Journal of Experimental Psychology:
Learning, Memory, and Cognition
2018, Vol. 44, No. 11, 1814–1823

0278-7393/18/$12.00 http://dx.doi.org/10.1037/xlm0000550

Stimulus Effects and the Mediation of Recognition Memory

Ryan M. McAdoo, Kylie N. Key, and Scott D. Gronlund
University of Oklahoma

Two broad approaches characterize the type of evidence that mediates recognition memory: discrete state and continuous. Discrete-state models posit a thresholded memory process that provides accurate information about an item (it is detected) or, failing that, no mnemonic information about the item. Continuous models, in contrast, posit the existence of graded mnemonic information about an item. Evidence favoring 1 approach over the other has been mixed, suggesting the possibility that the mediation of recognition memory may be adaptable and influenced by other factors. We tested this possibility with 2 experiments that varied the semantic similarity of word targets and fillers. Experiment 1, which used semantically similar fillers, displayed evidence of continuous mediation (contrary to Kellen & Klauer, 2015), whereas Experiment 2, which used semantically dissimilar fillers, displayed evidence of discrete mediation. The results have implications for basic theories of recognition memory, as well as for theories of applied domains like eyewitness identification.

Keywords: recognition memory, computational modeling, discrete mediation, continuous mediation

Recognition memory serves many important functions, ranging from recognizing the answer on a multiple-choice test to meteorologists' issuing a tornado warning because they recognize a familiar atmospheric signature to choosing a suspect from a lineup. Cognitive psychologists have spent decades studying the mechanisms involved in simple and complex recognition memory tasks. One avenue of research utilizes mathematical and computational models of memory (for a review see Clark & Gronlund, 1996; also Hintzman, 1991; McClelland, 2009). Some models, like Clark's (2003) WITNESS model, simulate complex events, like choosing a face from a lineup. Others are more foundational, examining the cognitive representation of memory evidence. It is the latter of these that is the focus of the present research. Specifically, we explore whether the memory evidence that is made available for consideration is mediated in an all-or-none, discrete manner (indicating detection or not) or in a continuous manner (reflecting a gradient of latent strength). Although continuous models have come to dominate the literature (Luce, 1997), a resurgence of interest in discrete-state models has occurred (e.g., Bröder & Schütz, 2009).

Research exploring discrete and continuous model classes is important because complex models of memory and decision-making should rely on the correct foundational representation. Of course, this assumes that recognition memory is *always* mediated continuously or discretely. Alternatively, the cognitive representation of recognition memory evidence, or at least the way people weigh this evidence, may be stimulus-dependent (e.g., dependent on the semantic relationships among fillers and targets), task-specific (e.g., *old*−*new,* 2-alternative forced choice [2AFC]), or subject to strategic influences or individual differences. For example, continuous mediation occurs when participants are asked to rank words and faces for the likelihood of having been previously studied (Kellen & Klauer, 2014; McAdoo & Gronlund, 2016), but discrete mediation appears to occur in related recognition tasks like *old*−*new* and 2AFC paradigms that involve confidence judgments (Kellen & Klauer, 2015). The mixed findings are also supported by Kapucu, Macmillan, and Rotello (2010), who compared individual fits of a continuous model—signal detection theory—and Yonelinas's (1994) dual-process model that includes both discrete and continuous contributions. Kapucu et al. concluded that participants used varying strategies, with some participants relying on pure continuous processes and others supplementing a continuous process with a discrete one (dual-process). Our goal was to continue this exploration of discrete and continuous mediation by examining how stimulus relationships between fillers and targets influence the mediation of recognition memory.

We begin by reviewing prototypical discrete-state and continuous models of recognition memory. Next, we review extant tests between the two model classes, including receiver operating characteristic (ROC) analysis, which is thought to support continuous mediation, and two conditional probability measures (Kellen & Klauer, 2014, 2015) that reach opposing conclusions. We then present two new experiments that show that stimulus relationships do indeed influence recognition memory mediation.

Discrete-State and Continuous Models

The typical recognition memory task requires one to indicate whether a stimulus item has been previously encountered. In a standard *old*−*new* paradigm, participants study a series of stimuli (e.g., words, pictures, faces) and, following a brief delay, complete a series of tests that consist of a mixture of previously encountered stimuli (*old*) and stimuli that had not been previously encountered (*new*). Participants indicate whether a presented test item is *old* or *new.* This can be done by way of a simple dichotomous response

This article was published Online First April 19, 2018.

Ryan M. McAdoo, Kylie N. Key, and Scott D. Gronlund, Department of Psychology, University of Oklahoma.

Correspondence concerning this article should be addressed to Ryan M. McAdoo, Department of Psychology, University of Oklahoma, 455 West Lindsey Street, Norman, OK 73019. E-mail: ryan.m.mcadoo-1@ou.edu

(old, new) or by indicating one's confidence (e.g., 1–6 rating) that an item is *old* or *new*. A similar paradigm presents participants with *old* and *new* item pairs (or multiple-item arrays), and the participant is instructed to indicate which item is *old* (*n*-alternative forced choice [*n*AFC]). Performance on both tasks is measured by calculating the proportion of *old* items that are correctly endorsed as *old* (hit rate), and the proportion of *new* items incorrectly endorsed as *old* (false alarm rate).

Discrete-state models assume that the memory evidence made available to the cognitive level of representation is assessed in an all-or-none fashion. An example of the discrete-state class of models is the two-high threshold (2HT) model (Bröder, Kellen, Schütz, & Rohrmeier, 2013). The 2HT model (see Figure 1) assumes that when presented with an *old* item, a participant either detects the item correctly as *old* (with probability D_O) or fails to detect the item (with probability $1 - D_O$), in which case it is assumed that the participant has no mnemonic information available upon which to base a decision and can only guess whether the item is *old* (with probability g) or *new* (with probability $1 - g$). These guesses are strategic and reflect response biases, or one's willingness to endorse an item as *old*. Decisions involving *new* items are governed in the same way, such that a *new* item may be detected as *new* (D_N) or, in the absence of detection ($1 - D_N$), guessed to be *old* (g) or *new* ($1 - g$). Additional parameters govern the probability of selecting a particular level of confidence (denoted as d_{Oi} and g_{Oi} in Figure 1 for detect-*old* and guess-*old,* respectively, and d_{Ni} and g_{Ni} for detect-*new* and guess-*new,* respectively). Different versions of discrete-state models have been proposed (e.g., 1-high threshold; low threshold, Luce, 1963), but they all share an assumption regarding the all-or-none representation of memory evidence: If an item is not detected (as *old* or *new*), there is no mnemonic information available to the participant for that item (complete information loss). Contrast this with continuous models of memory, which assume that there are gradients of information available for both *old* and *new* items. We describe this class of models next.

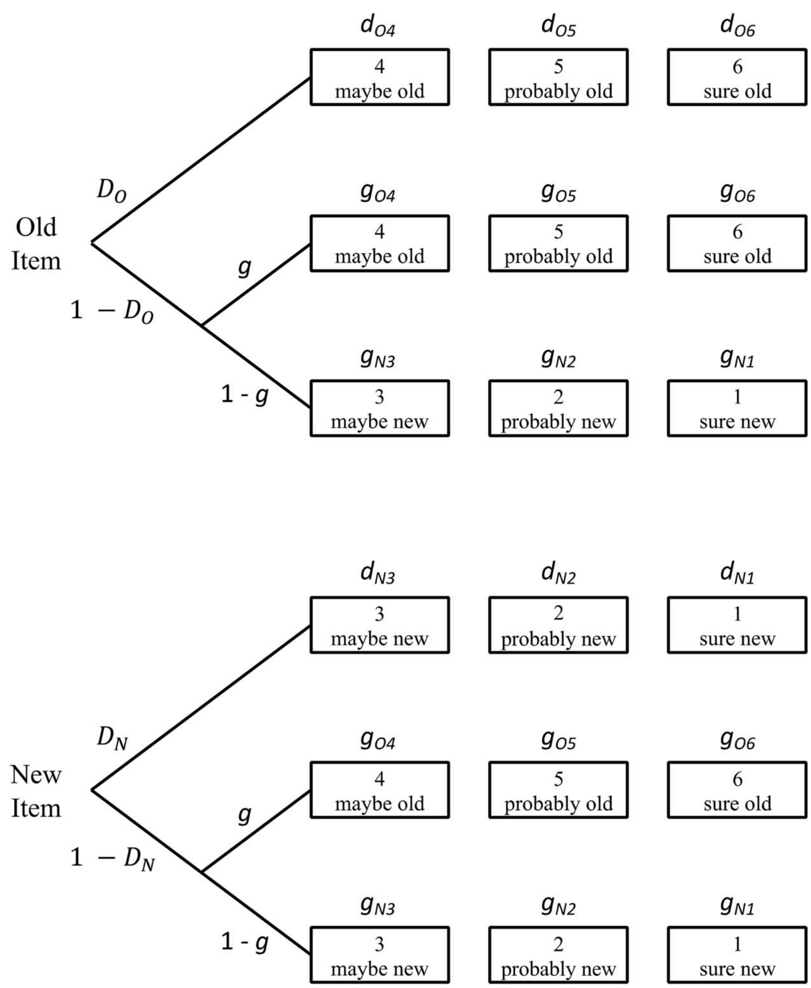

Figure 1. Two-high threshold model. Absent detection of *old* (D_O) or *new* (D_N) items, participants may correctly or incorrectly guess the status of the item (g), reflecting complete information loss. Confidence ratings are mapped by probability parameters, $d_{O4\text{-}6}$, $d_{N1\text{-}3}$ for detection and $g_{O4\text{-}6}$, $g_{N1\text{-}3}$ for guessing states. Confidence mapping parameters are strength independent, meaning that they are unaffected by manipulations of encoding strength, predicting $\theta_{[2|3]}^{weak} = \theta_{[2|3]}^{strong}$.

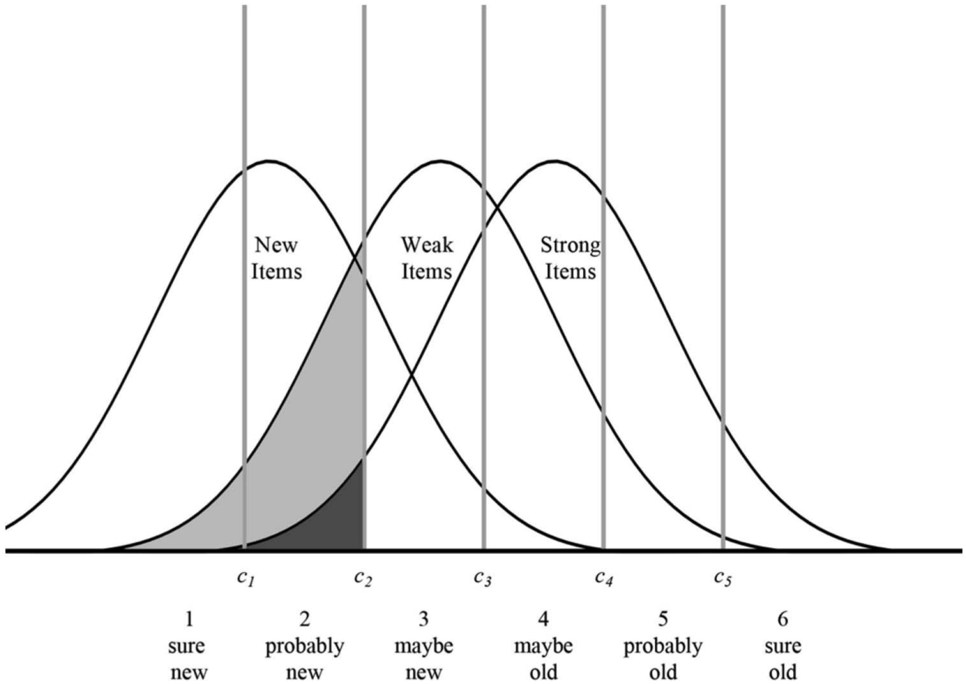

Figure 2. Equal-variance signal detection theory. Shaded gray regions depict the probability of a confidence rating of 1 (*sure new*) or 2 (*probably new*) for *old* items. Shaded areas (light gray = weak targets, dark gray = strong targets) denotes the probability of a target's being rated 1 (*sure new*) or 2 (*probably new*), which is greater for weak than for strong targets. This is because the mean of the strong distribution is greater on average than the mean of the weak distribution, whereas criteria mapping remains constant (strength-independent). This relationship gives rise to the prediction $\theta^{weak}_{[2|3]} > \theta^{strong}_{[2|3]}$. Note that any shape of continuous distribution predicts this same relationship, as long as the strong item distribution is farther from the *new* item distribution than the weak item distribution is from the *new* distribution.

Signal detection theory (SDT; Egan, 1958) is the prototypical example of a continuous model (see Figure 2). SDT assumes that the strengths of memory evidence governing *new* (unstudied) and *old* (studied) items are represented as overlapping *normal* distributions. In the case of an equal-variance signal detection model (depicted in Figure 2), the standard deviations of the *old* and *new* distributions are the same. *Old* items should, on average, have stronger memory strengths than should *new* items, and in Figure 2, the *old* distribution is shifted upward on the *y*-axis to reflect this. (In fact, Figure 2 depicts two classes of *old* items, reflecting strong and weak encoding, which will become relevant later.) When a particular item is presented to a participant, the strength of that item is assessed and compared to a decision criterion. If asked to provide confidence in one's decision, SDT utilizes a set of criteria that correspond to varying levels of confidence in both *new* and *old* decisions (denoted as c_i in Figure 2; note that c_{1-3} correspond to confidence in *new* decisions and c_{4-6} correspond to confidence in *old* decisions). Given these confidence mappings, if an item's strength is greater than c_{4-6}, the item is endorsed as *old* (and *new* if within the boundaries of c_1–c_3).[1] Although Figure 2 depicts equal variance SDT as our prototype, any SDT model makes the same predictions regarding our critical measure (to be defined later), including models that do not assume *normal* distributions of evidence. In addition to purely discrete or continuous mediation, models have been proposed that include continuous and discrete components, like Yonelinas's (1994) dual-process model. This model assumes a discrete *recollection* process, whereby an item is recognized with high confidence, and a continuous *familiarity* process, translating to varying degrees of confidence. However, in the present research, we consider only models that posit continuous or discrete mediation, not both.

The most prominent evidence in favor of continuous mediation involves accounting for various aspects of receiver operator characteristic (ROC) curves (e.g., shape: Malmberg, 2002; slope: Ratcliff, Sheu, & Gronlund, 1992). In a review, Wixted (2007) concluded that continuous models more accurately matched the empirical patterns emanating from ROC analysis. However, objections have been raised to using ROC analysis to adjudicate between discrete and continuous models. In the next section, we review ROC evidence and the concerns raised against it, then discuss two alternative tests that have been developed.

[1] For more information about signal detection theory, see Macmillan and Creelman (2005) or Kellen and Klauer (in press).

Evidence of Discrete-State Versus Continuous Mediation

ROC Analysis

We review only one aspect of ROC analysis—shape, which has been used extensively to test between discrete-state and continuous models. ROC curves plot the hit and false alarm rates across varying levels of response bias (assessed, typically, via confidence). Predictions regarding the shape of ROC curves differ between discrete-state models, which predict linear ROC curves, and continuous models, which predict curvilinear ROC curves. Empirical tests of these predictions reveal ROC curves that are almost always curvilinear, indicating strong evidence in favor of continuous mediation (see reviews by Wixted, 2007; Yonelinas & Parks, 2007). Recently, however, it has been pointed out that relaxing response-mapping assumptions can result in discrete-state models that can produce curvilinear ROC curves. Consequently, ROC curvature does not provide a sufficient critical test between the two model classes (see Bröder & Schütz, 2009; Erdfelder & Buchner, 1998; Malmberg, 2002; Province & Rouder, 2012).

Another method of producing ROC curves is to directly manipulate response bias rather than using confidence ratings as a proxy measure. Unfortunately, this method also seems unable to definitively test between discrete and continuous models. Bröder and Schütz (2009) argued that confidence-rating ROC curvature is a product of variability in confidence scales and that only ROCs produced by manipulating bias (binary response ROCs) are a valid means to test between discrete-state and continuous models. In a series of three experiments, the authors manipulated response bias directly and determined that the 2HT fit binary response ROCs better than did SDT. But Dubé, Starns, Rotello, and Ratcliff (2012) showed that a signal detection model fit binary response ROCs (and reaction time [RT] data) better than did the 2HT model. However, Kellen, Klauer, and Bröder (2013) countered, arguing that discrete-state models fit individual ROCs better than do continuous models when model complexity is considered using measures of minimum description length. The mixed evidence from ROC analysis is not restricted to *old*−*new* recognition tasks. Kellen et al. (2015) fit an unequal-variance SDT and the 2HT model to 2AFC confidence data. They reported, for both word and picture conditions, that the 2HT model outperformed the unequal-variance SDT most of the time. In sum, given the inconsistencies associated with testing between discrete-state and continuous models of recognition memory using ROC data, other critical tests are required. We discuss two of these tests next.

Conditional Probability

Kellen and Klauer (2014) proposed a test using a ranking paradigm and a conditional probability measure that relied on minimal model assumptions. In their study, Kellen and Klauer had participants study a list that contained words presented either once in a weak encoding condition or three times in a strong encoding condition. At test, participants were presented with one *old* word (hereafter, target) and two (Experiment 2) or three (Experiment 1) *new* words (hereafter, fillers). The participants were asked to rank each item from most likely to have been seen before to least likely to have been seen before. The conditional probability (c_2) that a target would be ranked second (given that it was not ranked first) was calculated for each participant.

Kellen and Klauer (2014) proved that discrete-state and continuous models made different predictions regarding the relationship of c_2 values for weak versus strong items. Discrete-state models predict that $c_2^{\text{weak}} = c_2^{\text{strong}}$; if a target is not detected (ranked first), it does not matter whether that target was weakly or strongly encoded; a participant can only guess among the remaining items because no mnemonic information exists regarding items that are not detected. Therefore, the probability of a weak undetected target's being ranked second is the same as the probability of a strong undetected target's being ranked second. In contrast, continuous models predict that $c_2^{\text{weak}} < c_2^{\text{strong}}$. If a target is not recognized as the most likely to have been seen, there still exists memory evidence about the target available to the participant to evaluate. Because strong targets possess stronger evidence, on average, than do weak targets, the probability of the strong targets' being ranked second (given that they were not ranked first) will tend to be greater than for the weak targets.

Kellen and Klauer (2014) found evidence of continuous mediation using this paradigm with words as the test stimuli. McAdoo and Gronlund (2016) replicated these results using faces. However, D. Kellen, Erdfelder, Malmberg, Dubé, and Criss (2016) successfully fit a discrete-state model, the low-threshold model (LTM), to Kellen and Klauer's c_2 ranking data, and found no difference in measures of goodness of fit between the LTM and SDT. We applied the LTM to the McAdoo and Gronlund (2016) data and also found little difference between the LTM and SDT (we return to this point later). These conflicting results necessitate further explorations of continuous and discrete-state models.

Confidence Ratings

Kellen and Klauer (2015) developed another critical test that relied on confidence rating data. Using previously published data from nine *old*−*new* and 2AFC recognition experiments, the authors focused on *old* items that had been judged *new* via a confidence scale that ranged from 1 (*Sure New*) to 6 (*Sure Old*). As in Kellen and Klauer (2014), items in these studies had been either weakly or strongly encoded. Because strength manipulations affect only the probability of detecting an item, if an item fails to be detected, the confidence ratings assigned to these *old* items (i.e., misses) are strength-independent (Province & Rouder, 2012, referred to this as conditional independence, but see Chen, Starns, & Rotello, 2015, for a possible violation of this assumption). Therefore, discrete-state models predict that the probability of an *old* item's being rated 1 or 2 (given it was rated 1, 2, or 3, hereafter $\theta_{[2|3]}$) are the same for weak and strong items. Continuous models, on the other hand, predict that $\theta_{[2|3]}$ is greater for weak items than for strong items. As shown in Figure 2 (see the shaded areas of the weak and strong probability density functions), $\theta_{[2|3]}$is greater (larger area in light gray) for weak items than for strong items (smaller area in dark gray), assuming that the criteria do not vary between weak and strong items. This is because the strength manipulation affects the mean of the *old* item distributions but not the confidence criteria placement in relation to these distributions. That makes the distribution of confidence ratings for *new* items not independent of strength.[2] Kellen and Klauer's (2015) analysis concluded in favor of discrete-state mediation.

The mixed evidence for discrete and continuous mediation is surprising if one expects recognition memory mediation to be unfailingly discrete or continuous. But instead, it is possible that how recognition memory is mediated is adaptable, contingent on a variety of factors. For example, Kellen and Klauer (2015) suggested that task differences could play a role; the c_2 ranking task encourages people to use graded information, whereas confidence ratings may not. But we chose to hold the task constant in our present study and instead chose to examine the effect of the relationships among the tested stimuli.

Present Study

The two experiments in the present study varied stimulus relationships (the similarity between fillers and targets). Our choice to examine this factor was inspired by a finding that arose when we fit the LTM and SDT to the data from McAdoo and Gronlund (2016). We found that the LTM fit the data better than did SDT, despite the c_2 data's supporting continuous mediation (LTM better for 64% of McAdoo & Gronlund's, 2016, participants in Experiment 1 and 52% of participants in Experiment 2, as measured by the G^2 statistic). We also noticed an interesting result that motivated the particular manipulation we conduct. The LTM was unable to produce c_2 values that were less than .5 (if there were three options to be ranked; less than .33 if there were four options to be ranked). However, empirical values of c_2 did sometimes fall below .5 (43% of McAdoo & Gronlund's, 2016, participants in Experiment 1; 27% of the participants in Experiment 2). According to the LTM, c_2 cannot fall below .5 (given three options), because it would mean that targets that are not ranked first are more likely to be ranked third than ranked second. This could occur only if fillers are sometimes better matches to memory than are some of the targets. This highlighted the potential impact of stimulus effects on recognition mediation.

The present study explores the effect that filler−target similarity has on the mediation of recognition memory in a confidence rating task. In his review of models of recognition memory, Malmberg (2008) discussed the concept of *efficiency,* which in this context refers to the idea that participants utilize a recognition strategy that can achieve the desired accuracy in the shortest amount of time. But efficiency does not mean optimality. Rather, we posit efficiency as akin to satisficing—in a given context, participants will adopt a strategy that produces responses that are "accurate enough." We applied this concept to motivate the hypotheses of the current study. For example, in one context, deciding among stimuli relationships may require participants to use continuous mediation to reach some level of desired accuracy, whereas discrete mediation may be sufficient, given different stimuli relationships, to reach a "good enough" level of accuracy.

We designed two experiments that varied filler−target similarity. Both experiments used repetition as the strength manipulation (1× vs. 3×) and utilized single-item tests and confidence judgments (like Kellen & Klauer, 2015). Figure 3 summarizes our manipulations and hypotheses. Experiment 1 uses semantically similar targets and fillers (top half of Figure 3). Suppose a participant studies a list of words that includes *frog* and *chair.* At test, the participant will encounter fillers that are similar to studied targets (e.g., *table* in Figure 3). Semantically similar fillers will elicit some degree of familiarity because of their relation to a studied target (*chair* and *table*). Over the course of the test phase, this may cause participants to more carefully consider the strength of items, resulting in use of the confidence scale in a manner that produces $\theta_{[2|3]}$ estimates consistent with a continuous model. In other words, to be efficient, participants need to use graded information when making confidence judgments *because of* the high similarity of the fillers to the targets. This would result in estimates of $\theta_{[2|3]}$ that reflect the continuous mediation of memory ($\theta_{[2|3]}^{weak} > \theta_{[2|3]}^{strong}$), contrary to the results of Kellen and Klauer (2015).

	Study Phase	Test Phase	Mediation
Experiment 1	…	…	Continuous
	frog	**frog**	
	…	…	
	table	*chair*	
	…	…	
Experiment 2	…	…	Discrete
	frog	**frog**	
	…	…	
	table	*apple*	
	…	…	

Figure 3. Target words are denoted by bolded text and fillers by italicized text. For Experiments 1 and 2, participants studied target words like *frog* and *table.* At test, participants in Experiment 1 were tested on *frog* and on a semantically similar filler, *chair.* Participants in Experiment 2 were also tested on *frog* and on a semantically dissimilar filler, *apple.* We predicted that participants would rely on continuous mediation in Experiment 1 and discrete mediation in Experiment 2.

Experiment 2 used semantically *dissimilar* (or idiosyncratic) targets and fillers (bottom half of Figure 3). Assume that participants study the same target words (including *frog* and *chair*), but at test, participants encounter fillers that are *not* similar to targets (represented by *apple* in Figure 3). In this case, fillers elicit little to no familiarity in relation to any targets. Across test trials, a discretely mediated strategy may be sufficient to achieve reasonable efficiency. In other words, the use of graded information would not be necessary (or maybe even possible, which we discuss

[2] Note that the $\theta_{[2|3]}$ predictions for continuous mediation are not unique to the Gaussian distribution. Any distribution of *graded information* makes similar predictions (e.g., ex-Gaussian, uniform), as long as the weak and strong distributions are separated (see Kellen & Klauer, 2015, for proofs).

later) to maximize accuracy. This would lead to patterns of $\theta_{[2|3]}$ that are consistent with discrete mediation ($\theta_{[2|3]}^{weak} = \theta_{[2|3]}^{strong}$).

Method

General Procedure

The experiments took place in a room with five cubicles, each with a personal computer. Data collection was conducted using E-Prime 2.0 (Psychology Software Tools, 2012). Participants first gave consent and provided demographic information. The experiment began with a study phase, which consisted of a series of words. Participants studied 100 unique words for 750 ms each, with a 500-ms fixation cross between stimuli. Fifty of these words were presented once for a weak encoding manipulation, and 50 were presented thrice for a strong encoding manipulation. This was followed by a distractor task lasting approximately 5 min (40 trials indicating whether two presented numbers summed to a third number). Next, participants completed a test phase, wherein previously studied targets (weak and strong) and new fillers were presented in a random order. The participants rated the likelihood of having seen an item before on a scale ranging from 1 (*Sure New*) to 6 (*Sure Old*).[3] To conclude, we debriefed, thanked, and dismissed each participant. All experiments were approved by the University of Oklahoma Institutional Review Board and followed American Psychological Association ethical guidelines (see the Method sections for each experiment for details unique to that experiment).

Experiment 1

Experiment 1 was designed to test the effect of semantically similar, or confusable, targets and fillers on the mediation of recognition memory in a confidence-rating task. Fillers were chosen to be strongly associated to one of the targets (see the detailed method later). We hypothesized that presenting confusable fillers would cause participants to exhibit continuously mediated patterns of recognition, reflecting the use of graded information in the absence of positive recognition.

Participants. Participants (N = 90) were University of Oklahoma introductory psychology students, who completed this study in exchange for partial course credit. They were mostly female (N = 72), with an average age of 18.8 years. Self-reported ethnicity was Caucasian (74%), American Indian/Alaska Native (7%), Asian (11%), African American (1%), and No Response (6%).

Materials. Words were drawn from the University of South Florida Free Association Norms database (Nelson, McEvoy, & Schreiber, 1998).[4] All words from the database were downloaded and sorted to include only English nouns. Only nouns that had associated words with forward-strength scores of at least .40 (at least 40% of participants provided the same associated word when prompted) were selected. We sorted this *new* list of nouns by cue frequency (per Kučera & Francis, 1967) and selected only nouns with a frequency of at least 3 (this cutoff was arbitrarily chosen; average frequency score = 35.8). This final list was then randomized, and the first 100 nouns were chosen to be included as targets. The top associate (measured by the forward-strength score) for each of the 100 target nouns was found in Appendix B in the Nelson et al. (1998) database and served as semantically similar fillers (fillers with strong association to targets). For example, if *arrow* were a target, the word with the top association score, *bow,* would be selected as its semantically similar filler.

Experiment 2

Experiment 2 was designed to test the effect of semantically dissimilar, or idiosyncratic, fillers on the mediation of recognition memory. Fillers were chosen to have weak associations with targets (see the detailed method later). We hypothesized that idiosyncratic fillers would result in participants' exhibiting discretely mediated recognition.

Participants. University of Oklahoma (N = 79) and Canisius College (N = 6) introductory psychology students completed this study in exchange for partial course credit. They were mainly female (N = 65), with an average age of 19.4 years. Self-reported ethnicity was Caucasian (66%), American Indian/Alaska Native (7%), Asian (13%), African American (4%), Middle Eastern (4%), and No Response (6%).

Materials. Target words for Experiment 2 were identical to those in Experiment 1 (participants who participated in Experiment 1 were excluded from Experiment 2). Fillers were taken from Appendix D of the Nelson et al. (1998) database. This appendix provides idiosyncratic words generated by only one participant during free association. This ensured that only a weak association existed between targets and fillers. For example, if *arrow* were served as a target, *thief* would serve as its idiosyncratic filler.

Results

Experiment 1

Raw data are available through the Open Science Framework at https://osf.io/hp9ft/. Table 1 shows the summary statistics and significance tests for Experiments 1 and 2.[5] The strength manipulation for Experiment 1 was successful, with the average hit rate (proportion of targets rated 4, 5, or 6) significantly greater for strong (M = .71, SD = .17) than for weak (M = .54, SD = .16), $t(89) = 11.84$, $p < .001$, Cohen's d = 1.03, targets. The mean false alarm rate was .41 (SD = .19). Bayesian analysis (Kruschke's, 2013, Bayesian Analysis Supersedes the t-Test [BEST] method) revealed the same pattern. Markov chain Monte Carlo (MCMC) estimation predicted an average (mean) difference of the means (*Diff*) to be .18 (P[$Diff > 0$] > 99.9%, 95% highest density interval [HDI: .15, .21]). The interval does not include 0, indicating a strong posterior likelihood that the mean hit rates are different for

[3] The full confidence scale was 1 (*Sure New*), 2 (*probably new*), 3 (*maybe new*), 4 (*maybe old*), 5 (*probably old*), and 6 (*Sure Old*).

[4] For information on how the database was compiled, visit http://w3.usf.edu/FreeAssociation/

[5] Kellen and Klauer (2015) proved that discrete and continuous models predicted that patterns of $\theta_{[1|2]}$ (probability of a rating of 1, given a 1 or 2) were identical to $\theta_{[2|3]}$. Many participants did not report a confidence of 1 for misses, making estimates of $\theta_{[1|2]}$ more unstable than estimates of $\theta_{[2|3]}$. In Experiment 1, 16.5% of participants gave no ratings of 1 for weak targets, and 34% of participants gave no ratings of 1 for strong targets. In Experiment 2, 29% of participants gave no ratings of 1 for weak targets, and 44% of participants gave no ratings of 1 for strong targets. Because of the lack of data and the resulting variability in $\theta_{[1|2]}$ estimates this situation creates, we focused on $\theta_{[2|3]}$ in the present analyses.

Table 1
Descriptive and Inferential Statistics for Experiments 1 and 2

Variable	n	*M* (*SD*) Weak	Strong	t	*Diff* [95% HDI]	P(*Diff* > 0)	BF
Hit rate							
Exp. 1	90	.54 (.16)	.71 (.17)	11.84***	.18 [.15, .21]	>99.9%	>100
Exp. 2	84	.59 (.17)	.75 (.19)	11.14***	.16 [.10, .22]	>99.9%	>100
$\theta_{[2\|3]}$ (all)							
Exp. 1	87	.67 (.24)	.61 (.31)	3.22***	.07 [−.02, .15]	94.2%	13.81
Exp. 2	81	.68 (.25)	.67 (.29)	.30	.01 [−.08, .09]	54.4%	.13
$\theta_{[2\|3]}$ (diagnostic)							
Exp. 1	80	.67 (.25)	.60 (.31)	3.60***	.08 [−.01, .17]	95.5%	41.34
Exp. 2	72	.67 (.26)	.66 (.30)	.79	.02 [−.07, .11]	64.0%	.18

Note. *Diff* = difference (referring to the mean posterior estimate of the difference of the means); HDI = highest density interval (referring to the 95% highest density interval of the posterior distribution obtained using the Bayesian Analysis Supersedes the *t*-Test method); BF = Bayes factors; Exp. = experiment.
*** $p = .001$.

strong and weak items.[6] We also calculated Bayes factors (BF) using the method developed by Rouder, Speckman, Sun, Morey, and Iverson (2009); we used the criteria developed by Jeffreys (1961) for interpretation. The BF for differences in hit rate for Experiment 1 was >100, indicating decisive evidence for a difference between weak and strong hit rates.

Of primary interest is the relationship between $\theta_{[2|3]}^{weak}$ and $\theta_{[2|3]}^{strong}$. Three participants could not be analyzed due to providing no ratings of 1, 2, or 3 for strong items. A paired-samples t test revealed a significant difference between $\theta_{[2|3]}^{weak}$ ($M = .67$, $SD = .24$) and $\theta_{[2|3]}^{strong}$ ($M = .61$, $SD = .3$), $t(86) = 3.22$, $p = .001$, Cohen's $d = .22$. BEST analysis estimated a mean *Diff* of .07 (P[Diff > 0] = 94.2%, 95% HDI [−.02, .15]), which does include 0 but strongly favors a difference greater than 0. BF was equal to 13.81, indicating strong evidence for a difference in $\theta_{[2|3]}^{weak}$ and $\theta_{[2|3]}^{strong}$. This evidence indicates support for continuous mediation of memory, contrary to the results of Kellen and Klauer (2015). Like these authors, we also separately analyzed only participants who responded to the strength manipulation (hereafter, diagnostic participants). Seven participants were removed, leaving 80 diagnostic participants (89% of the original sample). A paired-samples t test supported a difference in $\theta_{[2|3]}^{weak}$ ($M = .67$, $SD = .25$) and $\theta_{[2|3]}^{strong}$ ($M = .60$, $SD = .31$), $t(86) = 3.59$ $p < .001$, Cohen's $d = .25$. BEST analysis estimated a mean *Diff* of .08 (P[*Diff* > 0] = 95.5%, 95% HDI [−.01, .17]). Although the HDI contained 0, the posterior probability that *Diff* is greater than 0 was high (95.5%). The BF was equal to 41.34, indicating very strong evidence for a difference in $\theta_{[2|3]}^{weak}$ and $\theta_{[2|3]}^{strong}$ for diagnostic participants. Results of Experiment 1 strongly supported continuous mediation in recognition memory when fillers resembled targets.

Experiment 2

As in Experiment 1, the strength manipulation for Experiment 2 was successful, with the average hit rate greater for strong ($M = .75$, $SD = .19$) than for weak ($M = .59$, $SD = .17$) targets, based on a paired-samples t test, $t(83) = 11.14$, $p < .001$, Cohen's $d = .89$. The mean false alarm rate was .37 ($SD = .20$). BEST analysis estimated a mean *Diff* to be .16 (P[*Diff* > 0] > 99.9%, 95% HDI [.10, .22]), which does not include 0, indicating strong posterior likelihood that the mean hit rates are different for strong and weak items. BF was >100, indicating decisive evidence in favor a difference in weak and strong hit rates in Experiment 2.

Our primary interest was in the relationship between $\theta_{[2|3]}^{weak}$ and $\theta_{[2|3]}^{strong}$. Three participants could not be analyzed, due to no ratings of 1, 2, or 3 for strong items. There was no significant difference between $\theta_{[2|3]}^{weak}$ ($M = .68$, $SD = .25$) and $\theta_{[2|3]}^{strong}$ ($M = .67$, $SD = .29$), $t(80) = .30$, $p = .77$. BEST analysis estimated a mean *Diff* of .01 (P[*Diff* > 0] = 54.4%, 95% HDI [−.08, .09]), which includes 0. BF was equal to .13, indicating substantial evidence in favor of the null hypothesis of no difference between $\theta_{[2|3]}^{weak}$ and $\theta_{[2|3]}^{strong}$.This evidence indicates support for discrete mediation of memory, contrary to the results of Experiment 1. As in Experiment 1 and in Kellen and Klauer (2015), we separately analyzed only diagnostic participants. Nine participants were removed, leaving 72 diagnostic participants (87% of the original sample). There was still no difference between $\theta_{[2|3]}^{weak}$ ($M = .67$, $SD = .26$) and $\theta_{[2|3]}^{strong}$ ($M = .66$, $SD = .30$), $t(71) = .79$, $p = .42$. BEST analysis estimated a mean *Diff* of .02 ([P(*Diff* > 0] = 64.0%, 95% HDI [−.07, .11]). BF was equal to .18, indicating substantial evidence in favor the null hypothesis of no difference between $\theta_{[2|3]}^{weak}$ and $\theta_{[2|3]}^{strong}$. Unlike Experiment 1, Experiment 2 revealed strong evidence for the discrete mediation of recognition memory.

Discussion

Using a critical test with minimal model assumptions developed by Kellen and Klauer (2015), Experiment 1 showed that semantically similar targets and fillers prompted participants to utilize continuous mediation, whereas semantically dissimilar targets and fillers in Experiment 2 prompted participants to rely on discrete mediation. The construct of efficiency (Malmberg, 2008) is one possible rationale for why these differences in stimulus relationships induce these differences in mediation. When fillers are semantically similar to targets, it may be efficient (and necessary) for participants to adopt continuous mediation, assessing the latent

[6] The highest density interval is the interval in which 95% of parameter estimates fall in the posterior distribution estimated by Markov chain Monte Carlo.

strength of fillers and targets, to achieve an acceptable level of accuracy. But when fillers are semantically dissimilar to targets, it may be efficient (and sufficient) to adopt discrete mediation, treating misses as instances of complete information loss. The similarity relationship between targets and fillers appears to influence how recognition memory is mediated.

Note again that our interpretation of efficiency does not require that participants maximize accuracy, only that participants seek a level of accuracy that they deem adequate. That is, participants may engage in satisficing, whereby they select a strategy that can produce reasonable accuracy for the task at hand. Likewise, efficiency does not necessitate that the level of performance be equal between the two experiments (and it was not). Assuming an SDT measurement model and examining only the diagnostic participants, we measured discriminability via d' in Experiments 1 and 2. For d'_{weak}, there were reliable differences between Experiments 1 ($M = .40$, $SD = .46$) and 2 ($M = .68$, $SD = .57$), $t(154) = 3.37$, $p < .001$, Cohen's $d = 0.54$, and for d'_{strong}, there were also reliable differences between Experiments 1 ($M = 1.00$, $SD = .60$) and 2 ($M = 1.36$, $SD = .80$), $t(154) = 3.10$, $p < .005$, Cohen's $d = .50$. This raises the possibility that task difficulty may be another factor to consider in the context of discrete and continuous mediation. Participants in the harder task (Experiment 1) may have had to utilize graded information to reach satisfactory levels of accuracy. Future experiments should strive to equate discriminability performance across tasks to test this explanation.

Other types of target−filler similarity could be manipulated to understand the range of possible stimulus effects. For example, will orthographic similarity (number of letters that words share; e.g., *cat* and *bat* vs. *can* and *bat*) influence mediation differently from semantic similarity? Prior research has shown that orthographic similarity has stronger effects on memory performance than does semantic similarity (Gillund & Shiffrin, 1984). Different recognition tasks may also influence mediation. Province and Rouder (2012), one of the studies reanalyzed by Kellen and Klauer (2015), used semantically similar fillers and targets but found evidence that favored discrete mediation (contrary to the results of our Experiment 1). However, Province and Rouder used a 2AFC paradigm, whereas we used an *old−new* paradigm. Task differences like this also may explain why Kellen and Klauer (2014) and McAdoo and Gronlund (2016) found evidence of continuous mediation using a ranking task, but Kellen and Klauer (2015) found evidence of discrete mediation using a confidence rating task. Studies that separately vary both stimuli relationships and task difference can isolate these effects.

Future Directions and Implications

Figure 4 plots differences in hit rates versus $\theta_{[2|3]}$ for weak and strong items for each participant. Panel A shows the results for Experiment 1, and Panel B shows the results for Experiment 2. Evidence for continuous mediation is signaled if a majority of the participants fall in the upper left quadrant (gray-shaded points), depicting a successful strength manipulation and $\theta^{\text{weak}}_{[2|3]}$ that is greater than $\theta^{\text{strong}}_{[2|3]}$. Although more participants in Experiment 1 displayed evidence of continuous mediation (53.4%) than in Experiment 2 (40%), there clearly was a good deal of variability across the two experiments.[7] Although our hypothesis was supported statistically, the variability depicted in Figure 4 (also apparent in Kellen & Klauer, 2015, Figure 6) hints that some participants in Experiment 1, when presented with semantically similar fillers and targets, appeared to rely on discrete mediation. Conversely, some participants in Experiment 2, despite the semantically dissimilar fillers, appeared to rely on continuous mediation. A goal for future research is to determine what portion of this variability is statistical noise and what portion has theoretical import.

One way to disentangle the potential contributions to the variability in Figure 4 is to specify the responsible psychological processes. We propose that semantically similar fillers cause participants to more meaningfully assign confidence ratings and semantically dissimilar fillers cause participants to use the confidence scale less effectively. However, we have no way of knowing whether that is how participants approached this task without some form of verbal protocol analysis or additional dependent measures (e.g., electroencephalogram, *remember−know* judgments). In future studies involving this and related paradigms, assessments of participants' strategies would produce richer data sets that allow researchers to tease apart variables that may impact individual differences in mediation strategies, as well as provide a more complete picture of why stimuli similarity impacts mediation.

The current results have important implications for basic research and applied settings. Understanding the foundational representations of memory should serve as the starting point for complex models of memory and decision-making. If recognition memory mediation is influenced by tasks and stimuli, these factors need to be incorporated into complex process models. Applied fields, like eyewitness memory, will benefit from a better understanding of memory mediation. For example, some eyewitness memory researchers have claimed that when eyewitnesses view a lineup, they rely on guessing processes when a face is not immediately recognized (Lindsay & Wells, 1985; Wells, Steblay, & Dysart, 2012). The results of McAdoo and Gronlund (2016) intimated that this was not the case and that memory was continuously mediated in the absence of strong recollection (using the c_2 paradigm). As a consequence, McAdoo and Gronlund argued that it was discrediting to claim that eyewitnesses guessed (a discrete process); rather, eyewitnesses respond with differing levels of confidence, based on assessing continuously varying latent memory strengths. But are there circumstances in which an eyewitness utilizes a discretely mediated strategy that incorporates guessing (e.g., when viewing a lineup that contains fillers that do not resemble the suspect)? If so, what factors mitigate these strategies? One solution suggested by the present research is to use similar fillers (already a good idea for other reasons). Similar fillers should be more likely to induce a reliance on continuously mediated strategies, which would result in more meaningful assignments of lineup decisions to confidence ratings.

[7] We simulated a signal detection model, using a performance level comparable to our data, and varied criteria positioning to produce more versus fewer misses. The majority of simulated subjects always fall in the upper-left-hand quadrant (signaling continuous mediation), but criteria positioning that resulted in fewer misses produced much more variability (i.e., points that fell outside the upper-left quadrant). An alternative version of Figure 4, which includes diagnostic participants who made 10 more misses, shows much less variability.

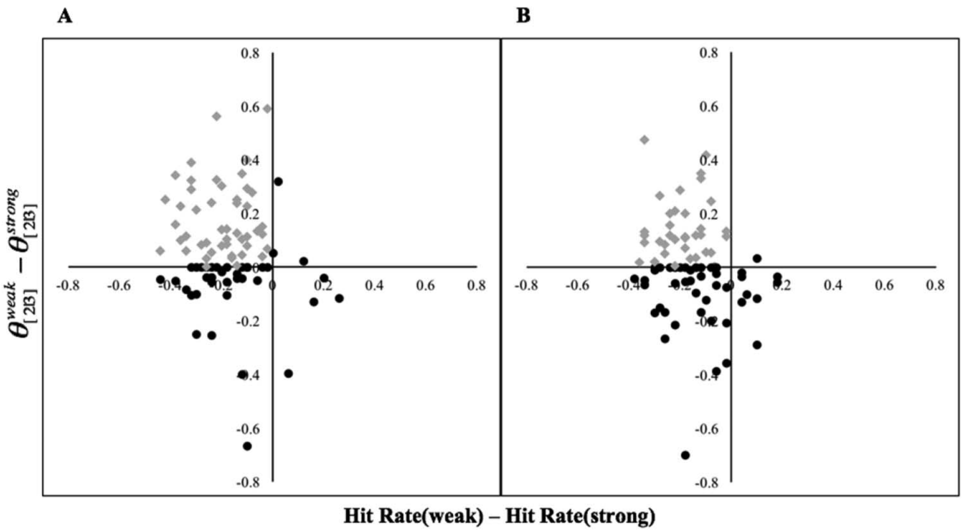

Figure 4. Scatterplots indicating differences in hit rates versus $\theta_{[2|3]}$ for each participant in Experiment 1 (Panel A) and Experiment 2 (Panel B). Gray diamonds indicate data consistent with continuous mediation (i.e., $\theta^{weak}_{[2|3]} - \theta^{strong}_{[2|3]} < 0$ and Hit Rate(weak) $-$ Hit Rate(strong) > 0).

Confidence ratings play an important role in eyewitness identification research (for recent reviews see Wixted, Mickes, Clark, Gronlund, & Roediger, 2015; Wixted & Wells, 2017) and in the assessment of eyewitness performance using ROC analysis (Gronlund, Wixted, & Mickes, 2014; Wixted & Mickes, 2012). Most relevant given the current findings are lineup decisions from nonchoosers, those who reject a lineup. The confidence−accuracy relationship has been shown to be weak for nonchoosers (Brewer & Wells, 2006; Sporer, Penrod, Read, & Cutler, 1995), meaning that nonchoosers' confidence is not a good proxy for their accuracy. This could be because nonchoosers are discretizing their confidence judgments. Particularly at the low end of the confidence scale, eyewitnesses (even those who choose from the lineup) often exhibit underconfidence (more accurate than confident). This is what one would expect if participants were not meaningfully assigning responses to confidence bins over this portion of the scale. In other words, eyewitnesses may find it sufficient to discretize their low confidence judgments, especially if they surmise that only high confidence ratings are likely to be used in court cases (Cutler, Penrod, & Dexter, 1990). In sum, our results suggest a potential role of stimulus similarity on the measurement of eyewitness confidence, which would hold policy implications for the collection and utilization of eyewitness evidence by the criminal justice system.

Conclusion

Given the large body of research regarding the question of whether recognition memory is mediated discretely or continuously, one would think researchers had come to a consensus. However, the evidence in favor of one class of models over the other remains mixed. This suggests to us that the question is not "is recognition memory continuous or discrete?" but rather "*when* and *why* is recognition memory continuous or discrete?" We manipulated one variable that appears to influence mediation: stimuli similarity. We found that recognition memory is mediated continuously when targets and fillers are semantically similar but is mediated discretely when targets and fillers are semantically dissimilar. This finding has implications for improving recognition memory theory, as well as for improving the understanding of applied domains like eyewitness identification.

References

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

Bröder, A., Kellen, D., Schütz, J., & Rohrmeier, C. (2013). Validating a two-high-threshold measurement model for confidence rating data in recognition. *Memory, 21,* 916–944. http://dx.doi.org/10.1080/09658211.2013.767348

Bröder, A., & Schütz, J. (2009). Recognition ROCs are curvilinear—or are they? On premature arguments against the two-high-threshold model of recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 587–606. http://dx.doi.org/10.1037/a0015279

Chen, T., Starns, J. J., & Rotello, C. M. (2015). A violation of the conditional independence assumption in the two-high-threshold model of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41,* 1215–1222. http://dx.doi.org/10.1037/xlm0000077

Clark, S. E. (2003). A memory and decision model for eyewitness identification. *Applied Cognitive Psychology, 17,* 629–654. http://dx.doi.org/10.1002/acp.891

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review, 3,* 37–60.

Cutler, B. L., Penrod, S. D., & Dexter, H. R. (1990). Juror sensitivity to eyewitness identification evidence. *Law and Human Behavior, 14,* 185–191. http://dx.doi.org/10.1007/BF01062972

Dubé, C., Starns, J. J., Rotello, C. M., & Ratcliff, R. (2012). Beyond ROC curvature: Strength effects and response time data support continuous-evidence models of recognition memory. *Journal of Memory and Language, 67,* 389–406. http://dx.doi.org/10.1016/j.jml.2012.06.002

Egan, J. P. (1958). *Recognition memory and the operating characteristic* (Technical Note AFCRC-TN-58–51). Bloomington, IN: Indiana University, Hearing and Communication Laboratory.

Erdfelder, E., & Buchner, A. (1998). Comment: Process-dissociation measurement models: Threshold theory or detection theory? *Journal of Experimental Psychology: General, 127,* 83–96. http://dx.doi.org/10.1037/0096-3445.127.1.83

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67. http://dx.doi.org/10.1037/0033-295X.91.1.1

Gronlund, S. D., Wixted, J. T., & Mickes, L. (2014). Evaluating eyewitness identification procedures using receiver operating characteristic analysis. *Current Directions in Psychological Science, 23,* 3–10. http://dx.doi.org/10.1177/0963721413498891

Hintzman, D. L. (1991). Why are formal models useful in psychology? In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp. 39–56). Hillsdale, NJ: Erlbaum.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, United Kingdom: Oxford University Press.

Kapucu, A., Macmillan, N. A., & Rotello, C. M. (2010). Positive and negative remember judgments and ROCs in the plurals paradigm: Evidence for alternative decision strategies. *Memory & Cognition, 38,* 541–554. http://dx.doi.org/10.3758/MC.38.5.541

Kellen, D., & Klauer, K. C. (in press). Elementary Signal Detection and Threshold Theory. *The Stevens' Handbook of Experimental Psychology and Cognitive Neurocience*.

Kellen, D., & Klauer, K. C. (2014). Discrete-state and continuous models of recognition memory: Testing core properties under minimal assumptions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40,* 1795–1804. http://dx.doi.org/10.1037/xlm0000016

Kellen, D., & Klauer, K. C. (2015). Signal detection and threshold modeling of confidence-rating ROCs: A critical test with minimal assumptions. *Psychological Review, 122,* 542–557. http://dx.doi.org/10.1037/a0039251

Kellen, D., Klauer, K. C., & Bröder, A. (2013). Recognition memory models and binary-response ROCs: A comparison by minimum description length. *Psychonomic Bulletin & Review, 20,* 693–719. http://dx.doi.org/10.3758/s13423-013-0407-2

Kellen, D., Singmann, H., Vogt, J., & Klauer, K. C. (2015). Further evidence for discrete-state mediation in recognition memory. *Experimental Psychology, 62,* 40–53. http://dx.doi.org/10.1027/1618-3169/a000272

Kellen, D., Erdfelder, E., Malmberg, K. J., Dubé, C., & Criss, A. H. (2016). The ignored alternative: An application of Luce's low-threshold model to recognition memory. *Journal of Mathematical Psychology, 75,* 86–95. http://dx.doi.org/10.1016/j.jmp.2016.03.001

Kruschke, J. K. (2013). Bayesian estimation supersedes the *t* test. *Journal of Experimental Psychology: General, 142,* 573–603. http://dx.doi.org/10.1037/a0029146

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Hanover, NH: Dartmouth Publishing Group.

Lindsay, R. C. L., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology, 70,* 556–564. http://dx.doi.org/10.1037/0021-9010.70.3.556

Luce, R. D. (1963). A threshold theory for simple detection experiments. *Psychological Review, 70,* 61–79. http://dx.doi.org/10.1037/h0039723

Luce, R. D. (1997). Several unresolved conceptual problems of mathematical psychology. *Journal of Mathematical Psychology, 41,* 79–87. http://dx.doi.org/10.1006/jmps.1997.1150

Malmberg, K. J. (2002). On the form of ROCs constructed from confidence ratings. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 380–387. http://dx.doi.org/10.1037/0278-7393.28.2.380

Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitve Psychology, 57,* 335–384.

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

McAdoo, R. M., & Gronlund, S. D. (2016). Relative judgment theory and the mediation of facial recognition: Implications for theories of eyewitness identification. *Cognitive Research: Principles and Implications, 1,* 11. http://dx.doi.org/10.1186/s41235-016-0014-7

McClelland, J. L. (2009). The place of modeling in cognitive science. *Topics in Cognitive Science, 1,* 11–38. http://dx.doi.org/10.1111/j.1756-8765.2008.01003.x

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*. Retrieved January 8, 2008, from http://w3.usf.edu/FreeAssociation/

Province, J. M., & Rouder, J. N. (2012). Evidence for discrete-state processing in recognition memory. *Proceedings of the National Academy of Sciences of the United States of America, 109,* 14357–14362. http://dx.doi.org/10.1073/pnas.1103880109

Psychology Software Tools. (2012). E-Prime 2.0. [Computer Software]. Pittsburgh, PA: Author.

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

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16,* 225–237. http://dx.doi.org/10.3758/PBR.16.2.225

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

Wells, G. L., Steblay, N. K., & Dysart, J. E. (2012). Eyewitness identification reforms: Are suggestiveness-induced hits and guesses true hits? *Perspectives on Psychological Science, 7,* 264–271. http://dx.doi.org/10.1177/1745691612443368

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review, 114,* 152–176. http://dx.doi.org/10.1037/0033-295X.114.1.152

Wixted, J. T., & Mickes, L. (2012). The field of eyewitness memory should abandon probative value and embrace receiver operating characteristic analysis. *Perspectives on Psychological Science, 7,* 275–278.

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

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

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,* 1341–1354. http://dx.doi.org/10.1037/0278-7393.20.6.1341

Yonelinas, A. P., & Parks, C. M. (2007). Receiver operating characteristics (ROCs) in recognition memory: A review. *Psychological Bulletin, 133,* 800–832. http://dx.doi.org/10.1037/0033-2909.133.5.800

Received August 15, 2017
Revision received November 28, 2017
Accepted December 29, 2017 ■