UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

RECOGNITION MEMORY IS FUNDAMENTALLY CONTINUOUS, AND STRATEGIC DISCRETIZATION DOES NOT CHANGE THIS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

In partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

RYAN MICHAEL MCADOO
Norman, Oklahoma
2019

RECOGNITION MEMORY IS FUNDAMENTALLY CONTINUOUS, AND STRATEGIC DISCRETIZATION DOES NOT CHANGE THIS

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF PSYCHOLOGY

BY

Dr. Scott Gronlund, Chair

Dr. Sepideh Stewart

Dr. Edward Cokely

Dr. Robert Terry

Dr. Michael Wenger

This dissertation is dedicated to my late father, John Richard McAdoo,
my future wife, Lisa Ann De Stefano,
and my beautiful daughter, Daniela

Acknowledgements

I would like to thank Dr. Scott Gronlund for all his invaluable guidance and advice throughout my graduate career. His skills, encouragement, and dedication have provided me with the tools and skills I require for my future in science. I look forward to continuing to view him as not only a wonderful mentor, but a trusted colleague.

I would also like to thank the members of my committee for their feedback on this project and their continuing support in my future endeavors. Thank you, as well, to my undergraduate research assistants who made data collection for this project possible and to my graduate level lab mates for their endless willingness to discuss research ideas.

I would like to thank my family. My mother, siblings, and grandparents who never stopped encouraging me and believing in me, even though they don’t quite understand what I do.

Most of all, I want to thank my future wife, Lisa De Stefano for her endless patience and support from day one. I have been so lucky to take this journey with her. She has been counselor, proof-reader, and cheerleader. Thank you. And finally I want to thank my daughter, Daniela, for reminding me what is important in life, and always being the voice of wisdom at such a young age. Love you both to Titan and back.

Table of Contents

List of Tables

List of Figures

Abstract

Recognition memory research has long focused on whether it is mediated by discrete or continuous processes. Recent research has shown that the picture is more complex. Recognition memory is not continuous *or* discrete, but may be treated as either, depending on a confluence of internal and external factors. This strategic discretization assumes that the memory signal available to the decision makers is fundamentally continuous, but this has not been empirically supported. Experiment 1 tested this assumption and found that recognition memory is fundamentally continuous. Experiment 2 was designed to test whether strategic discretization changes this signal from continuous to discrete and found that it did not. The results of these studies further solidifies the understanding of how recognition memory is mediated, and also suggests future directions for answering important, applied questions.

Introduction

Recognition memory decisions play important roles in our everyday lives. These range from identifying your favorite actor or actress in a movie, to more life-changing events such as choosing a perpetrator from a police lineup. For decades, cognitive psychologists have been attempting to discern how we make these recognition-based decisions. One avenue of research involves using formal, computational and mathematical models to describe unseen processes that underlie these memory decisions (see McClelland, 2009 for review). The most fundamental of these models attempt to ascertain how mnemonic evidence is represented to the cognitive level of analysis. It is these types of models, and their implications, that form the foundation of the following work.

Two classes of models have been examined regarding how recognition memory evidence is represented at the cognitive level of analysis: *discrete* and *continuous*. Discrete models assume that evidence is available in an all-or-none manner. Continuous models assume that memory is available along a continuum of graded evidence. To illustrate the difference, one can use the metaphor of recognition memory evidence being like the illumination from a lightbulb. Discrete models in this metaphor operate like a standard on-off switch. One either has a memory of a stimulus (switch is on, light is lit) or no memory of the stimulus (switch is off, light is dark). Continuous models operate like a dimmer switch; stimuli possess gradations of evidence. Some stimuli elicit a strong memory (dimmer switch is high, bulb is bright), whereas others elicit a weaker memory (dimmer switch is low, bulb is dim). Understanding, and predicting when, recognition memory evidence is discrete or continuous, has important implications for how researchers understand the functioning of recognition memory, both theoretically and practically.

For many years, recognition memory research focused on determining whether memory is *always* mediated as continuous evidence, discrete evidence, or some mixture of two (e.g., some dual-process models). But recent research (reviewed below) suggests that how memory is mediated depends on a number of factors including choice architecture and stimulus relationships. This has led to the creation of a new framework of recognition memory that envisions a recognition memory decision as dynamic process, one in which mediation is influenced by factors both external and internal to the decision maker. This raises the question of what exactly prior research was telling us about the true nature of the mnemonic evidence underlying recognition memory. To develop this point, I need to distinguish between the inherent nature of the underlying evidence and how that evidence can be strategically *treated*. Extending the lightbulb metaphor a bit further makes clear what I mean. Consider someone who possesses a dimmer switch, but either has the switch turned on all the way or off all the way. In this circumstance, the amount of illumination *possible* is continuous, but it is *treated* as discrete. Recognition memory may operate in the same way: recognition memory evidence may be fundamentally continuous, and available to the decision maker, but treated as discrete when strategically beneficial. Note that this possibility makes more sense than the alternative – that recognition evidence is inherently discrete but may be treated as continuous. If memory is inherently discrete, it is all-or-none, and cannot be examined continuously (e.g., one cannot get different grades of illumination with only a binary on-off switch).

One purpose of the current study is to provide evidence for the hypothesis that memory is inherently continuous but can be strategically discretized. Such a finding would help reconcile conflicting literature that finds evidence for both discrete and continuous mediation, and points to a more complete understanding of how recognition memory tasks operate. Another, related purpose, is to investigate whether discretization of continuous memory evidence for strategic reasons changes this evidence from graded to all-or-none.

The remainder of this paper proceeds as follows. First, I will outline in more detail the hallmarks of discrete and continuous models using the two-high-threshold model (2HTM) and signal detection theory (SDT) as prototypical discrete and continuous models, respectively. Next, I will discuss recent work that suggests that memory evidence may be treated as discrete *or* continuous, depending on various circumstances. Third, I will outline a framework of recognition memory that integrates these new findings and begins to reconcile conflicting research. I will then discuss evidence supporting the hypothesis that memory may be fundamentally continuous but can be strategically discretized by participants. Two experiments designed to address this hypothesis and enhance understanding of the underlying cognitive mechanisms involved in recognition memory decisions, are detailed. Finally, I discuss implications of this new framework of recognition memory for both basic recognition memory research and applied research including recognition memory and medical decision making.

Discrete and continuous models

Although there are many instantiations of discrete and continuous models (e.g., one-high-threshold; low-threshold, Luce, 1963; diffusion model, Ratcliff, 1978; general recognition theory, Ashby & Townsend, 1986), the major distinction between discrete and continuous models (and the distinction that will inform the following work) is whether memory evidence is assessed in an all-or-none (discrete) or graded (continuous) manner. However, it is prudent to describe two specific, prototypical models (2HTM and SDT) in order to ground discussion of discrete and continuous models.

For context, consider a simple recognition memory experiment. In this experiment, a participant studies a series of words one at a time. Some of these words are studied one time, and others are studied multiple times. The words that are studied multiple times are assumed to be encoded better than words that are only studied once. I will hereafter refer to these as "weak" and "strong" words, respectively. Later, during a test phase, words that have been studied before, called *targets,* and words that have not been studied, called *fillers*, are presented to participants one at a time. The participant must decide whether a presented word is "old" or "new" (i.e., "familiar" or "unfamiliar"). Researchers are interested in two measures of performance from these tasks. The first is participants' ability to correctly classify targets as "old" and fillers as "new". This is known as *discriminability*. The second measure is the propensity for participants to respond "old" to *any* stimulus. Response bias is usually assumed to be independent of discriminability (but see Benjamin, Diaz, & Wee, 2009), because it does not reflect memory ability per se, but rather the participant's willingness to respond. The models described next specify the processes used by participants to make recognition memory decisions by positing parameters that govern discriminability and response bias.

First, I will discuss the two-high-threshold model (Figure 1; see Bröder, Kellen, Schütz, & Rohrmeier, 2013). The 2HTM posits that targets are either detected, and therefore endorsed correctly as "old" (with probability D_O), or they are not detected, which is accompanied by complete information loss (with probability $1 - D_O$). Strong targets are more likely to be detected, and consequently, D_O will be greater for strong than for weak targets. If a target is not detected, it is assumed that there is no memory for that target, reflecting the all-or-none nature of discrete models. When targets are not detected, one may still correctly guess that a target is "old" (with probability g), or incorrectly guess that a target is "new" (with probability $1 - g$). The

2HTM also assumes that fillers may be detected as “new” (with probability D_N) or not (with probability $1 – D_N$). When not detected, fillers may be incorrectly guessed to be “old” (probability g) or correctly guessed to be “new” (probability $1 – g$). Discriminability is described by the parameters D_O and D_N (detection of targets and fillers, respectively). Response bias is described by the parameter g.

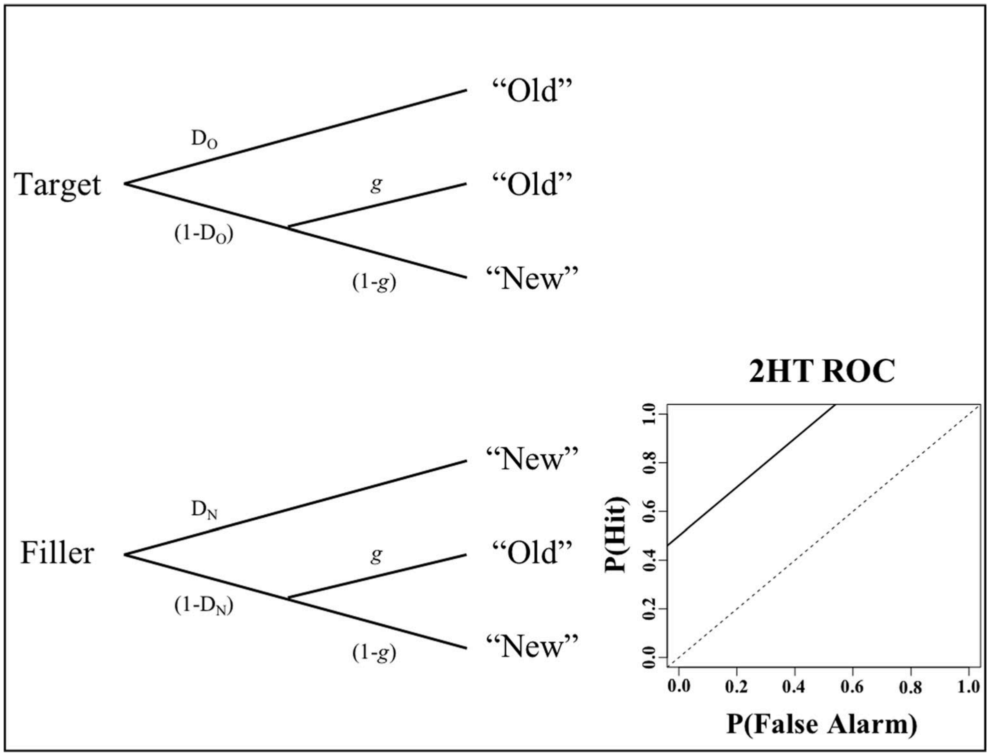


Figure 1. The two-high threshold model (2HTM). Targets are either detected as “old” with probability D_O, or are not detected with probability $(1 – D_O)$ and then guessed to be “old” with probability g, or incorrectly guessed as “new” with probability $(1 – g)$. Fillers are similarly detected as “new” (D_N) or, failing detection ($1 - D_N$), guessed to be “old” (g) or “new” $(1 – g)$. The inset depicts the linear receiver operating characteristic (ROC) curve predicted by the 2HTM (solid line).

Signal detection theory (Figure 2, e.g., Macmillan & Creelman, 2005) posits that memory evidence exits on a continuum. All stimuli, targets and fillers, are assumed to possess at least some latent memory strength. You can plot all these possible strengths along an *x*-axis, as seen in Figure 2. Overlapping Gaussian distributions represent the latent memory strengths of targets and fillers. Distributions with greater average strength (e.g., words studied three times) are shifted further right on the *x*-axis than those with weaker average strength (e.g., words studied one time). Items are evaluated by comparing the latent strength of the item to a response criterion, *C*. If the strength of that item, be it a target or a filler, is greater than *C*, it is endorsed as "old" (and "new" if the memory strength is less than *C*). Discriminability in SDT is governed by the degree of overlap between latent strength distributions of targets and fillers, and can be assessed using *d'*. Figure 2 depicts how *d'* increases as the strength of the memory distribution increases ($d'_{strong} > d'_{weak}$). Response bias is represented as *C*, or the point in which an item is judged to be strong enough to elicit an "old" response.

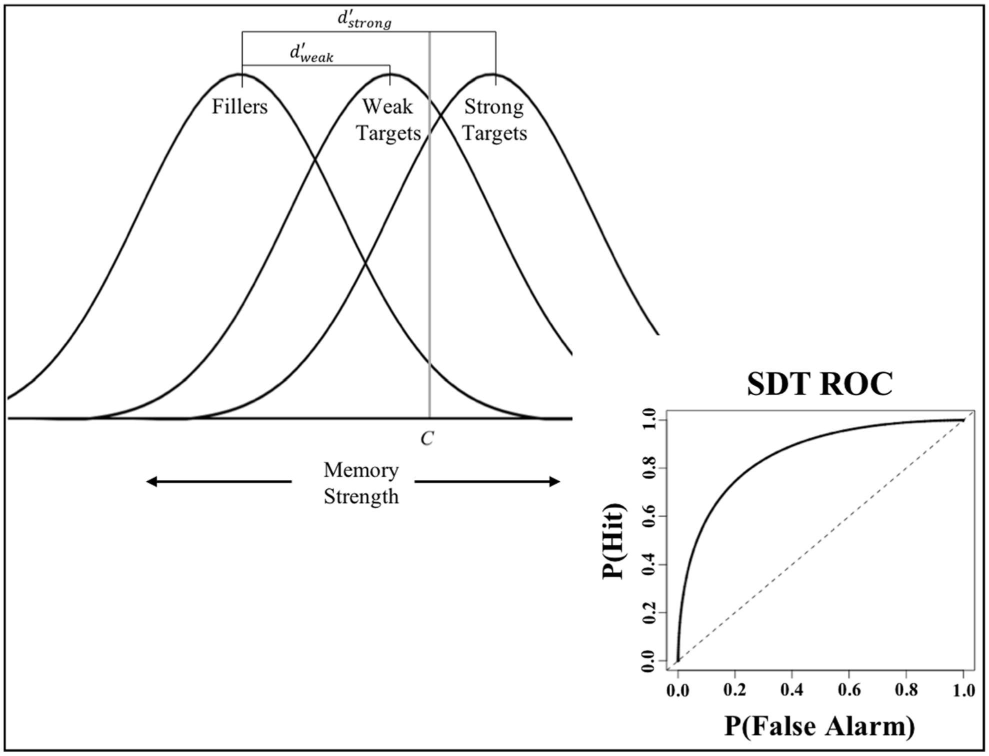


Figure 2. Equal-variance signal detection theory (SDT). Graded memory strength (evidence) is represented on the *x*-axis. Fillers, weak targets, and strong targets have variable memory strength represented as overlapping Gaussian distributions. When an item is presented, its strength is compared to a criterion (*C*). If the strength falls above *C*, it is endorsed as "old", and as "new" if the strength falls below *C*. *d'* is a function of the distance between target and filler distributions, and represents discriminability. The inset depicts the curvilinear ROC predicted by SDT (solid line).

Two additional assumptions, which differ between discrete and continuous models, are worth explicit mention. First, the *certainty assumption* states that any item detected in discrete models will be endorsed as "old" if it is a target or "new" if it is a filler with only high confidence. Continuous models do not share this assumption, and items that have strengths greater than *C* can be endorsed as "old" with a range of confidence, with higher confidence associated with greater strength. Another assumption, *conditional independence*, arises from the all-or-none nature of discrete models. Under this assumption, if targets are encoded with varying levels of strength, strongly and weakly encoded targets will experience the same, complete information loss if not detected. Continuous models again do not share this assumption. If a strong target elicits a strength below *C*, and is endorsed as "new", it will still possess more memory strength, on average, than weak targets that elicit a memory strength below *C*. I turn now to a discussion of the evidence marshalled for and against these two classes of models, and critical tests developed in response to concerns raised about traditional assessments of model performance.

Critical tests of discrete and continuous models

For many years, continuous models were favored over discrete models (Egan, 1958; Krantz, 1969; Luce, 1997; cf., Luce, 1963). This was due primarily to conclusions drawn from receiver operating characteristic (ROC) evidence (see Wixted, 2007, for a review). ROC curves are plots of hits (targets correctly endorsed as "old") and false alarms (fillers incorrectly endorsed as "old") over a range of response biases (*C* and *g* parameters in SDT and the 2HTM, respectively). Response bias is often measured using confidence ratings, with higher confidence associated with more conservative response bias (and vice versa). The 2HTM predicts ROC curves that are linear (see Figure 1 inset) and SDT predicts ROC curves that are curvilinear (see

Figure 2 inset). Empirical ROCs are almost always curvilinear – evidence which seems to favor continuous models (Swets, 1986). However, researchers have made the argument that the shape of ROC curves is not a conclusive test between these classes of models.

Bröder and Schütz (2009) (see also, Krantz, 1969; Malmberg, 2002) argued that the linear ROC predicted by most discrete models are the direct result of the certainty assumption made by many of these models. When the assumption is relaxed, discrete models are able to predict the curvilinear ROCs obtained when using confidence ratings to measure performance and response bias. Bröder and Schütz also demonstrated that by varying response bias directly, through payoff manipulations, the 2HTM fit empirical ROCs better than did SDT. Therefore, the authors concluded that ROC shape is not a conclusive test between these model classes. Other authors (Kellen, Klauer, & Bröder, 2013; Malejka & Bröder, 2019; Province & Rouder, 2012) demonstrated similar results. Given these inconclusive results regarding ROC shape, and the dangers of mono-method biases, other critical tests of discrete and continuous models have been proposed. These include a ranking task (Kellen & Klauer, 2014; McAdoo & Gronlund, 2016; McAdoo & Gronlund, in press; McAdoo, Key, & Gronlund, 2019; Parks & Yonelinas, 2009), the examination of error reaction times (Province & Rouder, 2012; Starns, Dubé, & Frelinger, 2018), and a confidence-rating task (Kellen & Klauer, 2015; McAdoo, Key, & Gronlund, 2018; 2019). I will outline the confidence rating task in detail, as it is used in the present studies.

Confidence rating task

Kellen and Klauer (2015) proposed a critical test between discrete and continuous models, reanalyzing nine previously published studies that had used strength encoding manipulations and tested participants using a six-point confidence scale. The key data the authors examined involved targets endorsed as "new" (misses) from which they computed the conditional probability that these targets received a confidence rating of 1 or 2, given that it was a miss (i.e., rated 1, 2, or 3; $P(1, 2 \mid 1, 2, 3) = \theta$). Kellen and Klauer proved that, because of the aforementioned conditional independence assumption, discrete models predict that $\theta^{weak} = \theta^{strong}$. Because missed targets experience *complete information loss*, strongly and weakly encoded targets are equally likely to receive ratings of 1 or 2 if they are endorsed as “new”. Continuous models, which do *not* adhere to the conditional independence assumption, predict that $\theta^{weak} > \theta^{strong}$. When strongly encoded targets are endorsed as “new”, they are less likely to receive ratings of 1 or 2 because these items will, on average, have more mnemonic evidence than weak targets that are endorsed as “new” (i.e., conditional independence is violated).

Kellen and Klauer (2015) found evidence of discrete mediation using the θ measure. McAdoo et al. (2018; 2019) also found evidence favoring discrete mediation, but only when targets and fillers were semantically dissimilar (e.g., ARROW and THEIF). When fillers and targets were semantically similar (e.g., ARROW and BOW), McAdoo et al. (2018) found evidence for *continuous* mediation. Malmberg and Xu (2007) found that discrete strategies were used more by participants who engaged in rating paradigms (such as those in Kellen and Klauer, and McAdoo et al.) than in yes-no recognition decisions. Finally, Malejka and Bröder (2019) found that individual ROCs could be fit using a model that allowed for discrete (rectangular) distributions of evidence, continuous (normal) distributions of evidence, and mixtures of both, and found that some participants’ data were best explained by discrete mediation, some by continuous mediation, and some by a mixture of the two.

In sum, findings from McAdoo et al. (2018) and others (Malejka & Bröder, 2019; McAdoo et al., 2019; McAdoo & Gronlund, in press), suggest that recognition memory is not fundamentally discrete *or* continuous – accepting one model to the exclusion of the other.

Rather, the situation is more complex (and distinct, though not necessarily divergent, from dual-process theories, which I return to in the General Discussion). This situation was outlined by McAdoo et al. (2019) in a proposed framework that attempts to reconcile the aforementioned results.

A new framework of recognition memory

McAdoo et al. (2019) introduced a general framework of recognition memory describing how recognition evidence can be *treated* in a discrete or continuous manner. This "differential mediation" occurs through the influence of various internal and external factors. Some examples include target-filler similarity (McAdoo et al., 2018), encoding strength (McAdoo & Gronlund, in press), and task demands (McAdoo et al., 2019). The proposed framework is depicted in Figure 3.

In the larger, left plate of Figure 3, a control process (Atkinson & Shiffrin, 1968) governs whether recognition evidence is treated as discrete or continuous. The control process is influenced by variables specific to the participant (top plate), and those specific to the testing environment (right plate). A key factor that impacts the functioning of the control process is efficiency. Malmberg (2008) described efficiency as influencing how mnemonic evidence is treated as a function of what is required to achieve an adequate level of performance in a reasonable length of time (an idea related to satisficing, Krosnick, 1991).

A series of recent studies have contributed to a better understanding of the right plate in Figure 3. McAdoo et al. (2018) demonstrated the impact of stimulus similarity on the proposed control process. When completing a confidence-rating task, participants either viewed targets and fillers that were semantically similar (ARROW and BOW) or semantically dissimilar (ARROW and THIEF). The authors found that when targets and fillers were similar, participants displayed patterns that signaled *continuous* evidence (i.e., $\theta^{weak} > \theta^{strong}$). When targets and fillers were dissimilar, participants displayed patterns that signaled *discrete* evidence (i.e., $\theta^{weak} = \theta^{strong}$). The authors speculated that the results were due in part to efficiency, wherein participants who were receiving the more confusable targets and fillers needed to evaluate continuous evidence in order to achieve a sufficient level of performance, but participants receiving the dissimilar stimuli were able to achieve efficiency while treating the evidence as discrete. The underlying evidence for the stimuli was not assumed to be represented as continuous or discrete, but rather the underlying evidence was *treated as* continuous or discrete depending on target filler similarity.

McAdoo et al. (2019) demonstrated that different *tasks* can induce the treatment of evidence as continuous or discrete and expanded upon the idea first presented in McAdoo et al. (2018). In this study, participants completed both a ranking task (Kellen & Klauer, 2014; McAdoo & Gronlund, 2016) and a confidence rating task. The authors found that participants utilized continuous mediation during the ranking task and the *same* participants utilized discrete mediation during the confidence rating task (dissimilar targets and fillers were used). Note that all encoding and target-filler similarity was held constant between the two conditions, indicating that the task itself was the main factor behind the differential mediation within subjects.

Finally, McAdoo and Gronlund (in press) showed that the strength of targets also impacts how recognition memory is mediated. The authors fit seven studies that employed a ranking task using a continuous model, SDT, and a discrete model, Luce's (1963) low-threshold model (LTM). Both models fit these data equally well quantitatively, but fit different aspects of the data. Specifically, SDT was able to fit data from weakly encoded targets best while the LTM was able to fit data from strongly encoded targets best. The authors speculated that metacognitive judgments of target strength at retrieval led to the use of discrete mediation when targets were strong, and continuous mediation when targets were weak.

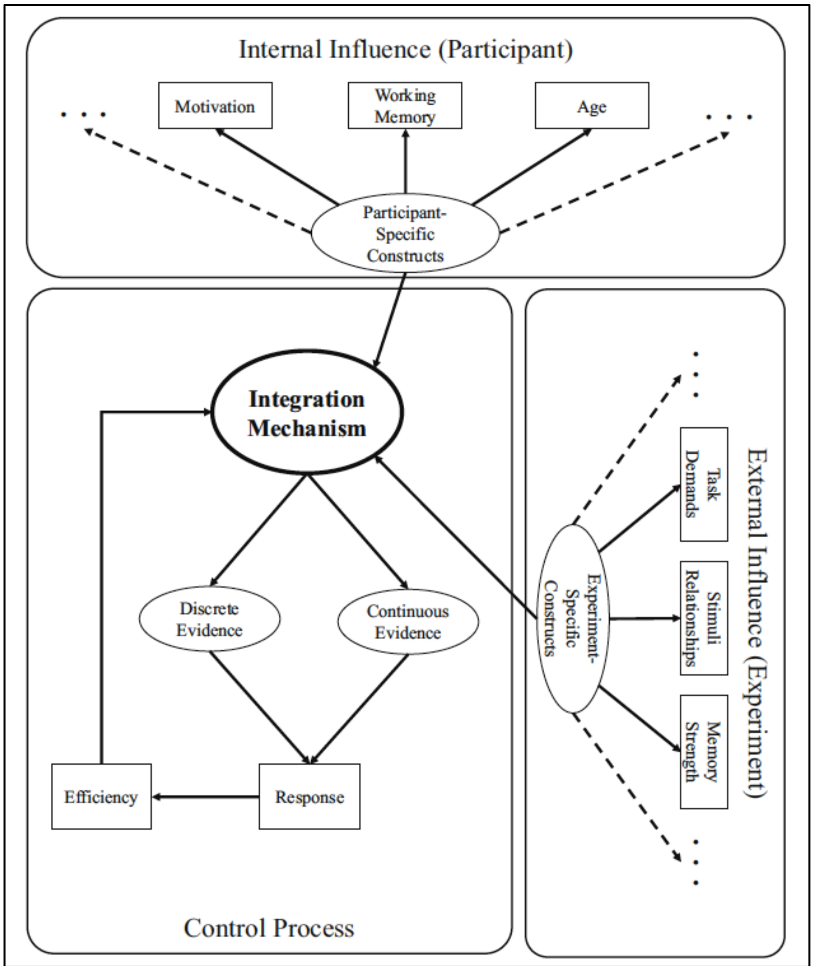


Figure 3. Framework for the mediation of recognition memory (from McAdoo et al., 2019). In the larger, left plate a control process utilizes an integration mechanism to determine whether memory evidence is treated as discrete or continuous, depending on a number of influences. The right plate depicts external, task-specific constructs that influence the control process; the top plate depicts internal, participant-specific constructs that influence it. Reprinted from McAdoo, Key, and Gronlund (2019) with permission from Springer Nature.

Current studies

This brings us to the main questions of interest for the current studies. It has been established that evidence for discrete or continuous mediation can be obtained by changing the structure of the task environment (see right panel of Figure 3). What is not clear is whether recognition memory is fundamentally continuous before a decision is made and, similarly, whether discretization changes that evidence. It seems reasonable to assume that in order for recognition memory to be treated as continuous *or* discrete (strategically), it must be continuously represented at the start of a task (e.g., at the start of a confidence rating task). This is because continuous evidence can be discretized, but discrete evidence cannot be "continualized". This is akin to a dimmer switch that is only set to on or off, to borrow from the earlier metaphor. However, the framework proposed by McAdoo et al. (2019) does not make this assumption explicit, even though it implies this possibility. According to the framework proposed by McAdoo et al., as the participant makes a recognition decision, discretization may occur as a function of efficiency (e.g., if the targets and fillers are dissimilar). Then the question becomes whether this strategic discretization fundamentally changes the underlying evidence from being continuously represented to being discrete. To borrow from the lightbulb metaphor one last time, does discretization change the dimmer switch into a light switch? The focus of the current studies is thus twofold; first to demonstrate that recognition memory, before it is discretized, is fundamentally continuous, and second, to determine whether discretization changes the nature of this evidence to a discrete representation.

It has been shown that successful retrieval can have an effect on memory traces. For example, the long literature supporting a testing effect (see Rowland, 2014 for meta-analysis) suggests that effortful processing of remembered stimuli helps strengthen it upon re-encoding. However, memory traces are not always changed for the better. Hupbach, Gomez, and Nadel (2009) found that by retrieving a memory, the process of reconsolidation and updating can cause source confusion because the memory needs to be re-encoded, and is therefore subject to change as a result of this noisy process. Given what is known about the changes that can occur to memory traces upon successful retrieval, it is reasonable to consider whether discretization of recognition memory affects that evidence, changing it from being continuously represented to being discretely represented. Experiment 2 was designed to investigate this possibility.

In sum, recognition memory may be treated as continuous or discrete, depending on a number of external (McAdoo & Gronlund, 2016; McAdoo, Key & Gronlund, 2018, 2019) and internal factors. The framework depicted in Figure 3 describes how this may occur through the functioning of a control process. Two unanswered questions are investigated in the current studies. First, Experiment 1 is designed to determine whether recognition memory evidence is inherently continuous, but is treated as discrete when discretization strategies are employed. Experiment 2 investigates whether discretization fundamentally changes the underlying mnemonic evidence.

Experiment 1

Experiment 1 was designed to demonstrate that recognition memory evidence is fundamentally continuous. I do so by conducting an experiment for which evidence for discrete mediation has been found in the past, but utilize a measure that might be more sensitive to the underlying evidence than the six-point scale used previously (Kellen and Klauer, 2015; McAdoo et al., 2018; 2019). I hypothesize that when θ is calculated using data from this sliding scale, it will signal continuous mediation, indicating that the evidence available to participants during this task is fundamentally continuous, but is subject to strategic discretization (see McAdoo et al., 2018; 2019).

Method

Participants

University of Oklahoma introductory psychology students (*N* = 69) completed this study in exchange for partial course credit. Five participants were excluded for not following instructions or due to technical difficulties. One additional participant was excluded for response patterns that did not allow for calculation of the dependent variable (i.e., no misses in one or both encoding conditions). This left a final *N* = 63. Participants were mostly self-identified female (*N* = 42) with a median age of 19 years. Participants' self-reported ethnicity broke down as Caucasian (*N* = 36), African American (*N* = 5), Hispanic (*N* = 10), Asian (*N* = 8), Middle Eastern (*N* = 1), Native American (*N* = 2), Pacific Islander (*N* = 1), and No Response (*N* = 2).

Materials

The stimuli were the same as used in McAdoo, Key, and Gronlund (2018; 2019). They were drawn from the Nelson, McEvoy, and Schreiber (1998) University of South Florida Free Association Norms database. The words were English nouns with a forward-strength of .40 (meaning 40% of participants provided the same freely associated word), and a word-frequency count (Kučera & Francis, 1967) of at least three. The list was randomized and the first one hundred items were chosen to be targets. Appendix D of the Nelson et al. database contains idiosyncratic associations for each of the one hundred chosen target words. Idiosyncratic associations are words that only one participant associated with the target word when prompted. As an example, for the target word ARROW, the word THIEF served as the idiosyncratic filler. The final list of words contained one hundred targets and one hundred semantically dissimilar fillers. Fillers were validated such that, to the best of my ability, a filler that was dissimilar to one target was not similar to any other target(s) or filler(s). At the beginning of each session (i.e., for every participant, to reduce potential for specific stimuli effects), the list of one hundred targets was randomized and the first half (50 words) were chosen to be weak targets and the second half (50 words) were chosen to be strong targets. Strong targets were duplicated twice so that the final list contained 100 unique targets, 50 presented one time, and 50 presented three times.

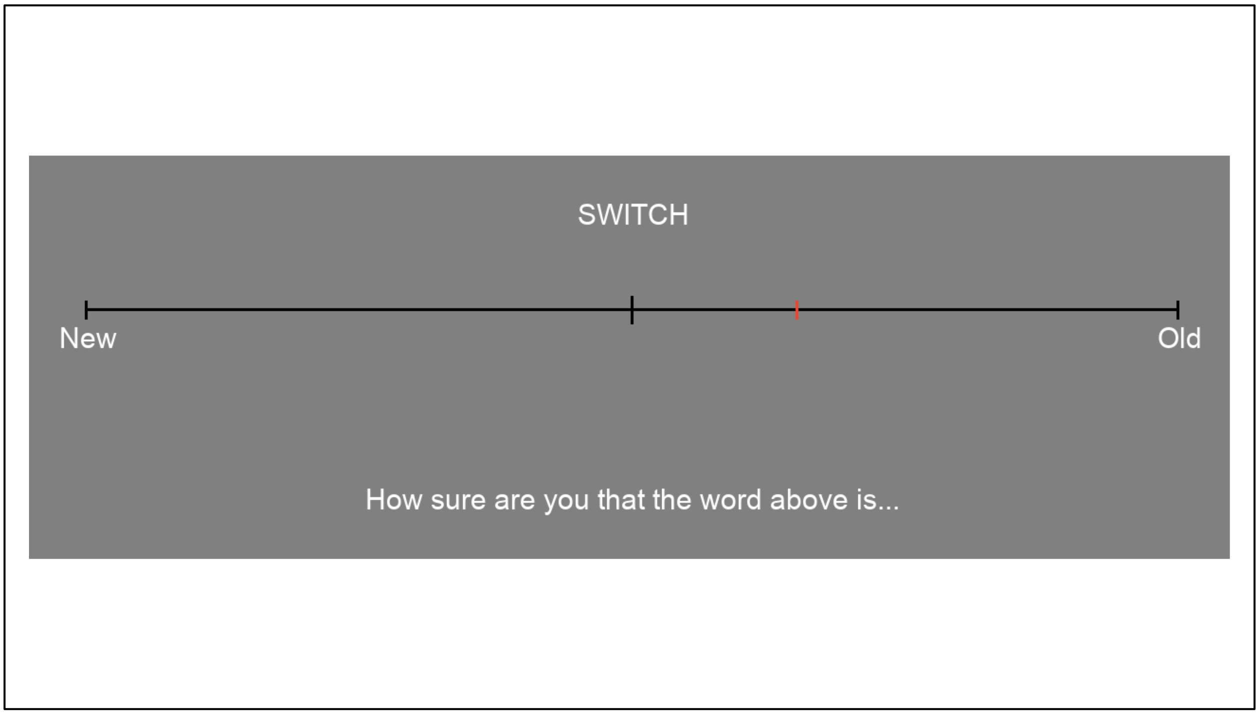
Procedure

The procedure was reviewed and approved by the University of Oklahoma Institutional Review Board and followed American Psychological Association ethical guidelines. The experiment took place in a room with four cubicles, each with a personal computer. Data collection and stimulus presentation were controlled using MATLAB (MathWorks, 2018) using the Psychophysics Toolbox extensions (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007).

First, participants consented and provided their age before beginning the experiment. This was followed by the study phase. Participants were instructed that they would view a series of words, some once and some multiple times, to study the words by any method they chose, and that good performance was achievable even if it felt overwhelming. The study list (50 weak, 50 strong) was randomized before presentation. Each word was then presented for 1,000-ms followed by a 500-ms fixation cross. Following the study phase, participants completed a distractor task consisting of 20 math problems and taking about eight minutes.

Participants were then told they would be viewing a series of words, some of which they had previously studied and some of which they had not. The participants were told to indicate on a slider (depicted in Figure 5) how sure they were that the presented word was “old” (studied) or “new” (not studied). Examples were given describing how placing the slider closer to “old” or “new” indicated that the participant was more certain that the word was “old” or “new”, and less certain the further from the ends the slider was placed. Once a participant indicated he or she understood the instructions, the test phase began. Participants placed their mouse along the slider (which did not contain any markings) and clicked the mouse button once they had placed it where they wished. This continued for a total of 200 trials (50 weak targets, 50 strong targets, and 100 fillers presented in random order).

After the test phase, participants provided self-reported demographic information and were debriefed.



How sure are you that the word above is...

Figure 4. Example of sliding confidence scale used in Experiment 1. The red line indicates a position where a participant may place the slider if he or she was somewhat confident that the word SWITCH was “old”.

Results

Table 1 displays the relevant descriptive and inferential statistics for Experiment 1. Inferential statistics were computed using Bayesian paired samples *t*-tests within the JASP statistical analysis software (JASP Team, 2019). In standard frequentist hypothesis testing, one reports how likely it would be to obtain data given the null is true [P(Data | Null)] (see

Wasserstein & Lazar, 2016 for a discussion on the *p*-value and its misconceptions). Bayesian hypothesis testing reports how likely a hypothesis is given one's data [P(Hypothesis | Data)]. One can express this probability as a likelihood ratio between two competing hypotheses, a Bayes Factor (BF). When BFs are reported here, the competing hypotheses will be specified, and the likelihood of one over the other will be spelled out. I will use the Jeffreys (1961) criteria for interpreting Bayes factors. Responses from the confidence slider were binned in order to calculate θ (see Figure 4). Although participants did not see the numerical results of their choices, the program reported distances from the center of the slider as numbers between -100 and 100 (with negative numbers indicating responses made for "new" and positive for "old").[1]

Table 1. *Descriptive and inferential statistics for Experiment 1.*

		M(SD)			
Variable	***N***	***Weak***	***Strong***	***H_1***	***BF_{10}***
Performance					
Hit rate	64	.57 (0.15)	.79 (0.16)	$Hit^{weak} < Hit^{strong}$	>1000
False alarm rate	64	.25 (0.16)		NA	NA
θ					
Full group	63	.68 (0.28)	.61 (0.30)	$\theta^{weak} < \theta^{strong}$	5.22
Diagnostic and stable	23	.67 (0.29)	.61 (0.24)	$\theta^{weak} < \theta^{strong}$	4.39

Performance

Hits occur when participants rate a target between 0 and 100 on the confidence slider (i.e., place their slider to the right of the middle line). The BF in favor of the hypotheses $Hit^{strong} > Hit^{weak}$ was > 1000, providing decisive evidence in favor of better memory in the strong encoding manipulation than the weak encoding manipulation. That is, it is 1000 times more likely that Hit^{strong} (M = .57 , SD = 0.15) is greater than Hit^{weak} (M = .79, SD = 0.16) than not.

---

[1] All data an stimuli are publicly available on the Open Science Framework at https://osf.io/yzgw2/

Given the successful encoding manipulation, I can confidently use the θ measure to test for discretization in the confidence-rating task.

θ: Six-Point Scale

First, slider scale responses were binned into six equal categories. θ was calculated by dividing the number of targets that were rated -100 – -67 and -66 – -34 and dividing by the total number of targets that were rated "new" (between -100 and 0). This was done for weak and strong targets for every participant. Evidence for continuous evidence would be found if $\theta^{weak} > \theta^{strong}$ and for discrete evidence if $\theta^{weak} = \theta^{strong}$. A BF in favor of the hypothesis $\theta^{weak} = \theta^{strong}$ was 0.38. This indicated that it was 2.63 times more likely that θ^{weak} (M = .68, SD = 0.28) did not equal θ^{strong} (M = .61, SD = 0.30). Another BF was calculated, this time in favor of the more theoretically focused hypothesis $\theta^{weak} > \theta^{strong}$ (the statistic reported in Table 1) and was found to be 5.22, indicating that θ^{weak} is 5.22 times more likely to be greater than θ^{strong} than not. These results suggest that, when tested with the confidence slider rather than a six-point scale, recognition evidence is continuously represented. This further supports the hypothesis that discretization is a strategic treatment of continuous evidence.

The validity of θ is influenced by the number of misses and a successful encoding manipulation. Therefore, participants with less than ten misses for both weak and strong targets, and a difference in weak and strong hit rates less than .10, were excluded from analysis to create a group of "diagnostic and stable" participants (see Kellen & Klauer, 2015; McAdoo et al. 2018; 2019; for similar exclusion). A BF for stable participants (*N* = 23) was 4.93 in favor of the hypothesis $\theta^{weak} > \theta^{strong}$. This indicated that it was 4.93 times more likely that θ^{weak} (M = .67, SD = 0.29) was greater than θ^{strong} (M = .609, SD = 0.243) than not, evidence in favor of continuous mediation.

Discussion

Experiment 1 revealed underlying continuous representation of recognition evidence in a task previously shown to encourage discretization. This was accomplished by utilizing a confidence slider that was more sensitive to this evidence than the six-point scale used in prior research. Responses were binned in order to calculate θ, and evidence of continuous mediation was found ($\theta^{weak} > \theta^{strong}$), supporting the hypothesis that underlying recognition memory evidence is likely inherently continuous, and is discretized when efficient.

Recall that McAdoo et al. (2018; 2019) found evidence for discrete mediation using the exact same task as Experiment 1, with only the scale with which responses were made differing. This has both theoretical and practical implications. I will discuss the practical implications in more detail in the General Discussion. However, one theoretical implication directly informs Experiment 2. With evidence that memory is fundamentally continuous, one may now ask the question – does discretization change this? Because the confidence-rating scale induces discretization, does that decision modify the nature of the mnemonic evidence, changing it from being continuous to being discrete? Experiment 2 tests this possibility, utilizing the six-point confidence scale that has been shown to induce strategic discretization, but adding a second test to determine if that decision changes the recognition evidence.

Experiment 2

Consider the following scenario. A participant studies the word *TABLE* one time and the word **FROG** three times. In this case, *TABLE* is weakly encoded and **FROG** is strongly encoded. Consider the possibilities if recognition memory for these items is inherently continuously represented (see Figure 5 for an outline of this idea). At test, the participant is shown the word *TABLE* on one trial and **FROG** on another. The participant must classify these words as being “old” or “new” using a six-point confidence scale ranging from “1 – sure new” to “6 – sure old”. In both test trials (one with *TABLE,* the other with **FROG**, embedded among many other test trials), the participant evaluates the evidence for the word, and in both cases this evidence falls below some response threshold and both test words are endorsed as “new”. Because of the confluence of internal and external factors described in Figure 3 (e.g., targets and fillers are dissimilar), prior research indicates that participants *treat* this evidence as discrete. As such, both *TABLE* and **FROG** sustained complete information loss, and the assumption of conditional independence holds. What is unknown at this point is whether this discretization strategy changes the underlying mnemonic evidence from continuous to discrete. That is, if memory is continuously represented, does treating it as discrete change the evidence such that it is updated to be discretely represented. A second test of the missed items (i.e., a second test of *TABLE* and **FROG** in the above scenario) can reveal whether continuous evidence is maintained following discretization. This is the goal of Experiment 2.

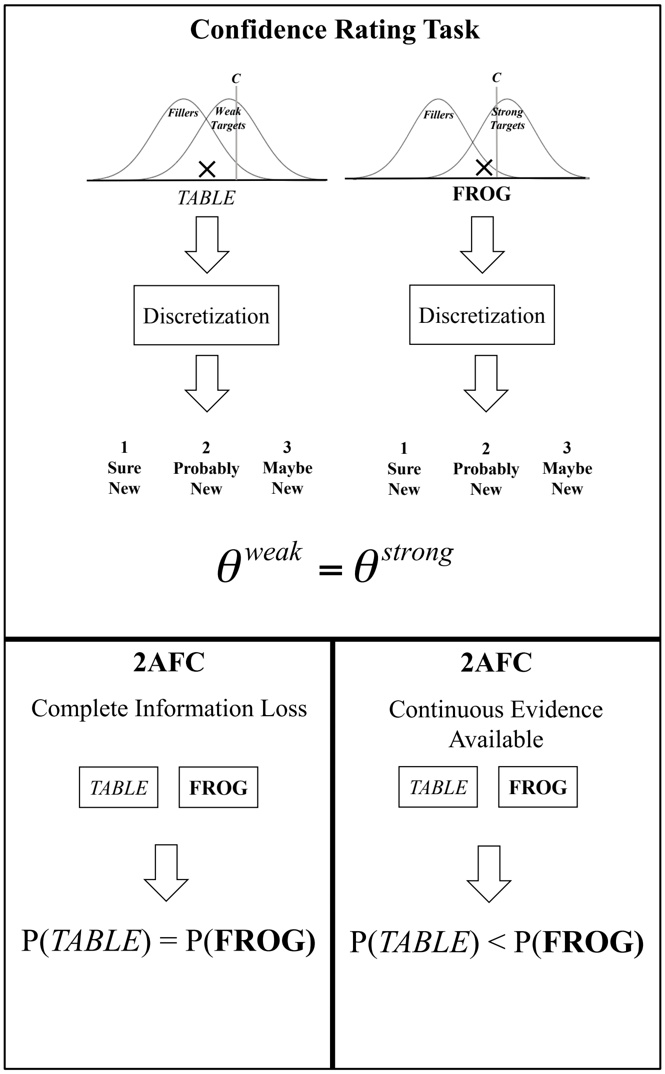


Figure 5. Possible patterns of 2AFC responses following a confidence-rating task wherein missed targets are discretized.

Method

Figure 6 outlines the general procedure for Experiment 2. In the second test, strong misses and weak misses will be paired in a 2AFC task. If discretization in the confidence rating task changes memory evidence from continuous to discrete, weak and strong misses will be chosen equally often in the 2AFC task. This is due to complete information loss when a target is missed (conditional independence holds). If, on the other hand, underlying evidence remains continuous despite θ signaling discretization, strong misses will be more likely to be chosen on average than weak misses, as they retain more evidence on average (conditional independence does not hold). Therefore, I predict in the 2AFC task: $H_{discrete}$: $P(Miss)^{strong} = P(Miss)^{weak}$ and $H_{continuous}$: $P(Miss)^{strong} > P(Miss)^{weak}$.

Participants

University of Oklahoma introductory psychology students (*N* = 54) completed this study in exchange for partial course credit. Eight participants were excluded for not following instructions or due to technical difficulties. An additional three participants were excluded for response patterns that did not allow for calculation of the dependent variable (i.e., no misses in one or both encoding conditions). This left a final *N* = 42. Participants were mostly female (*N* = 30) with a median age of 19 years. Participants' self-reported ethnicity broke down as Caucasian (*N = 33*), African American (*N* = 3), Hispanic (*N* = 4), Asian (*N* = 4), and Middle Eastern (*N* = 1).

Materials

The words used in the confidence rating task were the same used in Experiment 1. An additional 100 semantically dissimilar words were chosen from the Nelson et al. (1998) database as fillers for the second, 2AFC test. The final list of words contained 100 targets and 200 semantically dissimilar fillers. Fillers were again validated such that, to the best of my ability, a filler that was dissimilar to one target was not similar to any other target(s) or filler(s). At the beginning of each session (i.e., for every participant, to reduce potential for specific stimuli effects), the list of 100 targets was randomized and the first half (50 words) were chosen to be weak targets and the second half (50 words) were chosen to be strong targets.

Procedure

The procedure was reviewed and approved by the University of Oklahoma Institutional Review Board and followed American Psychological Association ethical guidelines. Figure 6 depicts the procedure for Experiment 2. The entire experiment took place in a room with four cubicles, each with a personal computer. Data collection and stimulus presentation were controlled using MATLAB.

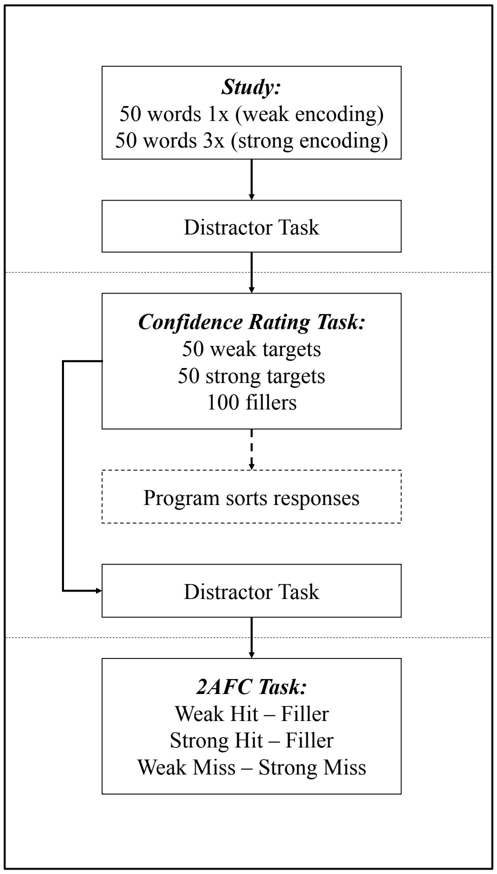


Figure 6. Procedure outline for Experiment 2. Participants study a list of words, followed by a distractor task. They then completed a confidence-rating task, followed by another distractor task, and finished with the 2AFC task.

The study phase for Experiment 2 was identical to Experiment 1. Following the distractor, participants completed the first test phase. One hundred targets (50 weak, 50 strong) and 100 fillers were presented to participants one at a time in random order (for a total number of

200 test trials). For each test trial, participants were instructed to rate how confident they were that the presented word was "new" or "old" using a six-point confidence scale using labeled keys on the computer keyboard. The points on the scale were: 1 – Sure New; 2 – Probably New; 3 – Maybe New; 4 – Maybe Old; 5 – Probably Old; 6 – Sure Old. As the participants completed this phase, the program kept track of responses, sorting targets into four categories: weak hits; strong hits; weak misses; and strong misses.

Following the confidence-rating task, participants completed another distractor task consisting of twenty math problems. They were then instructed that they would be shown two words side-by-side and that they needed to select the word they believed was most likely to have been seen in the first study phase. Three types of word pairs were presented amongst a total of fifty trials. The first type of word pair contained weak targets that were missed during the first test phase and strong targets that were missed during the first test phase: [weak miss – strong miss]. The number of these trials varied depending on how many misses the participant made and ranged from zero to 37 with a median of nine. The remaining trials were split in half with one half containing pairs of weak hits and new, idiosyncratic fillers, and the other with strong hits and new, idiosyncratic fillers: [weak/strong hit – filler]. The order of the tests was randomized. Participants indicated their choices by using keys labeled "left" and "right" on the computer keyboard. After completing all 50 2AFC trials, participants provided self-reported gender and race and were debriefed and dismissed.

Results

Table 2 displays the relevant descriptive and inferential statistics for Experiment 2. As in Experiment 1, inferential statistics were computed using Bayesian paired samples *t*-tests within the JASP statistical analysis software (JASP Team, 2019). I begin with performance in the confidence rating task. Then, I will describe θ results from the confidence-rating task and the [weak miss – strong miss] trials from the 2AFC task.

Table 2. *Descriptive and inferential statistics for performance,* θ, *and 2AFC*

Variable	***n***	***M(SD)*** ***Weak***	***M(SD)*** ***Strong***	***H_1***	***BF_{10}***
Performance					
Hit rate	45	.58 (0.17)	.78 (0.16)	$Hit^{weak} < Hit^{strong}$	>1000
False alarm rate		.28 (0.20)		NA	NA
θ					
Full group	42	.74 (0.22)	.68 (0.31)	$\theta^{weak} = \theta^{strong}$	1.02
Diagnostic and stable	21	.77 (0.22)	.75 (0.24)	$\theta^{weak} = \theta^{strong}$	3.61
P(Miss)					
Full group	42	.21 (0.16)	.79 (0.16)	$P(Miss)^{weak} < P(Miss)^{strong}$	>1000
Diagnostic and stable	21	.25 (0.12)	.75 (0.12)	$P(Miss)^{weak} < P(Miss)^{strong}$	>1000

Performance

Hits occur in the confidence rating task when targets are rated “4 – Maybe Old”, “5 – Probably Old”, “6 – Sure Old”. The BF in favor of the hypotheses $Hit^{strong} > Hit^{weak}$ was > 1000, providing decisive evidence in favor of better memory in the strong encoding manipulation than the weak encoding manipulation. That is, it is 1000 times more likely that Hit^{strong} (M = .78, SD = 0.16) is greater than Hit^{weak} (M = .58, SD = 0.17) than not. Given the successful encoding manipulation, I can confidently use the θ measure to test for discretization in the confidence rating task.

Confidence rating task

The θ measure is calculated by dividing the number of weak and strong targets rated 1 or 2 by the number of targets rated 1, 2, or 3 for each participant. This provided θ^{strong} and θ^{weak} estimates for every participant. Evidence for discretization is found when $\theta^{weak} = \theta^{strong}$ and for continuous evidence when $\theta^{weak} > \theta^{strong}$. A BF in favor of the hypothesis $\theta^{weak} = \theta^{strong}$ was 1.017.

This hypothesis was used as I was expecting to find that $\theta^{weak} = \theta^{strong}$ given prior research (McAdoo et al., 2018; 2019). That is, θ^{weak} (M = .74, SD = 0.22) and θ^{strong} (M = .68, SD = .31) are just as likely to be the same than different. A group of "diagnostic and stable" participants (those with at least a .10 difference in weak and strong hit rates and ten or more misses) was selected to increase the validity of the θ measure. Note, stable participants also provide at least ten [weak miss – strong miss] 2AFC trials. For stable participants (*n* = 21), a BF in favor of the hypothesis $\theta^{weak} = \theta^{strong}$ was 3.45, indicating strong evidence in favor of discretization. That is, θ^{weak} (M = .79, SD = 0.22) is 3.45 times more likely to be equal to θ^{strong} (M = .77, SD = 0.22) than not.

P(Miss)

The probability of choosing a weak missed target [$P(Miss)^{weak}$] versus a strong missed target [$P(Miss)^{strong}$] in 2AFC trials was calculated for each participant. For the full group, a BF in favor of the hypothesis $P(Miss)^{strong} > P(Miss)^{weak}$ was >1000, indicating decisive evidence in favor of the hypothesis. That is, $P(Miss)^{strong}$ (M = .79, SD = 0.16) is greater than 1000 times more likely to be greater than $P(Miss)^{weak}$ (M = .21, SD = 0.16) than not. The validity of the P(Miss) measure is enhanced by having a sufficient number of misses. A BF in favor of the hypothesis $P(Miss)^{strong} > P(Miss)^{weak}$ was >1000, indicating decisive evidence in favor of the hypothesis. That is, $P(Miss)^{strong}$ (M = .75, SD = 0.12) is more than 1000 times more likely to be greater than $P(Miss)^{weak}$ (M = .25, SD = 0.12) than not.

Effect of priors on Bayes factors

It is prudent to explore the use of different priors in calculation of Bayes factors given the noisiness of the confidence-rating task and to verify the robustness of conclusions drawn from these data. To address this concern, Figure 7 displays the robustness of the Bayes factors for the diagnostic and stable group analysis under different prior (Cauchy) distributions for Experiment 1 (top panel) and Experiment 2 (bottom panel). These figures are produced by JASP and allow one to determine how strong the evidence is in favor of the alternative/null hypotheses under different priors.

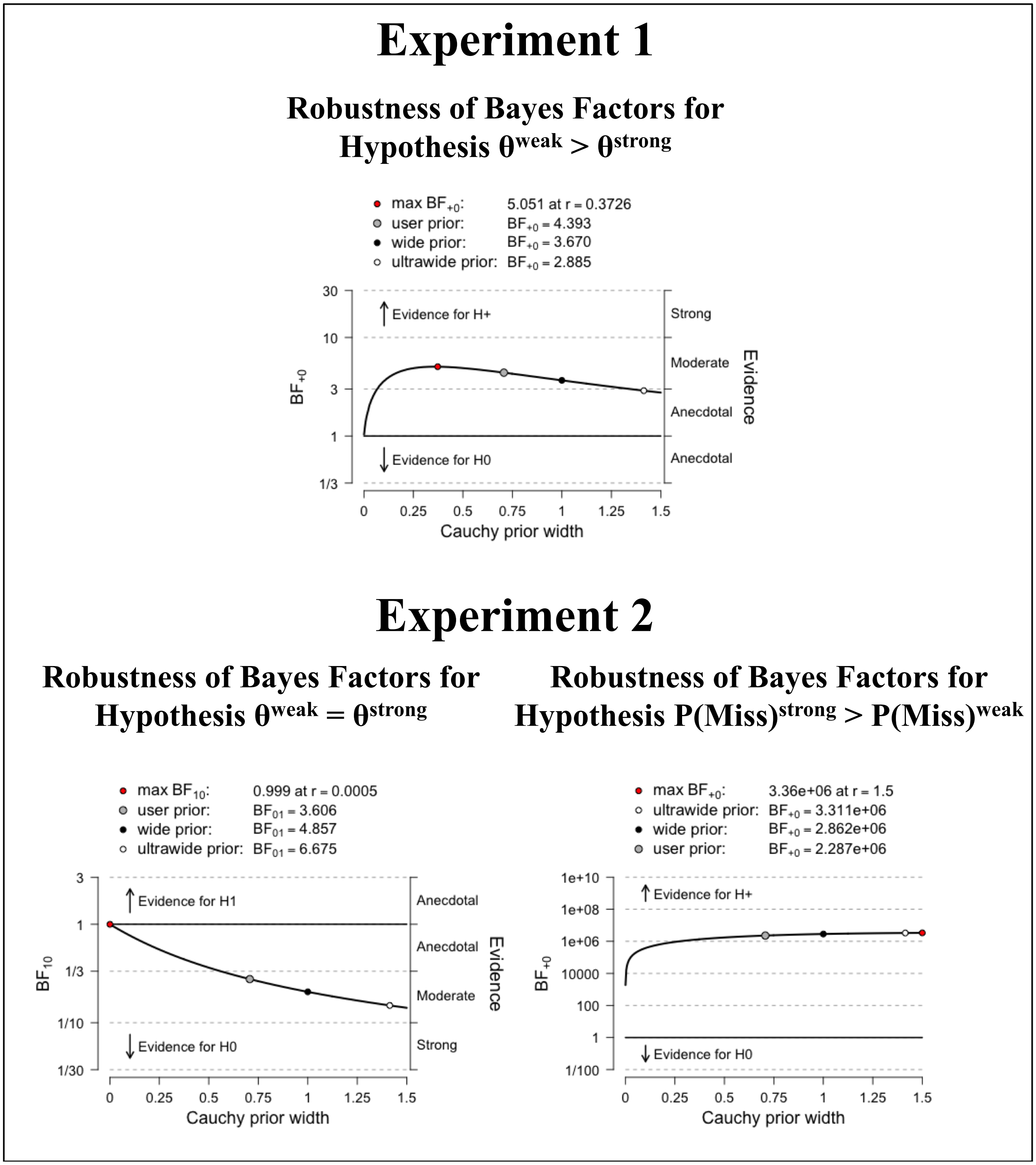


Figure 7. Robustness of Bayes factors for different Cauchy distribution priors from Exp. 1 (top panel) and Exp. 2 (bottom panel). Results reported in Table 1 (Exp. 1) and 2 (Exp. 2) are labeled with grey dots, with wider priors indicated by black and white dots.

Discussion

Experiment 2 was designed to determine whether continuous recognition memory evidence, is changed when discretization occurs at the time of test. Participants completed a confidence rating task designed as a critical measure of discrete or continuous evidence, followed by a 2AFC task in which weak and strong misses were paired. It was hypothesized that if discretization changes evidence from continuously to discretely represented, $P(\text{Miss})^{\text{strong}} = P(\text{Miss})^{\text{weak}}$. If, on the other hand, continuous evidence is retained even after discretization, $P(\text{Miss})^{\text{strong}} > P(\text{Miss})^{\text{weak}}$.

Evidence for discretization was found for the confidence rating task, when dissimilar targets and fillers were used, replicating results from McAdoo et al. (2018; 2019). More importantly, it was shown that continuous evidence was maintained even after the evidence had been discretized. This was accomplished by examining choices made in a 2AFC task that paired weak and strong misses. There was decisive evidence indicating that $P(\text{Miss})^{\text{strong}} > P(\text{Miss})^{\text{weak}}$. This the hypothesis that discretization does not change evidence that is initially represented continuously. The implications of this finding are discussed in more detail in the General Discussion.

General Discussion

The current studies were designed to develop further a new framework proposed by McAdoo et al. (2019) that posits that recognition memory can be mediated by either discrete or continuous evidence. Experiment 1 tested the previously implicit assumption that recognition evidence is *fundamentally* continuous but may be strategically discretized. Using a confidence-rating task that had previously been shown to induce discretization (McAdoo et al. 2018; 2019)—but replacing the six-point confidence scale with a slider—Experiment 1 demonstrated that recognition memory is likely to be fundamentally continuous. Returning to the six-point confidence scale known to induce discretization, Experiment 2 showed that discretizing does not induce a modification of the evidence: A 2AFC task pairing weak and strong targets judged as “new” (missed) showed that strong misses were strongly favored. The results of the current studies have both theoretical and practical implications, discussed next.

Implications for theory

The debate over whether memory is mediated by continuous or discrete processes has been a long and seemingly inconclusive one, with evidence in favor of both discrete and continuous models, as well as other evidence suggesting dual-process approaches (see Pazzaglia, Dubé, and Rotello, 2013, for a critical review, with the authors concluding that continuous models are the most useful). Given the shortcomings associated with ROC analysis, and the dangers of mono-method bias, new critical tests of the core assumptions (all-or-none vs. graded information) are needed (Kellen & Klauer, 2014; 2015; Parks & Yonelinas, 2009; Province & Rouder, 2012; Starns, Dubé, & Frelinger, 2018). But conclusions from these new critical tests also have been mixed. The McAdoo et al. (2019) framework proposes one possible reconciliation. The authors suggested that a control process (Atkinson & Shiffrin, 1968), under the control of the participant, allows for the use of either discrete or continuous mediation as a function of efficiency (Malmberg, 2008). What do prior conflicting results look like when viewed from this framework, and in conjunction with the results of the current studies?

First, it is important to understand at what level of analysis the framework proposed by McAdoo et al. (2019) is positioned. Models like SDT, the 2HT, and dual-process models such as Yonelinas’s (1994) model attempt to capture the underlying processes that give rise to the evidence that leads to recognition memory decisions. McAdoo et al.’s framework can be viewed as a small, but significant, step above that level. That is, this framework is relatively silent on how memory evidence is produced (neurologically, for example). Rather, it is designed to better understand and predict what happens to a memory signal once it has been received by the participant – whether it is treated as continuous, or discretized. Experiment 1 demonstrated that this signal, when it becomes available to the participant, is fundamentally continuous, but may be discretized once responding occurs. With this in mind, it is possible to examine previously proposed theories and models within the context of McAdoo et al.'s framework.

Dual-process models, or models of recognition memory that assume two processes contribute to decision making, provide a good starting point for this discussion. Atkinson and Juola's (1974) dual process model assumed that memory is continuous, but if an item has sufficient strength (either sufficiently high or low), participants can respond to the item immediately, without an extensive search of memory. This could result in discrete-like patterns of response in, for example, a confidence-rating task. Conversely, if an item's strength falls between the high and low criteria, participants will engage in an extensive, and slower, search of memory. Like Experiment 1, I assume, as Atkinson and Juola did, that memory evidence is fundamentally continuous, but may be subjected to discrete-like patterns of response. That is, when memory is strong enough, immediate, discrete responses can be made, which is what McAdoo and Gronlund (in press) speculated accounted for better fits of a discrete model to strong memory data. In fact, Atkinson and Juola also speculated that a control process influenced these decisions, just as McAdoo et al. did.

The results of the current studies suggest a possible reconciliation of results supporting conflicting dual-process models like Yonelinas (1994) and Wixted and Mickes (2010). Yonelinas posits that items may either be *recollected* (a discrete process, with probability R) or, failing this, a decision may rely on a continuous, familiarity process. Results favoring this model come from studies involving Remember/Know judgments (Gardiner & Richardson-Klavehn, 2000), with recollective processes associated with more "remember" judgments. Conversely, Wixted and Mickes proposed that memory is fundamentally continuous, and recollection, and subsequent "remember" judgments, are the result of items that were encoded very strongly. Therefore, unlike Yonelinas, Wixted and Mickes propose a completely continuous model of recognition memory. A key consideration of Wixted and Mickes' continuous dual-process model is that it reduces to a simple, unequal-variance SDT model. Reconciliation between Yonelinas's model, which assumes a separate discrete process, and Wixted and Mickes' model may be found within McAdoo et al.'s framework and the results of the current studies. First, recognition memory is fundamentally continuous according to Experiment 1, consistent with Wixted and Mickes. But memory can be *treated* a discrete (if it is successfully recollected or if it is very strong), which would give rise to patterns of data consistent with predictions of discrete processes, and consistent with data supporting Yonelinas's model (see Yonelinas, 2002 for an extensive review). Kapucu, Macmillan, and Rotello (2010) found support for this idea when they reported that participants engaging in a plurals paradigm (where plurals of studied words served as highly-confusable fillers), sometimes engaged in discrete response strategies in a remember-know judgment task in accordance to Yonelinas's dual-process model. Note that these authors also conclude that the signal received from the participants was fundamentally continuous, but strategically discreteized by some. In sum, viewing recognition memory as a more dynamic process, wherein fundamentally continuous evidence may be discretized when it suits the demands of the task/environment, offers a beginning at reconciling these two viewpoints.

The results of the current studies provide new ways of approaching the discrete versus continuous debate in the context of single-process models (Batchelder & Alexander, 2013; Bröder & Schütz, 2009; Dubé & Rotello, 2012; Pazzaglia, Dubé, & Rotello, 2013) as well. According to Experiment 1, formal models of recognition memory should assume fundamentally continuous evidence but allow this evidence to be discretized. Furthermore, if and when memory is discretized appears to be a complex interplay of many factors. Task that utilize random or dissimilar targets and fillers, for example, are likely to be discretized when numerical (e.g., Likert) confidence scales are used as a measure, but not when a graded confidence slider is used. Nor is this the whole story. Province and Rouder (2012) used a confidence slider on a 2AFC test and found evidence of discrete processes. However, Province and Rouder labeled their confidence sliders with “pretty sure”, “believe” and “guess” with anchors of 100% confidence at each end and rewarded participants with points, with more extreme confidence ratings associated with riskier point gains/losses. That the labeling of the slider in Province and Rouder likely made their scale more similar to the six-point scale used in previous studies that found discrete mediation may explain the differences in the current results, which found evidence of continuous mediation. Similarly, Parks and Yonelinas (2009) found evidence of discrete mediation when targets were studied by forming arbitrary associations among them. This seems to be contradictory to the findings of McAdoo et al. (2018), who found that similarity of items induced *continuous* mediation. However, Kellen and Klauer (2011) showed that the data from Parks and Yonelinas existed within a space accounted for by both the 2HT model and SDT (see their Figure 2) and concluded that it was not a diagnostic test for discrete or continuous evidence. In sum, whether, when, and how recognition memory is *treated* as discrete is still under investigation.

Experiment 2 offers some intriguing takeaways from a phenomenon not often tested in recognition memory – second tests of previously missed items. In one recent exception, Starns, Dubé, and Frelinger (2018) examined 2AFC reaction times for second tests. The authors paired studied words that had been missed with non-studied words that had been correctly rejected, studied words that had been correctly endorsed as "old," and non-studied words that had been incorrectly endorsed as "old". The authors concluded that the discrete 2HT model could not account for their results, but that the LTM and continuously mediated diffusion model (Ratcliff, Smith, Brown, & McKoon, 2016 for a review) could. I extended this finding by pairing strong and weak target misses following discretization of single-item test responses and found that recognition memory remains continuous despite strategic discretization. I view this as more evidence that when discretization occurs, it is not due to the lack of available continuous evidence, but the result of a desire to be strategically efficient. This conclusion has significant practical implications, which I explore next.

Implications for practice

Confidence has been shown to be a strong predictor of eyewitness identification accuracy when tested fairly (Wixted & Wells, 2017; for a recent review see Gronlund & Benjamin, 2018). Using calibration analyses, Juslin, Olsson, and Wilson (1996) were among the first to show that when eyewitnesses were highly confident, they were also highly accurate. The relationship between confidence and accuracy is not surprising under assumptions of continuous mediation. According to SDT, for example, higher values of evidence are directly related to higher levels of confidence. However, the aforementioned strong confidence-accuracy relationship only applies to those who *choose* from lineups. When examining non-choosers (people who reject lineups), the confidence-accuracy relationship is weak. That is, rejecting a lineup with high confidence does not indicate that the lineup *definitely* did not contain the perpetrator. If eyewitnesses rely on discrete mediation, it is not surprising that a strong confidence-accuracy relationships is not found due to the conditional independence assumption. When an item is detected, whether it be strong or weak, the probability of assigning *n*-confidence to that item is the same regardless of the item's strength. Only the probability of *entering* the detect state differs for strong and weak memory, what occurs once one is in that state is independent of strength. Could, then, the lack of a strong confidence-accuracy relationship for non-choosers be the result of discretization, and the presence of conditional independence?

The findings from Experiment 1 suggest that recognition memory is fundamentally continuous, but subject to strategic discretization. But the results also suggest that this evidence might be evaluated more precisely if the participant could be made not to discretize it (e.g., when an appropriately sensitive scale is used). Most eyewitness identification studies utilize confidence scales similar to the six-point scale used for calculation of θ. When confidence-accuracy relationships disappear when eyewitnesses reject lineups, this may signal that they are discretizing their decisions. In other words, when a lineup is rejected, the confidence in that decision is not a direct index of memory strength (and subsequent accuracy), because discrete mediation posits the assumption of conditional independence. However, given the evidence that recognition memory evidence is fundamentally continuous, it should be possible to encourage eyewitnesses to probe this evidence, even when rejecting a lineup, and render a rejection decision that is more informative for the police (i.e., more indicative of whether the suspect is innocent or not).

Why would one be interested in eyewitnesses who reject lineups? It is of clear forensic utility that suspects of crimes, who actually committed the crime, be accurately selected from lineups. Therefore, scientists are interested in whether eyewitnesses who choose from a lineup do so with high accuracy, and these "choosers" have formed the main group of interest for those who study confidence-accuracy relationships. But the accuracy of non-choosers is also important. Imagine an investigation where police are fairly certain they have the perpetrator of a crime. They include this suspect in a lineup and ask an eyewitness to the crime to make a decision. The eyewitness then rejects the lineup. One of two things could happen. First, investigators may believe the eyewitness, release the suspect, and pursue other lines of evidence. Or the investigators may consider the eyewitness to be unreliable (e.g., have a poor memory) and continue to investigate the suspect they have. Given the poor confidence-accuracy relationship for non-choosers, the latter option may be considered the better one. However, if eyewitnesses who reject lineups are discretizing, procedures that encourage a re-evaluation of the evidence may strengthen the confidence-accuracy relationship for non-choosers and bolster the utility of reject decision. The results of Experiment 2 also suggest that even following discretization, continuous evidence of one's decisions still exists, and potentially can be exploited. This is an open empirical question and a topic for future research.

As another example, consider the scene of an accident or a natural disaster. Super (1984) developed the Simple Triage and Rapid Treatment (START) procedure to guide first responders toward the effective and efficient treatment of injured parties. START is a fast-and-frugal model – a class of decision-making models that resemble discrete models of recognition memory (Gigerenzer & Todd, 1999). First responders may discretize graded evidence of symptoms when engaging in the START paradigm, trying to determine who to treat immediately versus who has to wait. They would take what is fundamentally a continuum of symptoms and pertinent variables (e.g., age, consciousness, mental coherence) and discretize the evidence in order to quickly and efficiently “sort” victims. Once the most serious cases are taken care of, the first responders may then turn their attention to the victims who were deemed “less serious”. However, it is likely these other victims have a graded continuum of symptoms. Because discretization of evidence does not change the underlying nature of that evidence, it should still be available for evaluation, and a new batch of “most serious” victims could be determined without having to go through the START procedure, or other discrete decision making, again.

Limitations

One limitation of the current studies, and previous studies using the confidence rating task, is the existence of large between-participant variation in responses. On average, Experiment 1 signaled continuous evidence, and Experiment 2 signaled discrete evidence. But this does not mean that every participant relied on continuous evidence in Experiment 1, or discrete evidence in Experiment 2. Moreover, these measures do not indicate that every *trial* is continuously or discretely mediated, only that, across all trials, mediation was, on average, one or the other. Therefore, it would be prudent to study trial-by-trial variations in how discretization occurs. It is possible that all participants begin by evaluating continuous evidence, and then discretize as the demands of the task allow. However, the opposite is also possible; that participants begin by discretizing, and then only consider continuous evidence once discretization is found to be inefficient. Given that many recognition decisions, including eyewitness identifications, are only made once, a better understanding of these across-trial and between-participants’ variations is crucial.

A related issue concerns the measurement of $P(\text{Miss})^{\text{weak}}$ and $P(\text{Miss})^{\text{strong}}$ from Experiment 2. As with θ, the measurement of these variables is done on average, across trials, and across participants. So, although continuous evidence appears to remain intact, this is likely not true for every discretized missed target. It is possible that some targets, when discretized, are changed and the evidence is converted to be all-or-none. Trial-by-trial mediation could possibly be assessed by investigation response latency times. Perhaps fast responses are more indicative of discrete mediation than slower. A principled way of measuring these latencies would need to be developed. It may also be possible to use signals from EEG or a related measurement technique to determine whether some trials exhibit discrete or continuous patterns of ERPs.

Conclusion

Although it appears that recognition memory is not always mediated by continuous or discrete evidence, it seems likely that, fundamentally, recognition memory evidence is continuous. But continuous evidence can be strategically *treated as discrete*. Furthermore, this discretization does not change the underlying evidence to become discretely represented (at least not on average). This means that the continuous evidence of discretized items remains intact, and available to the participant. The implications of the framework proposed by McAdoo et al. (2019), and the results of the current studies, advance theory in recognition memory, and have potential real-world consequences. Going forward, new research, grounded in a more complete understanding of how recognition memory decisions (and other, similar decisions) are made, can help push our understanding of these decisions in exciting new directions.

References

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93,* 154 – 179.

Atkinson, R. C., & Juola, J. F. (1974). *Search and decision processes in recognition memory.* WH Freeman.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human Memory: A proposed system and its control processes. In *Psychology of learning and motivation* (Vol. 2, pp. 89 – 195). Academic Press.

Batchelder, W. H., & Alexander, G. E. (2013). Discrete-state models: Comment on Pazzaglia, Dubé, and Rotello (2013). *Psychological bulletin, 139,* 1204 – 1212.

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

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433 – 436.

Bröder, A., Kellen, D., Schütz, J., & Rohrmeier, C. (2013). Validating a two-high-threshold measurement model for confidence rating data in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 587 – 606.

Bröder, A, & Schütz, J. (2009). Recognition ROCs are curvilinear – or are they? On premature arguments against the two-high-threshold model of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41,* 1215 -1222.

Dubé, C., & Rotello, C. M. (2012). Binary ROCs in perception and recognition memory are curved. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 130 – 151.

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

Gardiner, J. M., & Richardson-Klavehn, A. (2000). Remembering and knowing. In E. Tulving & F. I M. Craik (Eds.). *The Oxford handbook of memory* (pp. 229 – 244). Oxford University Press.

Gigerenzer, G., & Todd, P. M. (1999). Fast and frugal heuristics: The adaptive toolbox. In G. Gigerenzer, P. M. Todd, & the ABC Research Group (Eds.). *Simple heuristics that make us smart* (pp. 3 – 34). Oxford University Press.

Gronlund, S. D., & Benjamin, A. S. (2018). The new science of eyewitness memory. K. Federmeier (Ed.) *The Psychology of Learning and Motivation, 69*, (pp. 241-284), Elsevier Publishers.

Hupbach, A., Gomez, R., & Nadel, L. (2009). Episodic memory reconsolidation: Updating or source confusion? *Memory, 17,* 502 – 510.

JASP Team (2019). JASP (Version 0.8.6) [Computer Software].

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

Juslin, P., Olsson, N., & Winman, A. (1996). Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from the low confidence-accuracy correlation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1304 – 1316.

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 and Cognition, 38,* 541 – 554.

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.

Kellen, D., & Klauer, K. C. (2015). Signal detection and threshold modeling of confidence-rating ROCs: A critical test with minimal assumptions. *Psychological Review, 112,* 542 – 557.

Kellen, D., & Klauer, K. C. (2011). Evaluating models of recognition memory using first- and second-choice responses. *Journal of Mathematical Psychology, 55,* 251 – 266.

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.

Kleiner, M., Brainard, D., & Pelli, D., (2007). *What's new in Psychtoolbox-3?*. Perception 36 ECVP Abstract Supplement.

Krantz, D. H. (1969). Threshold theories of signal detection. *Psychological Review, 76,* 308 – 324.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5,* 213 – 236.

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

Luce, R. D. (1963). A threshold theory for simple detection experiments. *Psychological Review, 70,* 61 – 79.

Luce, R. D. (1997). Several unresolved conceptual problems of mathematical psychology. *Journal of Mathematical Psychology, 41,* 79 – 87.

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

Malejka, S. & Bröder, A. (2019). Exploring the shape of signal-detection distributions in individual recognition ROC data. *Journal of Memory and Language, 104*, 83 – 107.

Malmberg, K. J. (2002). On the form of ROCs constructed from confidence ratings. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 380 – 387.

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

The MathWorks, Inc. (2018). MATLAB (Version R2018a) [Computer Software].

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*, 1 – 11.

McAdoo, R. M., & Gronlund, S. D. (in press). Theoretical note: Exploring Luce's (1963) low-threshold model applied to recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

McAdoo, R. M., Key, K. N., & Gronlund, S. D. (2018). Stimulus effects and the mediation of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

McAdoo, R. M., Key, K. N., & Gronlund, S. D. (2019). Task effects determine whether recognition memory is mediated discretely or continuously. *Memory and Cognition*. Advanced online publication. 1 – 13.

McClelland, J. L. (2009). The place of modeling in cognitive science. *Topics in Cognitive Science, 1,* 11 – 38.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. http://www.usf.edu/FreeAssociation/.

Parks, C. M, & Yonelinas, A. P. (2009). Evidence for a memory threshold in second-choice recognition memory responses. *Proceedings of the National Academy of Sciences of the United States of the America, 106*, 11515 – 11519.

Pazzaglia, A. M., Dubé, C., & Rotello, C. M. (2013). A critical comparison of discrete-state and continuous models of recognition memory: Implications for recognition and beyond. *Psychological Bulletin, 139,* 1173 – 1203.

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.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59 – 108.

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Science, 20,* 260 - 281

Rowland, C.A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin, 140,* 1432 – 1465.

Starns, J. J., Dubé, C. & Frelinger, M. E. (2018). The speed of memory errors shows the influence of misleading information: Testing the diffusion model and discrete-state models. *Cognitive Psychology, 102,* 21 – 40.

Super, G. (1984). START: A triage training module. Newport Beach, CA: Hoag Memorial Hospital Presbyterian.

Swets, J. A. (1986). Form of empirical ROCs in discrimination and diagnostic tasks: Implications for theory and measurement of performance. *Psychological Bulletin, 99*, 181 – 198.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on p-values: Context, process, and purpose. *The American Statistician, 70,* 129 – 133.

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review, 114,* 152 – 176.

Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review, 114,* 152 – 176.

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.

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.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46,* 441 – 517.