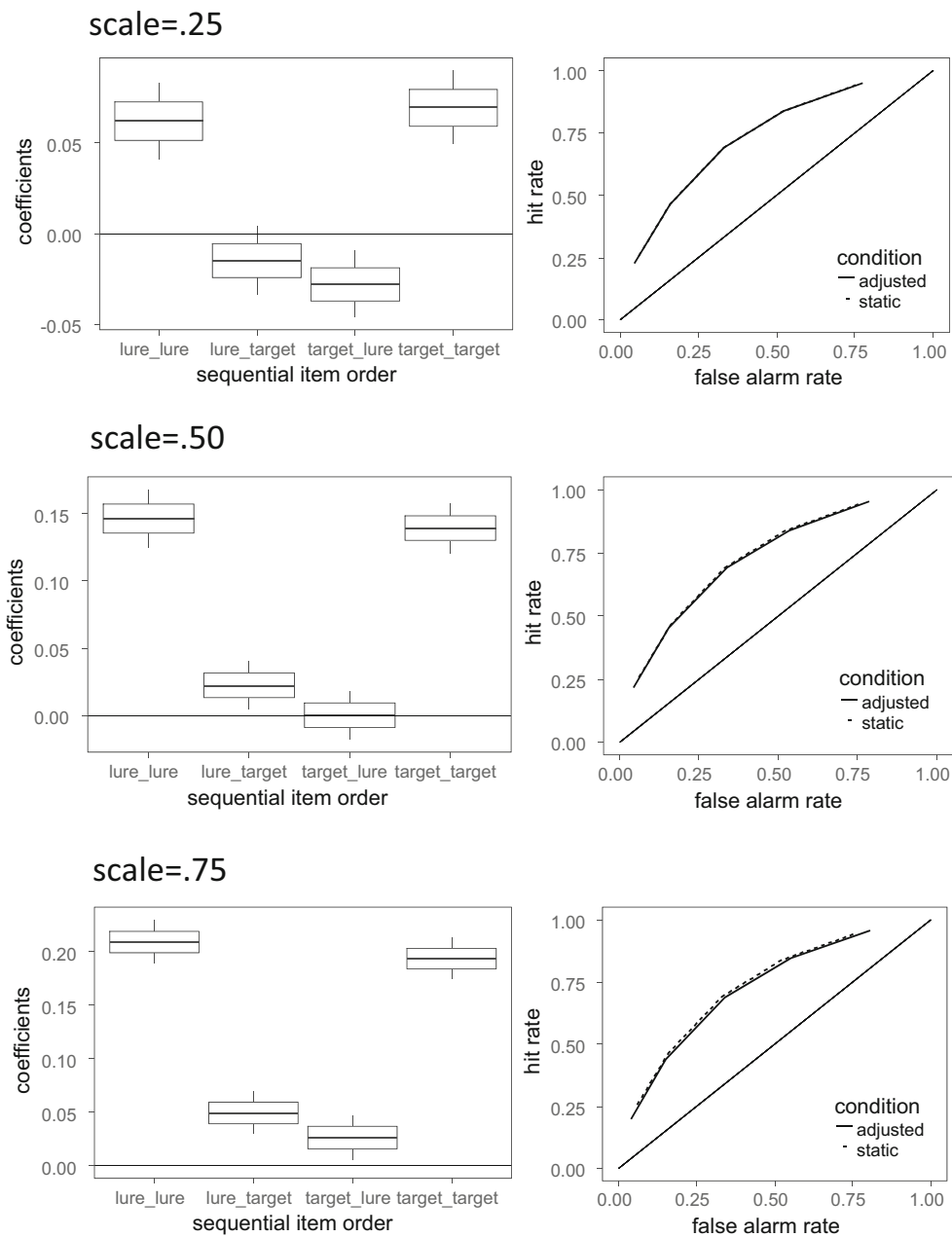
Fig. 11 Simulation results of sticky evidence model using three scaling parameters. Left panels demonstrate the behavior of linear model coefficients when predicting current confidence using accuracy and prior confidence for 100 fictive subjects and 600 test trials (300 targets and 300 lures) per subject. Right panels demonstrate the ROCs aggregated across the subjects using classifications based on either the unadjusted strength values or the adjusted, "sticky" values

trial increases the likelihood of responding "old" on the current trial by approximately 13.5%. This dependency occurs for the same reason as confidence carryover: when a current trial's evidence value is on one side of the central classification criterion, it has the potential to pull the subsequent trial's evidence value across this boundary if the two are close enough given the scaling parameter in place.

Although the sticky evidence model well captures the response-dependent confidence carryover effect, Figs. 5 and 8 suggest differences across recognition and perception in the empirical data. More specifically, in recognition, old_old sequences appear to yield confidence carryover that is smaller than new_new sequences yet larger than response-change sequences. Under one-dimensional signal detection accounts of recognition, this pattern might reflect the common assumption that old evidence is more variable than new evidence (e.g., Parks, 1966). To simulate this possibility, we reran the above simulation, but increased the old item evidence distribution's standard deviation to two, using a scaling parameter of .75 for evidence weighting. As Fig. 12 shows, this produces the
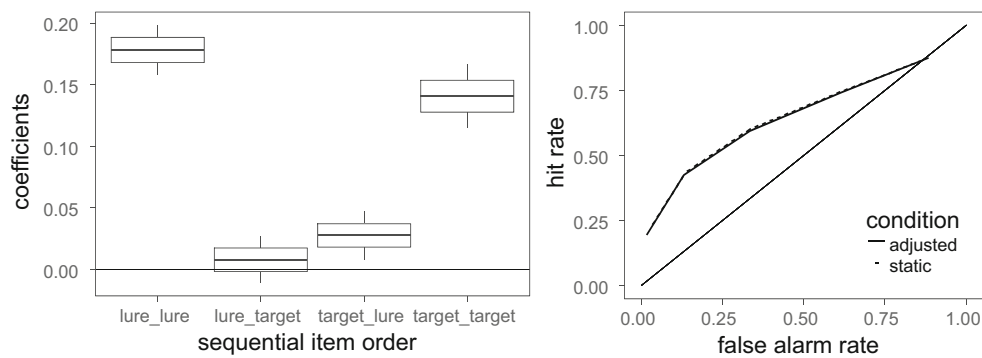
Fig. 12 Simulation of sticky evidence model using the unequal variance assumption. Left panel demonstrates the behavior of linear model coefficients when predicting current confidence using accuracy and prior confidence for 100 fictive subjects, and 600 test trials (300 targets and 300 lures) per subject. Right panel depicts the ROCs aggregated across the subjects using classifications based on either the unadjusted strength values, or the adjusted, "'sticky" values. For the simulation, the old item evidence distribution was set to a standard deviation of two

expected asymmetry in the ROC and it reduces the serial correlation for old_old sequences relative to new_new sequences. This occurs because as old item evidence becomes more variable, it increases the odds that a sequence of two old evidence values will be sufficiently far apart to minimally affect one another. The old_old sequence carryover effect can be further reduced by increasing the old item standard deviation beyond two; however, values that large are unusual in recognition data.

A second way to potentially lessen confidence carryover for old_old sequences is to assume a dual process model with a some-or-none recollection component (Yonelinas, 1994). Under this approach, some proportion of old items are assumed to trigger contextual recollective experiences, which in turn lead to highly confident endorsement of the items as studied. To simulate this assumption, we reran the simulation with a scaling parameter of .75 and assumed that a random 25% of the old item trials triggered contextual recollection. This was implemented by setting the strength of evidence for each recollection trial to the maximum observed value for the fictive subject, ensuring recollection trials would garner the highest confidence old classifications. As Fig. 13 shows, this reduces confidence carryover for the old_old sequences relative to the new_new and, as expected, produces the familiar asymmetric ROC of recognition memory. Conceptually, the recollection process limits the manifestation of the sticky evidence process because recollection overrides or overshadows familiarity as the basis for old responding, hence masking the sticky evidence process. Thus, both unequal variance and thresholded recollection assumptions are capable of producing the fundamental carryover phenomena reported here within a sticky evidence decision model.

Discussion

Experiment 1 replicates and extends the findings from the reanalysis of the extant data sets, demonstrating a confidence carryover effect that spans wholly different psychological domains. This effect cannot reflect simple motor priming, because subjects used different hands to make the recognition and gender judgments. In addition to this domain-general carryover phenomenon, we observed domain-specific carryover effects for the recognition and perceptual judgments. These effects are sensitive to repetitions versus changes of judgment, and they survive an intermediate, unrelated judgment; for example, recognition confidence carryover occurred despite the presence of an intervening perceptual judgment. Moreover, these domain-specific effects are not mediated by the intervening classification task. For example, Table 4 demonstrates that the influence of prior recognition confidence on current recognition confidence is essentially unchanged by the inclusion of the intervening gender classification confidence. The influence of prior recognition confidence on current recognition confidence is therefore not the result of its influence on the intermediate perceptual task. This same pattern holds for the domain-specific perceptual effect (not shown). Thus, the domain-specific and domain-general carryover phenomena appear to be unique influences on judgments in these tasks.

Because the domain specific carryover effect is highly sensitive to response change versus repetition, we developed a sticky evidence decision model potentially capable of producing key domain-specific carryover phenomena. This new model was motivated by several findings suggesting that the actual evidence in serial classification tasks may be sequentially dependent in an assimilative fashion. The model demonstrates the appropriate sequential response sensitivity and correctly produces serial correlation in both confidence and classification. Furthermore, two widely endorsed extensions of the basic signal detection decision model (unequal variance or thresholded recollection) yielded the correct declines in old–old sequential dependencies when incorporated into the sticky evidence model. Although the relative merits of the unequal variance versus thresholded recollection approaches remain highly debated, the point for the current study is that
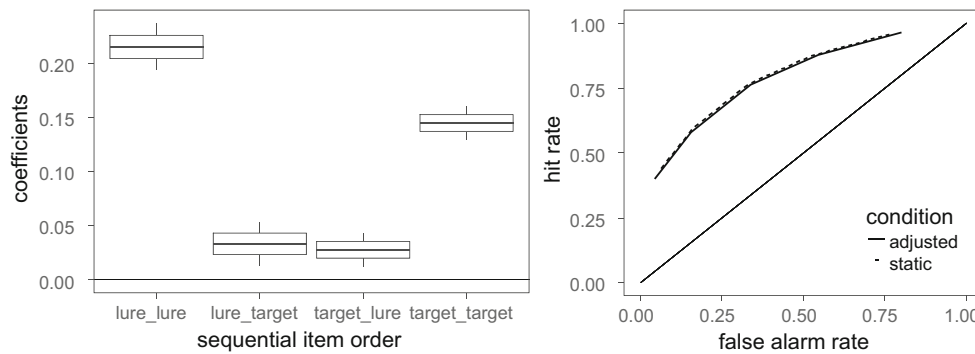
Fig. 13 Simulation of sticky evidence model using a thresholded recollection assumption. Left panel demonstrates the behavior of linear model coefficients when predicting current confidence using accuracy and prior confidence for 100 fictive subjects and 600 test trials (300 targets and 300 lures) per subject. Right panel depicts the ROCs aggregated across subjects, using classifications based on either the unadjusted strength values or the adjusted, "sticky" values. For the simulation, there was a .25 probability that each old item would trigger recollection, leading to maximum strength and hence high "old" confidence for the fictive subject

they are both in principle capable of yielding the correct pattern of confidence carryover during recognition when evidence is modeled as sequentially sticky. [1]

General discussion

The current findings demonstrate that confidence in a recognition judgment is a partial function of confidence in the preceding judgment. Confidence carryover occurred between consecutive recognition judgments in analyses of two existing data sets using different stimuli and response procedures. As highlighted by Malmberg and Annis (2012), sequential dependencies are inconsistent with current theories, such as signal detection theory, that assume classification and confidence are a sole function of the memory signal elicited by each test probe. Since test stimuli are typically randomized, these decision models assume that decision evidence, and hence the resulting judgments, will be sequentially independent.

Importantly, a new experiment with interleaved perceptual and recognition judgments revealed both domain-general and domain-specific confidence carryover effects. The present work is the first to document confidence carryover in recognition memory, and the first to demonstrate that confidence in recognition memory can be influenced by the confidence of an immediately preceding perceptual judgment. The domain-general carryover effect was demonstrated when the confidence of an immediately preceding perceptual judgment influenced that of a subsequent verbal recognition memory judgment (and vice versa). The fact that this phenomenon crosses fundamentally different

domains suggests it operates at a high level of abstraction and raises the question of whether the Rahnev et al. (2015) predictive model can accommodate these findings. Under that model, confidence carryover is assumed to result from learning that favorable viewing conditions facilitate numerosity discrimination regardless of whether shape or color is the category feature of interest. However, it is doubtful that that an environmental learning account alone can accommodate the domain-general carryover demonstrated in the current report, because verbal recognition memory and gender discrimination of faces would not be similarly limited by viewing conditions (except perhaps under extreme conditions, e.g., if the words could not be read). However, Rahnev et al. also suggested that conditions internal to the observer, such as headache or distractibility, might also influence decisions across different tasks in a predictive manner. This would require that the observer had learned that these sorts of internal states can generally impair performance and would traditionally fall under the rubric of metacognitive monitoring or awareness. This explanation would jointly accommodate domain-general carryover across perceptual and memory tasks, as was demonstrated here, and the confidence carryover observed in Rahnev et al. (2015) within perception.

The current study also demonstrated domain-specific confidence carryover. Domain-specific carryover appears to be a different phenomenon than domain-general carryover, for several reasons. First, it is sensitive to the repetition versus alteration of consecutive within-domain decisions. For example, in recognition memory, new_new sequences yielded the highest carryover effects, whereas changed judgments yielded the lowest carryover. Of course, the domain-general effect is, by construction, response independent. Second, even though domain-specific carryover spans an intermediate judgment from the other domain, the effect is not mediated by the confidence of this intermediate judgment (e.g., Table 4). For example, the confidence carryover from one memory judgment to the next did not require transmission via the intermediate perceptual judgment. The same direct within-domain influence was revealed for perception. Finally, as

---

[1] While post hoc model fitting is often conducted for decision models, its utility here would be limited, given there are no well-established competing models in this domain. Moreover, the exponential weighting function itself was used for convenience in order to capture a general principle, and there are likely a multitude of ways to instantiate the concept of sticky evidence at the decision level (see Roberts & Pashler, 2000). Thus, the simulations constitute a proof of a general principle that distance-dependent, assimilative effects for judgment evidence induce patterns similar those in the empirical data.

Table 4 shows, the domain-specific influence is stronger than the domain-general influence, even though it occurs one trial further back in time. An analogous pattern was observed when predicting current gender confidence. That is, the prior gender judgment exerted a stronger effect than the prior recognition judgment even though it occurred one trial further back in time (not shown). This pattern is remarkable because carryover effects generally diminish across time within a given domain (e.g., Malmberg & Annis, 2012). In a reanalysis of the extant data sets described above, we predicted current recognition confidence with prior recognition confidence at both recognition Lag 1 (i.e., the previous recognition judgment) and recognition Lag 2 (i.e., two recognition judgments back), and in all instances the coefficient for the two-back condition was smaller than for the one-back condition (not shown). These findings converge on the idea that domain-specific and domain-general carryover effects may be functionally dissociable.

Because the Rahnev et al. model is by construction response independent, while the domain-specific carryover effect was heavily response dependent, we developed a simple decision model that relied upon the construct of sticky evidence to see if it could generate the main characteristics of domain-specific carryover. This model is based on the idea that the recovered evidence is altered from trial to trial in an assimilative fashion within domains, an idea supported by several findings and frameworks. The sticky evidence model yielded the required pairwise response dependencies in confidence carryover (Figs. 11, 12, and 13), such that repeated responses yield more carryover than changed responses. Moreover, the model produces sequential dependencies in the classifications themselves, consistent with empirical data. Finally, the model was easily modified to accommodate the finding that old_old recognition sequences may show lower confidence carryover than new_new sequences by incorporating popular unequal variance or thresholded recollection assumptions. We note, however, that the construct of sticky evidence and the Rahnev predictive accounts are not exclusive. Under a joint model, one would assume that current evidence is altered by its similarity to prior evidence, and also that current decision criteria may be altered by the perceived ease of prior judgments. Thus, a full model would contain both a "sticky evidence" evidence process and some form of metacognitive learning linking internal states such as fatigue or distractibility to performance.

Future work

Given that confidence carryover appears to be a robust phenomenon across a wide range of recognition memory procedures, future work should examine conditions that increase or decrease the magnitude of the effect and further isolate domain-general from domain-specific carryover phenomena. For example, the persistence (or, inversely, the decay rate) of carryover is an open question. Rahnev et al. (2015) suggested that carryover does not fall off with increasing ITI, but in their procedure (as in ours), confidence was probed on every trial. If being asked to consider and report confidence produces carryover, then ITI may indeed be less relevant to carryover than the number of events between rendered confidence judgments. Requiring confidence judgments on only a subset of recognition decision trials (separated by varying intervals) would allow a test of whether the number of trials between confidence reports influences the degree of confidence carryover.

In addition, the impact of carryover on a given judgment may vary according to how clearly participants are able to assess their confidence in that judgment. When recognition decisions are easier, participants may have more information on which to base a confidence judgment, leaving less "room" for a biasing influence of prior confidence. Participants with higher recognition accuracy, then, should show less carryover than participants with lower accuracy. An analysis of the data from Experiment 1 offers preliminary support for this possibility: the correlation between a participant's overall recognition memory accuracy (measured as hits minus false alarms) and the magnitude of carryover was negative and near significance for old–old pairwise recognition decisions, $r(28) = -.35$, $p = .054$, but not for the three other possible recognition sequences ($ps > .62$). This accuracy measure also reliably correlated with magnitude of carryover from perception to subsequent recognition, $r(28) = -.37$, $p = .046$ (viz., the domain general effect). These findings are conceptually consistent with literature suggesting that subjects are better calibrated for "old" than for "new" decisions (e.g., Weber & Brewer, 2004) and may indicate that positive recognition evidence, and perhaps recollection specifically, mitigates confidence carryover.

Finally, it will be important to investigate how confidence carryover may manifest in other memory tasks and recognition designs. For example, does the effect occur in associative recognition or source memory attribution tasks, and is it stronger for sequential errors than correct responses? Paradigms such as these may be critical for determining whether the phenomenon is largely restricted to fluency-based or familiarity-based memory attributions that are presumably based on a fuzzy sense of evidence magnitude, as opposed to cases in which specific episodic information is sought and recovered.

Conclusion

Judgment confidence is a widely used measure in memory research, but the components of a confidence rating are not fully understood. Our findings provide the first demonstration of two confidence carryover phenomena during recognition memory. Extant data show that confidence carryover in recognition is robust to procedural and stimulus differences and occurs even when subjects are provided valid performance feedback. Finally, a new experiment with interleaved perceptual and memory

judgments suggests that confidence carryover is in part domain-general, occurring across judgments that are not statistically linked in the external environment. Domain-general carryover may therefore reflect more general metacognitive monitoring of internal states or levels of distractability. The data also suggested a domain-specific effect that spanned (and was not mediated by) the intervening judgment from the alternate domain. This effect varied as a function of judgment repetition versus alteration and was simulated by assuming that current recognition or perceptual evidence is pulled toward prior registered evidence from the same domain, when it is sufficiently similar across trials. Far from being solely a function of current memory or perceptual evidence, confidence ratings are the product of multiple influences, including general ease and specific evidence evoked in prior trials.

Author note Correspondence concerning this article should be addressed to Justin Kantner (justin.kantner@csun.edu), Department of Psychology, California State University, Northridge, Northridge, CA 91325, USA, or Ian G. Dobbins (idobbins@wustl.edu), Department of Psychological & Brain Sciences, Washington University in Saint Louis, Saint Louis, MO 63130, USA.

References

Annis, J., & Malmberg, K. J. (2013). A model of positive sequential dependencies in judgments of frequency. *Journal of Mathematical Psychology*, *57*(5), 225–236. doi:10.1016/j.jmp.2013.06.006

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Benjamin, A. S., Diaz, M., & Wee, S. (2009). Signal detection with criterion noise: Applications to recognition memory. *Psychological Review*, *116*(1), 84–115. doi:https://doi.org/10.1037/a0014351

Brown, S. D., Marley, A. J., Donkin, C., & Heathcote, A. (2008). An integrated model of choices and response times in absolute identification. *Psychological Review*, *115*(2), 396–425. doi:https://doi.org/10.1037/0033-295X.115.2.396

Criss, A. H., Malmberg, K. J., & Shiffrin, R. M. (2011). Output interference in recognition memory. *Journal of Memory and Language*, *64*(4), 316–326. doi:https://doi.org/10.1016/j.jml.2011.02.003

Duncan, K., Sadanand, A., & Davachi, L. (2012). Memory's penumbra: Episodic memory decisions induce lingering mnemonic biases. *Science (New York, N.Y.)*, *337*(6093), 485–487. doi:https://doi.org/10.1126/science.1221936

Endl, W., Walla, P., Lindinger, G., Lalouschek, W., Barth, F. G., Deecke, L., & Lang, W. (1998). Early cortical activation indicates preparation for retrieval of memory for faces: An event-related potential study. *Neuroscience Letters*, *240*(1), 58–60.

Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience, 17*(5), 738–743.

Han, S., & Dobbins, I. G. (2008). Examining recognition criterion rigidity during testing using a biased-feedback technique: Evidence for adaptive criterion learning. *Memory & Cognition*, *36*(4), 703–715. doi:https://doi.org/10.3758/MC.36.4.703

Han, S., & Dobbins, I. G. (2009). Regulating recognition decisions through incremental reinforcement learning. *Psychonomic Bulletin & Review, 16*(3), 469–474.

Kantner, J., & Lindsay, D. S. (2012). Response bias in recognition memory as a cognitive trait. *Memory & Cognition*, *40*(8), 1163–1177. doi:https://doi.org/10.3758/s13421-012-0226-0

Kucera, H., & Francis, W. N. (1967). Computational analysis of present-day American English. Providence, RI: Brown University Press.

Macmillan, N. A., & Creelman, C. (2005). Detection theory: A user's guide (2nd ed.). Mahwah, NJ: Erlbaum.

Malmberg, K. J., & Annis, J. (2012). On the relationship between memory and perception: Sequential dependencies in recognition memory testing. *Journal of Experimental Psychology: General*, *141*(2), 233–259. doi:https://doi.org/10.1037/a0025277

Mueller, S. T., & Weidemann, C. T. (2008). Decision noise: An explanation for observed violations of signal detection theory. *Psychonomic Bulletin & Review*, *15*(3), 465–494. doi:https://doi.org/10.3758/PBR.15.3.465

Parks, T. (1966). Signal-detectability theory of recognition-memory performance. *Psychological Review*, *73*(1), 44–58.

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, 162, 8–13. doi:https://doi.org/10.1016/j.jneumeth.2006.11.017

R Core Team. (2016). R: A language and environment for statistical computing [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rahnev, D., Koizumi, A., McCurdy, L. Y., Esposito, M. D., & Lau, H. (2015). Confidence leak in perceptual decision making. *Psychological Science*, *26*(11), 1664–1680. doi:https://doi.org/10.1177/0956797615595037

Ratcliff, R., & Starns, J. J. (2009). Modeling confidence and response time in recognition memory. *Psychological Review*, *116*(1), 59–83. doi:https://doi.org/10.1037/a0014086

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review, 107*(2), 358–367.

Stewart, N., Brown, G. A., & Chater, N. (2005). Absolute identification by relative judgment. *Psychological Review*, *112*(4), 881–911. doi:https://doi.org/10.1037/0033-295X.112.4.881

Stretch, V., & Wixted, J. T. (1998). Decision rules for recognition memory confidence judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1397–1410.

Treisman, M., & Williams, T. C. (1984). A theory of criterion setting with an application to sequential dependencies. *Psychological Review*, *91*(1), 68–111. doi:https://doi.org/10.1037/0033-295X.91.1.68

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.

Weber, N., & Brewer, N. (2004). Confidence-accuracy calibration in absolute and relative face recognition judgments. *Journal of Experimental Psychology: Applied*, *10*(3), 156–172.

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