Journal of Mathematical Psychology 75 (2016) 86–95

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

The ignored alternative: An application of Luce's low-threshold model to recognition memory

David Kellen [a], Edgar Erdfelder [b], Kenneth J. Malmberg [c,*], Chad Dubé [c], Amy H. Criss [d]

[a] Center for Cognitive and Decision Sciences, University of Basel, Switzerland
[b] Department of Psychology, University of Mannheim, Germany
[c] Department of Psychology, University of South Florida, United States
[d] Department of Psychology, Syracuse University, United States

ARTICLE INFO

Article history:
Available online 2 May 2016

Keywords:
High threshold models
Low threshold models
Signal detection theory
Recognition
ROCs

ABSTRACT

Recent years have seen an increased interest in models of recognition memory's decision stage. However, only a relatively narrow set of candidate models has been considered thus far, with comparisons being typically restricted to signal detection and high-threshold models. Here, we consider a third alternative, Luce's (1963) low-threshold model (LTM). We evaluated the LTM's predictions for existing Yes–No receiver-operating characteristic data (Dube et al., 2012) as well as data from K-alternative ranking tasks (Kellen and Klauer, 2014). The LTM, which to this point has been largely ignored in the recognition memory literature, turns out to perform at least as well as the most popular model in this domain, the Gaussian signal detection model. These results suggest future work concerning the decision stage of recognition should consider the LTM in addition to the continuous and discrete-state models that have dominated the literature so far.

One key aspect of R. Duncan Luce's seminal body of work is the focus on the basic constructs or "building blocks" underlying psychological theories. For example, Luce's later work concerned the formal representation of subjective intensities (e.g., brightness, loudness, monetary gains and losses), questioning the century-old assumption that such representations are additive (see Luce, 2000). According to Luce, the fact that traditional theories successfully account for most of the data at large does not mean that alternative accounts should be ignored (Luce, 2010). Due to their broad theoretical relevance, the theoretical results reported by Luce concerning this specific question have led to empirical work in distinct domains such as visual psychophysics (e.g., Steingrimsson, 2011) and economic decision-making (e.g., Davis-Stober & Brown, 2013).

The present work will focus on another fundamental question raised by Luce: Whether the subjective representations of stimuli that underlie individuals' judgments are *continuous* or *discrete*. This question, originally referring to psychophysical tasks in which individuals had to detect visual or auditory stimuli, led to one of Luce's most controversial contributions, the *low-threshold model* (LTM; Luce, 1963). In contrast to the popular class of signal detection (SDT) models that assume an infinite number of mental states ((Swets, Tanner, & Birdsall, 1961)), the LTM only postulated two distinct mental states – "detection" and "no detection" – which were separated by a fixed sensory threshold (for discussions on this notion, see Corso, 1963 and Rouder & Morey, 2009). Despite its simplicity, the LTM and the two mental states posited were sufficient to account for several important findings, therefore constituting a challenge to the widely-held notion that a continuous representation is necessary. Krantz (1969) discussed some of the applications and limitations of the model, and proposed an extended LTM that adressed the latter (see also Wickelgren, 1968).

The LTM received some attention shortly after its introduction (e.g., Krantz, 1969; Larkin, 1965; Lindner, 1968; Norman, 1964; Wickelgren, 1968; for a review, see Luce & Green, 1974), but in the last three decades it has been all but ignored in the literature (for a recent exception, see Hsu & Doble, 2015). To a large extent, this situation can be attributed to the enormous popularity of SDT models. According to Luce, sensory psychologists were perhaps not greatly interested in the fundamental question of whether sensation is discrete or continuous and yet were willing to dismiss the LTM without any clear empirical or formal rationale (Luce, 1997, p. 85–86). It is worth pointing out that from very early on

* Corresponding author.
E-mail addresses: davekellen@gmail.com (D. Kellen), malmberg@usf.edu (K.J. Malmberg).

http://dx.doi.org/10.1016/j.jmp.2016.03.001

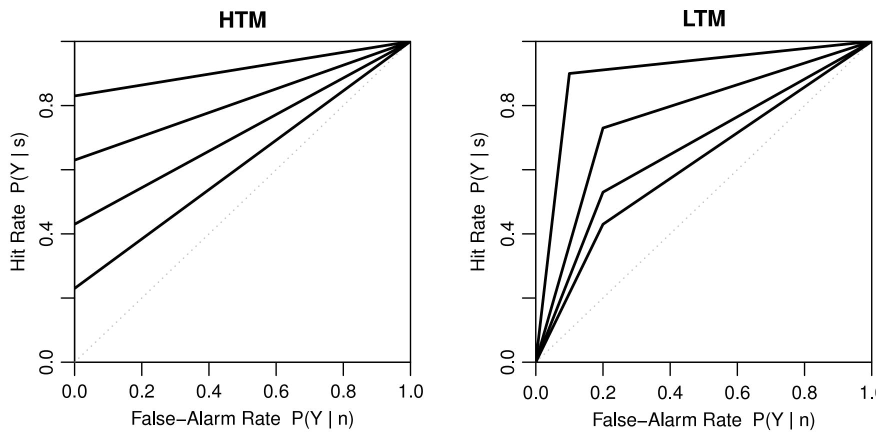

Fig. 1. Examples of ROCs as predicted by the HTM and LTM.

Luce himself admitted to holding a conflicting position in this debate (having "nagging doubts"; Luce, 1977, p. 468). Specifically, Luce was divided between the suspicion that the representation of certain sensory attributes is indeed discrete and the formal elegance and empirical success of continuous accounts (Luce, 1997, p. 85), including his own highly influential Choice Theory (Luce, 1959).

In the domain of recognition memory, distinguishing continuous and discrete accounts is a focal point of recent discussions (Batchelder & Alexander, 2013; Bröder, Kellen, Schütz, & Rohrmeier, 2013; Bröder & Schütz, 2009; Chechile, 2013; Chen, Starns, & Rotello, 2015; Dube, Rotello, & Pazzaglia, 2013; Dube & Rotello, 2012; Dube, Starns, Rotello& Ratliff, 2012; Kellen & Klauer, 2014, 2015; Kellen, Klauer, & Bröder, 2013; Malmberg, 2002; Pazzaglia, Dube, & Rotello, 2013 and Province & Rouder, 2012). Surprisingly, this debate has not considered low-threshold accounts and has instead focused on models assuming *high thresholds* (e.g., Bröder & Schütz, 2009). The purpose of the present manuscript is to fill that gap and evaluate the relative merits of the LTM in the domain of recognition memory. In addition to introducing a new element into a long-standing debate, we also honor the broad theoretical relevance of Luce's work.

First, we will introduce Luce's LTM and establish how it contrasts with typical high-threshold models. Second, very much along the lines of Luce (1963), we will evaluate the LTM's ability to account for previously published recognition data, namely Receiver Operating Characteristic (ROC) data and ranking judgments. As will be shown below, the LTM provides a reasonable account of both kinds of data, a success that beckons its consideration in further empirical and theoretical work.

1. High- and low-threshold models

In order to introduce both high- and low-threshold models in the context in which they were originally developed, let us assume a detection Yes–No (YN) task comprised of signal (s) stimulus trials and noise (n) trials in which participants are requested to indicate whether they have perceived a signal, answering "Yes" (Y) or "No" (N). Y responses to signal and noise trials are usually referred to as *hits* and *false alarms*, respectively. Initial discrete-state accounts of detection judgments (e.g., Blackwell, 1953 and Stevens, Morgan, & Volkmann, 1941) assumed a single high threshold that separates two mental states D (detection) and $\bar{D}$ (no detection). Due to the inherent assumption that state D can only be reached on trials in which a signal was presented (i.e., only signals can be "detected"), the threshold is commonly referred to as "high". If the threshold can be surpassed in noise trials as well then it would be referred to as "low".

In signal trials, the (high) threshold is surpassed with probability q_s, which in turn leads to a Y response. When detection fails with probabilities $1 - q_s$ and 1 in signal and noise trials, respectively, individuals are forced to guess whether a signal was presented or not, with response Y occurring with probability b. According to this *high-threshold model* (HTM), the probabilities of a Y response in signal and noise trials are given by

$$P(Y \mid s) = q_s + (1 - q_s)b, \quad (1)$$

$$P(Y \mid n) = b. \quad (2)$$

The relationship between hit and false-alarm probabilities when varying only response bias (i.e., the guessing parameter b) is referred to as the model's Receiver Operating Characteristic (ROC) function. Based on Eqs. (1) and (2) it is easy to see that the HTM's predicted yes–no ROC corresponds to a straight line that goes through points $(0, q_s)$ and $(1, 1)$ (see the left panel of Fig. 1). Yes–no ROC data can be obtained by selectively manipulating individuals' response biases, for instance, by varying the base rates of signal and noise trials across test blocks or by varying the payoffs associated with Y and N responses in both types of trials. In a seminal study by Swets et al. (1961), individuals' response biases in a visual detection task were varied by means of a payoff manipulation. The ROCs obtained from a sample of four participants are shown in Fig. 2. It is clear that the HTM's linear ROC predictions are grossly inconsistent with most of the individual data, with the "more conservative" ROC points (lower hits and false-alarm rates) being underestimated and the "more liberal" points (higher hits and false alarms) being overestimated. Instead, the data seem to be better captured by the SDT model (Green & Swets, 1966; Kellen & Klauer, in press and Swets et al., 1961).

In contrast to the HTM and its assumption of two mental states D and $\bar{D}$, the SDT model assumes that stimuli are presented across a continuous evidence axis, with Y judgments occurring when the evidence surpasses a response criterion κ. The criterion differs from the thresholds of LTMs and HTMs in that it is not fixed and may vary according to the individual's biases or goals. It is commonly assumed that the distributions of evidence values in signal and noise trials are Gaussian with parameters μ_s, σ_s, μ_n, and σ_n (for further details, see Green & Swets, 1966). According to the SDT model, hit and false-alarm probabilities are given by:

$$P(Y \mid s) = \Phi\left(\frac{\mu_s - \kappa}{\sigma_s}\right), \quad (3)$$

$$P(Y \mid n) = \Phi\left(\frac{\mu_n - \kappa}{\sigma_n}\right). \quad (4)$$

Luce (1963) argued that the results of Swets et al. (1961) did not demonstrate the necessity of a continuous account like SDT

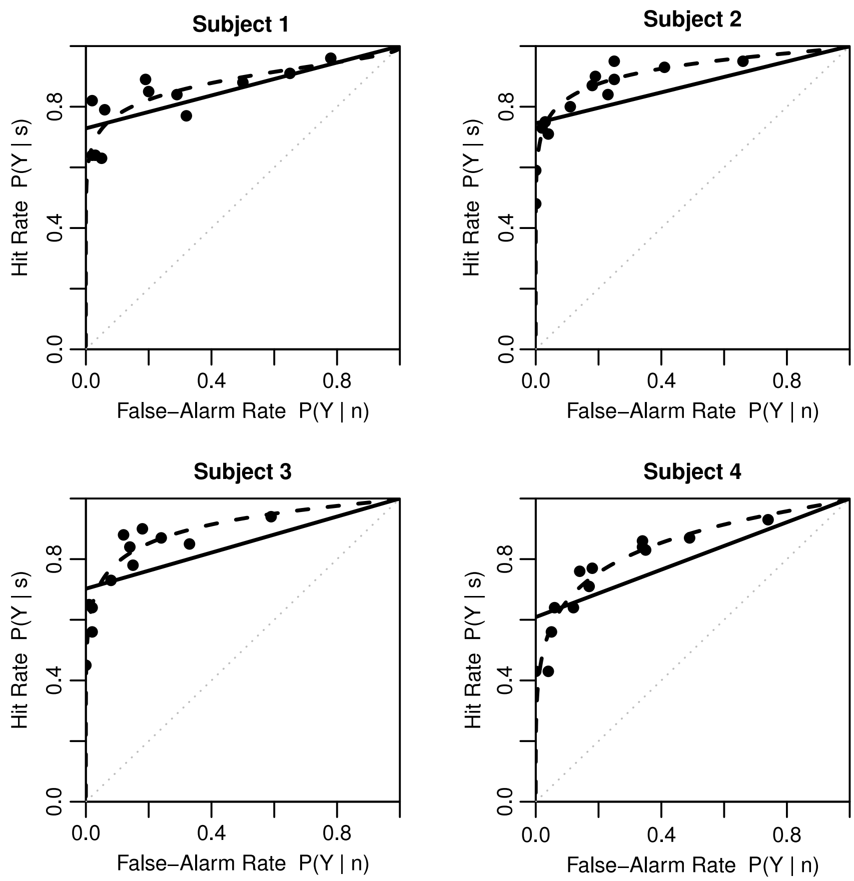

Fig. 2. ROC data from Swets et al. (1961). The solid line corresponds to the HTM predictions whereas the dashed line corresponds to the Gaussian SDT predictions. We fitted both models to the original data (reported in Table 2 of Luce, 1963) using the maximum-likelihood method.

and proposed the LTM as a viable threshold-based alternative. The LTM differs from the HTM in two important ways: First, it assumes a "low" threshold that can be reached in both signal and noise trials with probabilities q_s and q_n, with $q_s \geq q_n$. In other words, individuals might erroneously detect a "signal" when there is in fact only noise, although presumably at a lower rate than when a signal is in fact presented. Note that Luce (1963) did not make any assumptions regarding the exact way these sensory states are entered; he only proposed that the process of entering them is characterized by unknown stochastic processes. This is important because it means that the LTM is not only applicable to sensory decisions; it is equally applicable to recognition memory decisions, among others.

Second, the LTM assumes that the response mapping of mental state D depends on the individuals' responses biases: Even in the occurrence of detection, individuals can respond "No" in order to conform to the requirements of the task; for instance when adjusting the rate of Y responses according to the task's base rates or payoff schedules. The rationale here is that even though the model only assumes two sensory states, other non-sensory states affect their disposition toward each of the available response options, enabling their occurrence with some probability. Presumably the non-sensory states are independent of whatever states are relevant for the decision to be made, and therefore influence the performance in recognition memory tasks in the same manner as in sensory tasks. Therefore the following derivations are highly generalizable.[1]

According to the LTM, the hit and false-alarm probabilities are given by

$$P(Y \mid s) = \begin{cases} q_s(b+1), & \text{if } b < 0 \quad \text{(conservative)} \\ q_s + (1-q_s)b, & \text{if } b \geq 0 \quad \text{(liberal)} \end{cases} \tag{5}$$

$$P(Y \mid n) = \begin{cases} q_n(b+1), & \text{if } b < 0 \quad \text{(conservative)} \\ q_n + (1-q_n)b, & \text{if } b \geq 0 \quad \text{(liberal)} \end{cases} \tag{6}$$

with $-1 \leq b \leq 1$.

According to the LTM, individuals can express a "liberal" or "conservative" response strategy or mode. Under a liberal strategy, individuals always respond Y when state D is reached. When in state $\bar{D}$, response Y is produced via guessing with probability b. Under the liberal strategy, the minimum $P(Y \mid s)$ and $P(Y \mid n)$ are q_s and q_n, respectively (occurring when guessing probability b is zero), and the predicted ROC points correspond to a line segment between coordinates (q_s, q_n) and $(1, 1)$. Under the conservative strategy, all stimuli in state $\bar{D}$ are rejected, and stimuli in state D are only accepted with probability $b + 1$. The notion here is that when all stimuli of uncertain status are rejected, individuals express an increasing conservatism by starting to reject stimuli that were detected. This conservative strategy constrains hit and false-alarm probabilities to range from a minimum of 0 to q_s and q_n, respectively. These constraints yield a linear ROC segment from $(0, 0)$ to (q_n, q_s). Altogether, the LTM's predictions across response strategies and values of b yield a piecewise linear ROC comprised

[1] Note that the occurrence of "response reversals" (response N in state D) is not unreasonable or unexpected (e.g., Van Zandt & Maldonado-Molina, 2004), especially if individuals are somewhat aware that they will sometimes reach state D in noise trials. In fact, the well-known phenomenon of probability matching (e.g., Healy & Kubovy, 1981) suggests that such behavior is not uncommon.

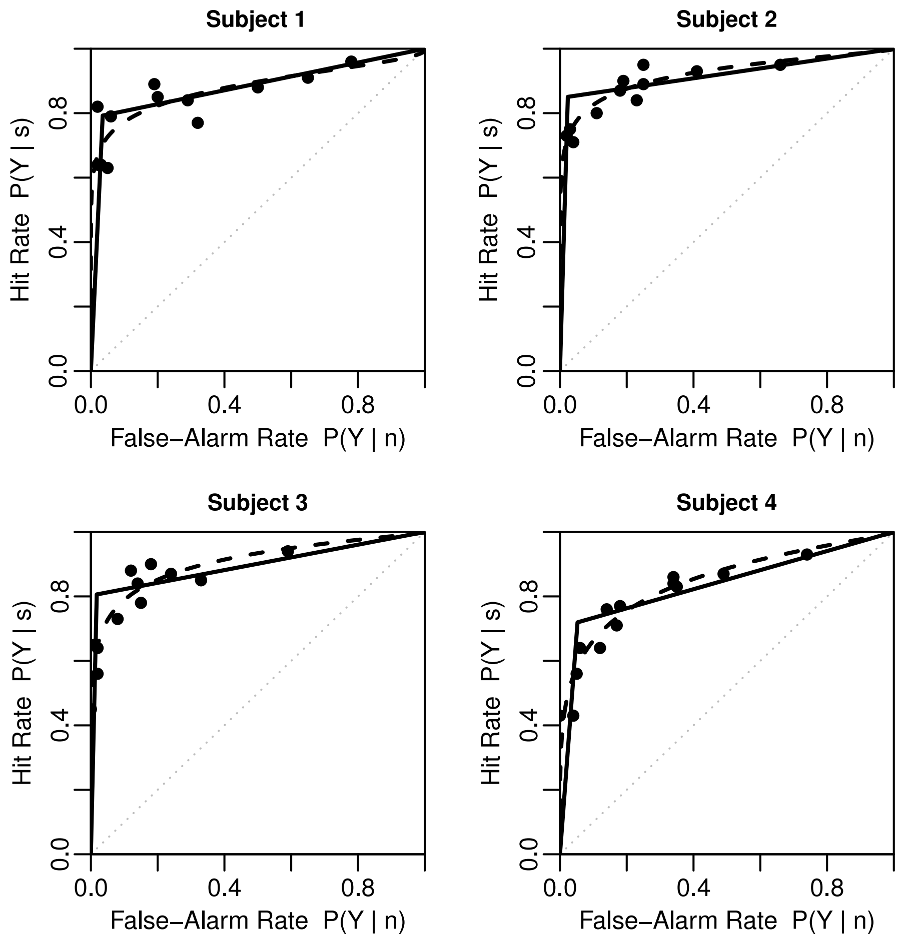

Fig. 3. ROC data from Swets et al. (1961). The solid line corresponds to the LTM predictions whereas the dashed line corresponds to the Gaussian SDT predictions. We fitted both models to the original data (reported in Table 2 of Luce, 1963) using the maximum-likelihood method.

of two branches that meet at the point (q_s, q_n) (see the right panel of Fig. 1).[2]

As shown in Fig. 3, when applied to the visual detection data from Swets et al. (1961) using the maximum-likelihood method, the LTM manages to provide a reasonable approximation. The most obvious misfit of the LTM to the data comes from the predicted "corner" at (q_n, q_s): Luce (1963) argued that such a mismatch between the fitted model and the data can emerge when there is a probabilistic adjustment of strategies across trials (i.e., individuals shift between the two ROC segments; see also Wickelgren, 1968, p. 128).[3] In light of these results, Luce (1963) concluded that a low-threshold account is not "clearly wrong" and therefore should not be automatically excluded in favor of a continuous SDT account.

1.1. 2AFC ROC data

Luce (1963) also discussed the LTM in the context of a two-alternative forced-choice (2AFC) task: During each trial, a signal and a noise stimulus are presented in two separate observation intervals ($\langle s, n\rangle$ and $\langle n, s\rangle$). The participants task is to correctly identify the interval in which the signal stimulus was included, with "1" and "2" denoting the responses "first observation interval" and "second observation interval", respectively. Luce (1963) assumed that individuals make their choices based on the joint mental states $\langle D, \bar{D}\rangle$, $\langle \bar{D}, D\rangle$, $\langle D, D\rangle$, and $\langle \bar{D}, \bar{D}\rangle$. In the first two joint mental states, the decision for the subject is rather easy, given that only one stimulus was detected (unless the experiment somehow imposes extreme response biases), but the latter two states require the subject to guess. Note that in the case of $\langle D, D\rangle$, individuals might end up choosing the noise stimulus even though the signal was detected as well. When $q_s(1-q_n) < P(1 \mid \langle s, n\rangle) \le 1 - q_s(1-q_n)$ and $q_n(1-q_s) < P(1 \mid \langle n, s\rangle) \le 1 - q_n(1-q_s)$, the probability of the first observation interval being selected as the one containing the signal in $\langle s, n\rangle$ and $\langle n, s\rangle$ trials is

$$P(1 \mid \langle s, n\rangle) = q_s(1-q_n) + q_s q_n v + (1-q_s)(1-q_n)w, \quad (7)$$

$$P(1 \mid \langle n, s\rangle) = q_n(1-q_s) + q_n q_s v + (1-q_n)(1-q_n)w, \quad (8)$$

where v and w are guessing parameters for states $\langle D, D\rangle$ and $\langle \bar{D}, \bar{D}\rangle$. It is easy to see that

$$P(1 \mid \langle s, n\rangle) = P(1 \mid \langle n, s\rangle) + q_s - q_n, \quad (9)$$

which corresponds to a linear ROC segment with slope 1 running from coordinates $\{q_n(1-q_s), q_s(1-q_n)\}$ to $\{1 - q_s(1-q_n),$

[2] The definition of the LTM differs superficially from Luce's (see Luce, 1963, p. 64). For example, according to Luce's definition, $P(Y \mid n) = q_n b'$ if $P(Y \mid n) < q_n$ (conservative), and $P(Y \mid n) = q_n + (1-q_n)b''$ if $P(Y \mid n) \ge q_n$ (liberal), with $0 \le b', b'' \le 1$. It is easy to see that if $b' = b + 1$ and $b'' = b$. The two parametrizations are equivalent, given that $P(Y \mid n) < q_n$ when $b < 0$ and $P(Y \mid n) \ge q_n$ when $b \ge 0$. We prefer the present definition of the LTM as it is more intuitive.

[3] The same way that the aggregation of data coming from two strategies can lead to mismatches between the model and the data, aggregating data from different participants also leads to distortions. Another issue is the unlikely assumption that trials are independent and identically distributed (i.e., that there are no sequential effects). Atkinson (1963) proposed a model in which detection probabilities change across trials, depending on the type of trials that were previously encountered, and its effect on the predicted ROCs.

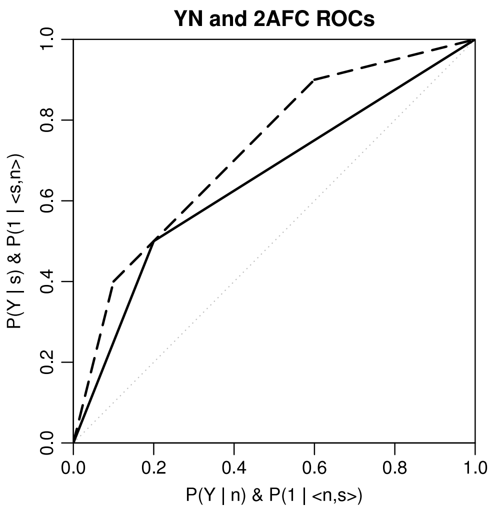

Fig. 4. Example of YN and 2AFC ROC (solid and dashed lines, respectively) as predicted by the LTM with parameters $q_s = 0.50$ and $q_n = 0.20$.

$1 - q_n(1 - q_s)\}$. Moreover, if one sets $v = w = \frac{q_s q_n}{q_s q_n + (1-q_s)(1-q_n)}$, it is easy to see that the YN ROC always meets the 2AFC ROC from below at coordinate $\{q_n, q_s\}$, which corresponds to the "corner" of the YN ROC. Examples of 2AFC and YN ROCs are given in Fig. 4.

Due to the lack of appropriate data, Luce did not conduct an evaluation of the LTM predictions for 2AFC data (in particular the existence of a linear segment with slope 1) nor its relationship with YN data. Fortunately, relevant data were reported by Atkinson and Kinchla (1965), who collected 2AFC ROCs obtained via a payoff manipulation. Fig. 5 shows the 2AFC ROC data from the eight participants that took part in this study, data that support the LTM prediction of a linear ROC segment with slope 1 (for similar data, see also Kinchla, Townsend, Vellott, & Atkinson, 1966 and Norman, 1964). Luce later expressed his frustration with the fact that the consistency between 2AFC data and the LTM predictions did not affect researchers' reluctance toward adopting threshold accounts (see Luce, 1997, pp. 84).

Now that we have described the LTM in the context in which it was originally proposed, we will turn to the domain of recognition memory in which the same SDT and HTM accounts proposed in the perception domain are usually compared. Importantly, consideration of the LTM has so far been almost completely absent from this domain.

2. Evaluating the LTM with recognition-memory data

Recognition memory is the discrimination of events one has experienced from events one has not experienced in a specific past context. This basic faculty is particularly important in testing models of human memory (for reviews, see Malmberg, 2008 and Raaijmakers & Shiffrin, 2002). In the laboratory, recognition memory is most often studied by presenting subjects with a sequence of stimuli to study, and then testing their memory for those stimuli via a yes–no, forced choice, or confidence-rating procedure. Over the past few decades, memory researchers have debated whether the basis of recognition judgments is a continuous representation of a prior occurrence of a stimulus or a discrete-state representation. The genesis of this debate can be traced back to early formal-modeling approaches to memory formation, representation, and retrieval (Gillund & Shiffrin, 1984). Among the first applications of a continuous account was the Gaussian SDT model, which succeeded in describing individuals' judgments about the previous experience of stimuli (Egan, Schulman, & Greenberg, 1959). Within the framework of memory models, the evidence axis in SDT represents the familiarity of tested stimuli.

However, like Luce (1963), some memory researchers had their doubts that the prior occurrence of a stimulus could only be described by a continuous model. For instance, some researchers proposed models assuming that the recognition of studied items sometimes occurs via the successful retrieval of episodic details (i.e., these models postulate the existence of discrete memory states; e.g., Atkinson & Juola, 1974 and Mandler, Pearlstone, & Koopmans, 1969). In the absence of episodic retrieval, recognition judgments are assumed to rely on the familiarity of the stimulus, as postulated by SDT. Due to the assumption of two forms of remembering these models are usually referred to as *dual-process models* (for reviews, see Malmberg, 2008; Wixted, 2007 and Yonelinas & Parks, 2007). Meanwhile, other researchers have argued that models that exclusively assume discrete states by means of high thresholds provide a sufficient account of recognition-memory judgments (e.g., Bröder & Schütz, 2009; Chechile, 2013; Kellen et al., 2013; Klauer & Kellen, 2010 and Province & Rouder, 2012).

So far, the recognition memory literature has focused on SDT, HTM, and dual-process models without considering the LTM (for exceptions, see Batchelder, Riefer, & Hu, 1994 and Bayen, Murnane, & Erdfelder, 1996). The LTM's assumption that new items can be detected as old makes it a particularly interesting candidate in the case of recognition memory given the predominance of phenomena involving the false remembering of new items (for a review, see Malmberg, 2008). But how does the LTM actually fare when fitted to recognition memory data? In an attempt to answer this question we examine the performance of LTM in fits to ranking and ROC data from recognition memory tasks (Dube et al., 2012; Kellen & Klauer, 2014). A critical property of the data coming from these two studies is that in both cases the Gaussian SDT model was found to provide a better account than the high-threshold competitor.[4] Also, both studies included a within-list study-strength manipulation (some of the items were presented multiple times during the study phase) that selectively affected the detection of old items, thereby imposing additional constraints on the model (Province & Rouder, 2012; see also Krantz, 1969, p. 312).[5]

Moreover, one novelty of the present analyses is that we also evaluate the LTM using data from a ranking task. The LTM has never been evaluated using this kind of data (Kellen & Klauer, 2014; see also Krantz, 1969, p. 316). In a ranking task, participants rank the items according to their likelihood of being the studied item. The analysis of ranking data is particularly relevant in the present context given that the HTM makes clear predictions that have been recently rejected in favor of SDT accounts (for details, see Kellen & Klauer, 2014).

Finally, in order to establish a reasonable reference point when evaluating the LTM's performance, we will compare its goodness of fit with that of the Gaussian SDT model. The purpose of this comparison is to allow us to see whether any failure of LTM is necessarily due to its assumption of discrete states. Although there are good reasons to believe that goodness-of-fit results do not provide the most sensible assessment of model performance (e.g., Myung, 2000), they are nevertheless able to indicate whether a model grossly fails to account for the present data.

4 The extended high threshold model used in the recognition-memory literature, the *two-high-threshold model*, assumes three memory states rather than two. Old items are detected as old and new items are detected as new with probabilities D_o and D_n, respectively. When both detection processes fail, items are in an uncertainty state and judged according to a guessing process (see Bröder & Schütz, 2009).

5 It should be noted that this selective influence of old-item detection is ensured by the fact that both weak and strong items (e.g., items studied once and thrice, respectively) are studied together in the same study phase, with both item types having no distinctive feature (e.g., a difference in color). When these conditions are not met, study-strength effects are expected to produce changes in both old- and new-item detection (see Criss, 2006, 2009, 2010).

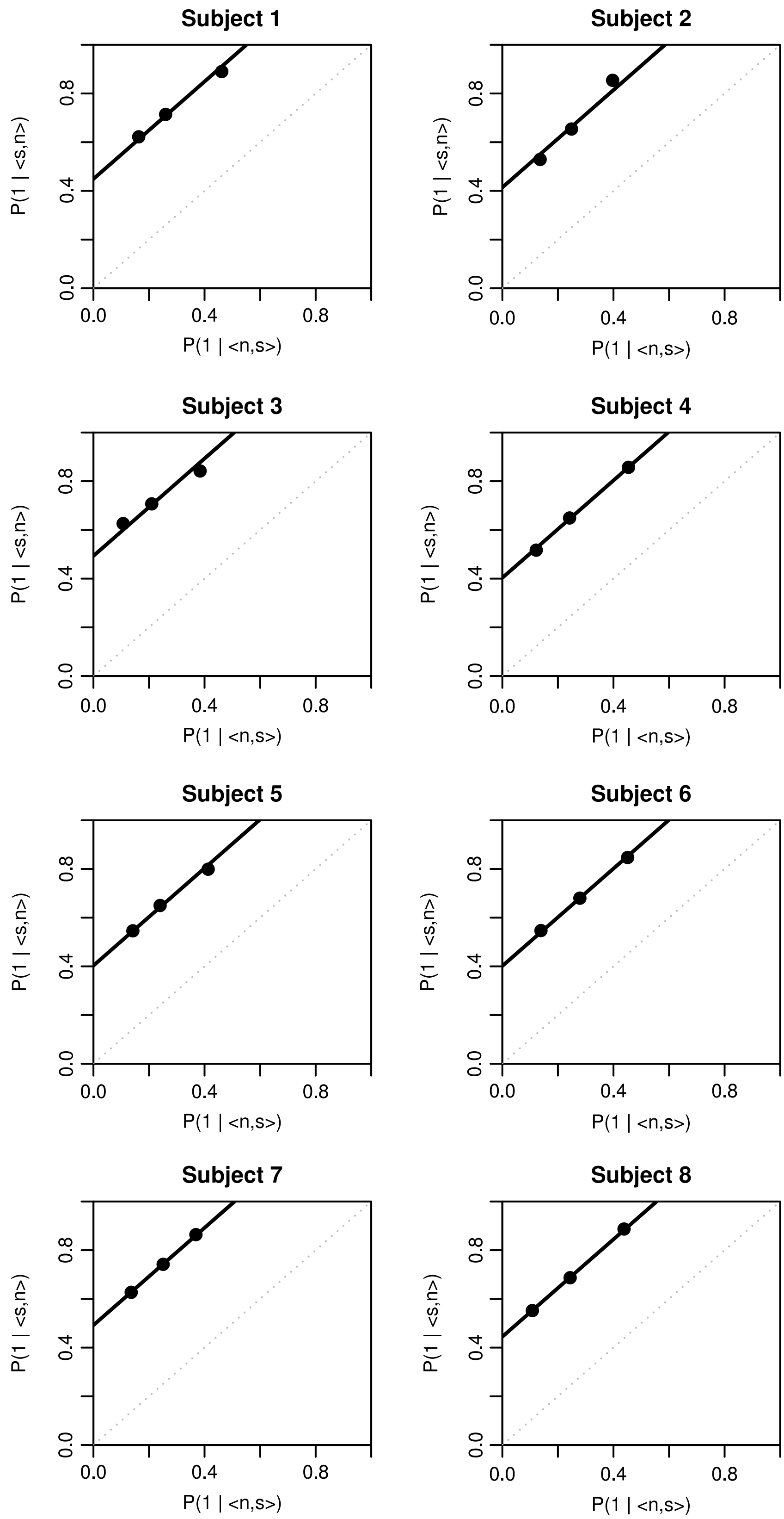

Fig. 5. ROC data from Atkinson and Kinchla (1965). The solid black lines correspond to the LTM predictions we obtained via maximum-likelihood estimation. For convenience, we fixed q_n to be zero.

2.1. Old–new ROC data

The ROCs analyzed here come from two experiments originally reported by Dube et al. (2012), who collected old–new (ON) recognition judgments across five response-bias conditions. The level of response bias was manipulated by varying the base rate of old and new items across different study-test blocks. In both experiments half of the old items were studied once and the

Table 1
Model fits.

Data	N	LTM		SDT	
		Summed G^2	$p < 0.05$	Summed G^2	$p < 0.05$
Dube et al. (2012, Exp 1)	21	191.97 (67%)	14%	216.20 (33%)	29%
Dube et al. (2012, Exp 2)	26	250.29 (73%)	19%	267.66 (27%)	27%
Kellen and Klauer (2014, Exp 1)	22	53.74 (55%)	0%	58.56 (45%)	0%
Kellen and Klauer (2014, Exp 2)	23	37.34 (52%)	17%	28.65 (48%)	13%

Note. N = number of participants. The values in parentheses correspond to the percentage of datasets for which the respective model provided the best fit. Columns "$p < 0.05$" indicate the percentage of datasets where a model produced statistically-significant misfits.

Table 2
Median parameter estimates (interquartile ranges).

Data	LTM			SDT		
	q_s^w	q_s^s	q_n	μ_s^w	μ_s^s	σ_s
Dube et al. (2012, Exp 1)	0.64 (0.10)	0.91 (0.11)	0.11 (0.12)	1.52 (1.35)	2.56 (3.13)	1.38 (2.17)
Dube et al. (2012, Exp 2)	0.57 (0.19)	0.90 (0.10)	0.08 (0.14)	1.23 (0.82)	2.38 (4.45)	1.39 (2.80)
Kellen and Klauer (2014, Exp 1)	0.34 (0.18)	0.63 (0.25)	0.09 (0.16)	0.40 (0.34)	0.90 (0.79)	1.24 (0.36)
Kellen and Klauer (2014, Exp 2)	0.52 (0.25)	0.84 (0.14)	0.07 (0.08)	1.04 (0.94)	2.09 (1.42)	1.49 (0.41)

other half multiple times (five and 10 times in Experiments 1 and 2, respectively). The experimental design used in both studies provides a total of fifteen degrees of freedom.

In order to fit the data from these two studies, the LTM requires eight free parameters: Three detection parameters for strong, weak, and new items (q_s^w, q_s^s, and q_n), and five response bias parameters b_i, $i = 1, \ldots, 5$. The Gaussian SDT model implemented here had the same number of parameters (see Dube et al., 2012): Two mean familiarities for weak and strong items (μ_s^w and μ_s^s), a common standard-deviation parameter for studied items (σ_s) and five response criteria κ_i (note that μ_n and σ_n are fixed to 0 and 1 without loss of generality). The LTM and SDT models were fitted with the R package MPTinR ((Singmann & Kellen, 2013)) using the maximum-likelihood method. Multiple fitting runs were conducted in order to avoid local minima.

As can be seen in Table 1, the overall fit results (quantified via the G^2 statistic) were reasonable, with the LTM slightly outperforming the SDT model in both experiments. Also, the LTM produced significant misfits ($p < 0.05$) less often that the SDT model. However, both models provided similar accounts as they tended to succeed and fail together (smallest G^2 correlation $r = 0.87$, largest $p < 0.001$). The median parameter estimates (see Table 2) from both models seem reasonable. For reference purposes, Fig. 6 depicts the ROC data that corresponded (the closest) to the LTM's median fits in the two experiments. Altogether, this analysis shows that the LTM fits this set of ROC data at least as well as the SDT model.

2.2. *Ranking judgments*

The ranking data come from two experiments with a total of forty-five participants. In each trial of a K-alternative ranking task, individuals ranked $K - 1$ new items and 1 old item according to their belief that the items were previously studied (Rank 1 being attributed to the item judged to be the one most likely to be old). Participants were informed that only one alternative in each trial was actually old. The probability that the old item is assigned rank i is denoted by R_i. Kellen and Klauer's (2014) Experiments 1 and 2 used a four- and a three-alternative ranking task, respectively. In addition to the number of alternatives per test trial, the two experiments differed in terms of the study time of each word (600 ms and 1200 ms in Experiments 1 and 2, respectively).

Like the previously-analyzed ROC experiments, both studies included a study-strength manipulation, with half of the old items in the study list being studied once (weak words) and the other half thrice (strong words). This study-strength manipulation played a critical role in analyses conducted by Kellen and Klauer (2014): Let c_2 be the conditional probability that the old item is assigned rank 2, given that it was not assigned rank 1; i.e., $c_2 = \frac{R_2}{1-R_1}$. The SDT model predicts c_2 to be larger for strong words than weak words ($c_2^w < c_2^s$) while the HTM predicts no difference whatsoever ($c_2^w = c_2^s$; see Kellen & Klauer, 2014, Appendix). The results from both experiments showed greater c_2 values for strong words, thereby rejecting the null-effect prediction established by the HTM.

The predictions of the LTM for the case of ranking judgments can be derived from Luce's (1963) approach for the case of 2AFC judgments: Higher rankings (lower i) are attributed to detected items; when k items are detected, the first k rankings are randomly attributed to them. The remaining ranks are randomly attributed to non-detected items. According to the LTM, the probability of an old item among $K - 1$ new items being assigned rank i is given by

$$R_i = q_s\xi_i + (1 - q_s)\eta_i, \tag{10}$$

with

$$\xi_i = \sum_{j=i}^{K} \frac{1}{j} \binom{K-1}{j-1} q_n^{j-1}(1-q_n)^{K-j}, \tag{11}$$

$$\eta_i = \sum_{m=0}^{i-1} \frac{1}{K-m} \binom{K-1}{m} q_n^{m}(1-q_n)^{K-m-1}. \tag{12}$$

It can be shown that the LTM is in line with SDT as it also predicts that c_2 will increase in the context of study-strength manipulations. This prediction stems from the fact that detected new items can be attributed rank 1 even when the old item was also detected. Let $f(q_s) = R_2(q_s, q_n)$ and $g(q_s) = 1 - R_1(q_s, q_n)$, and $c_2(q_s) = \frac{f(q_s)}{g(q_s)}$ for some fixed value of q_n. For example, when $K = 3$, the derivative $c_2'(q_s) = 3q_n \times (-q_n^2 + q_s q_n - 2q_s + 2q_n + 2)^{-2}$ is a strictly non-negative function, indicating that c_2 increases as a function of q_s. Equivalent results are found for other K.

Contrary to the case of ROC data no response-bias parameters are estimated here, only the "structural" parameters related to detection probabilities (q_s^w, q_s^s, and q_n) and the familiarity distributions (μ_s^w, μ_s^s, and σ_s). Experiments 1 and 2 provide six and four degrees of freedom, respectively, which are sufficient to estimate these parameters and test the models' goodness of fit. As shown in Tables 1 and 2, the LTM and the SDT models provided a similarly good account of the data and the observed c_2 (see Fig. 7). As with the ROC data, the two models produced similar accounts of the ranking judgments in the sense that they tended to fail and succeed in the same cases (smallest G^2 correlation $r = 0.50$, largest $p = 0.02$).

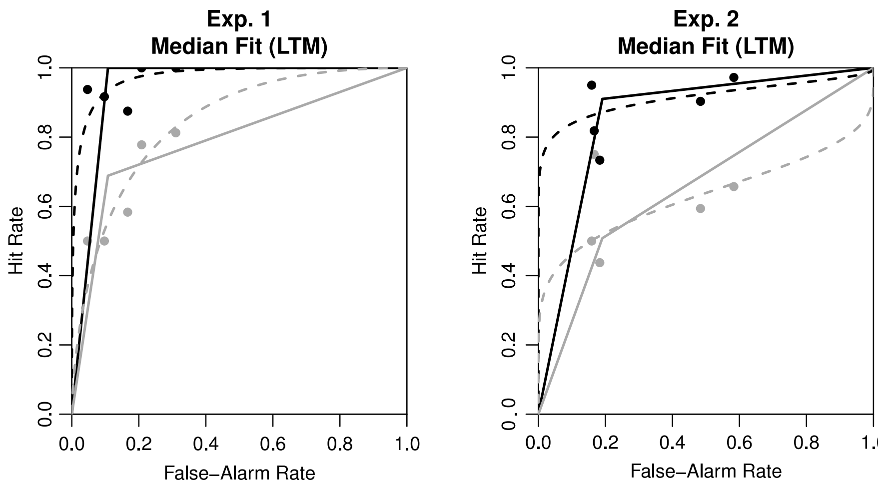

Fig. 6. Individual ROC data Dube et al. (2012) that correspond to the LTM's median fits (participants 16 and 14 in Experiments 1 and 2, respectively). The solid lines correspond to the LTM predictions whereas the dashed lines correspond to the Gaussian SDT predictions.

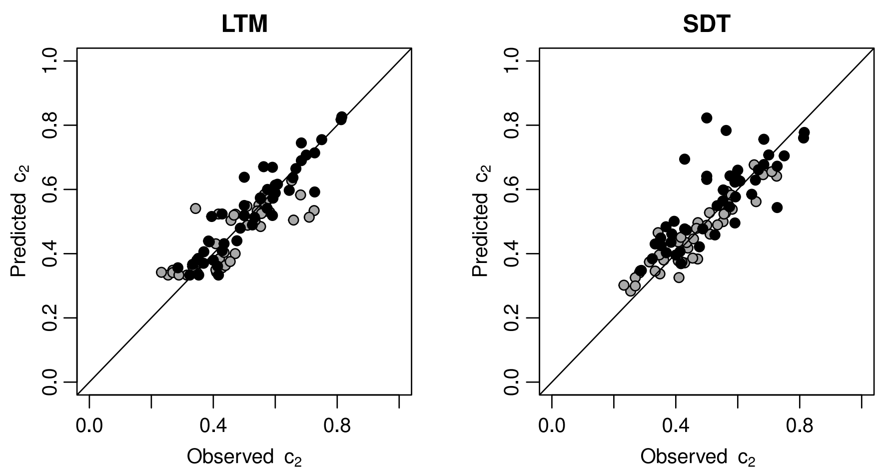

Fig. 7. Observed individual c_2 values for weak and strong words (in gray and black, respectively) plotted against their respective LTM and SDT predictions.

3. Discussion

Luce argued that the LTM did not receive a sufficient level of attention and scrutiny from researchers (Luce, 1997). With this in mind, the present work aimed at introducing the model to a new audience and evaluating its performance in the domain of recognition memory, where it has been all but ignored. As it turns out, the LTM is able to account for ROC and ranking data at least as well as the standard SDT model. We hope that these results will encourage researchers to consider this model in future work. The use of the LTM seems particularly promising in the characterization of associate and semantic false memories, phenomena that have led to the development of models incorporating erroneous retrieval processes (see Brainerd, Gomes, & Moran, 2014). The low thresholds assumed by the LTM could account for these false recognitions in a simple and straightforward manner.

Although the LTM provided a good account of the present data, further tests are required. In this respect, it is important to keep in mind that failures are to be expected. Also, that solutions have been proposed for some of these failures. For example, Chen et al. (2015) showed that threshold models cannot adequately account for study-strength manipulations that include conditions in which performance is virtually perfect. This failure was originally discussed by Krantz (1969), who proposed that in addition to $\bar{D}$ and D, there is a "super-detection" state D^* in which the individual is absolutely certain that the item was previously studied. This third state is associated with three important assumptions: (a) only old items can reach it; (b) old items in this state are always judged to be old; and (c) state D^* dominates D, which means that when two items are in these two states (e.g., in the context of a 2AFC task), the one in state D^* is always chosen to be the old one.

Moreover, some LTM failures could be attributed to the exact way mental states and biases are mapped onto observed responses. Luce (1963) assumed that N responses to detected items can only occur when the same answer is given to all items in $\bar{D}$ (i.e., when individuals assume a conservative strategy). As shown by Rouder, Province, Swagman, and Thiele (submitted for publication), a more general state-response mapping function can be introduced by allowing all responses to be produced under all states. Thus, in the case of the YN task, the probabilities of hits and false alarms correspond to:

$$P(Y \mid s) = q_s h + (1 - q_s)b, \tag{13}$$

$$P(Y \mid n) = q_n h + (1 - q_n)b, \tag{14}$$

where h and b are response-bias parameters. Rouder et al. (submitted for publication) showed that this model can account for *any single* ROC function but is nevertheless testable by means of its assumption of *conditional independence* (Province & Rouder, 2012; see also Krantz, 1969)[6]: Conditional on a state being reached,

[6] Note that the SDT model, when devoid of parametric assumptions, can describe *any* ROC data (Rouder et al., submitted for publication).

the probability of a given response is only dependent on the state-response mapping parameter (e.g., conditional on D, the probability of response Y is always h). When the experimental design introduces a study-strength manipulation that only affects the probability of reaching state D (as in the memory studies used in the present reanalysis), threshold models are bound to make more precise or constrained predictions (see Province & Rouder, 2012 and Rouder et al., submitted for publication).

Finally, it is important to note that the theorizing and modeling of the processes involved in memory occurs at a lower level than that of the LTM. Perhaps for this reason, *measurement models* such as the HTM and LTM have been applied to tasks as varied as perceptual classification and source memory under the mistaken assumption that they make no theoretical assumptions which could affect their measurement accuracy (Pazzaglia et al., 2013). On the other hand, as pointed out by Dube et al. (2013), measurement models such as the LTM and SDT model constrain the space of possible "*process*" *models* through their strong constraints on the nature of the process' output (e.g., the number of discrete states, the shape of the evidence distributions). These assumptions directly impact their accuracy and validity as measurement tools. Furthermore, measurement models may also have strong implications for lower level "process" models to the extent that the measurement models' general assumptions are justified for a given dataset. For instance, if a measurement model is disconfirmed by some new finding, then all members of the set of process models that the measurement model describes are also disconfirmed. Note that the converse does not necessarily hold because process models can be disconfirmed based on factors other than those that are incorporated into a particular measurement model.

In sum, the LTM turns out to provide a good account of existing recognition memory data. The model can be tailored to different types of recognition memory tasks, and it can be generalized in several directions, showing that the scope of threshold models of recognition is larger than previously thought.

Acknowledgment

We thank Bill Batchelder for his valuable comments.

References

Atkinson, R. C. (1963). A variable sensitivity theory of signal detection. *Psychological Review*, *70*, 91–106. http://dx.doi.org/10.1037/h0041428.

Atkinson, R. C., & Juola, J. F. (1974). Search and decision processes in recognition memory. In D. H. Krantz, R. D. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology, Vol. 1: Learning, memory and thinking* (pp. 243–293). San Francisco: Freeman.

Atkinson, R. C., & Kinchla, R. A. (1965). A learning model for forced-choice detection experiments. *British Journal of Mathematical and Statistical Psychology*, *18*, 183–206. http://dx.doi.org/10.1111/j.2044-8317.1965.tb00341.x.

Batchelder, W. H., & Alexander, G. E. (2013). Discrete-state models: Comment on Pazzaglia, Dube, and Rotello (2013). *Psychological Bulletin*, *139*, 1204–1212. http://dx.doi.org/10.1037/a0033894.

Batchelder, W. H., Riefer, D. M., & Hu, X. (1994). Measuring memory factors in source monitoring: Reply to Kinchla. *Psychological Review*, *101*, 172–176. http://dx.doi.org/10.1037//0033-295X.101.1.172.

Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 197–215. http://dx.doi.org/10.1037//0278-7393.22.1.197.

Blackwell, H.R. (1953). Psychological thresholds: Experimental studies of methods of measurement (Bulletin No. 36). University of Michigan, Engineering research Institute.

Brainerd, C. J., Gomes, C. F. A., & Moran, R. (2014). The two recollections. *Psychological Review*, *121*, 563–599. http://dx.doi.org/10.1037/a0037668.

Bröder, A., Kellen, D., Schütz, J., & Rohrmeier, C. (2013). Validating a two-high threshold model for confidence rating data in recognition memory. *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.

Chechile, R. A. (2013). A novel method for assessing rival models of recognition memory. *Journal of Mathematical Psychology*, *57*, 196–214. http://dx.doi.org/10.1016/j.jmp.2013.07.002.

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.

Corso, J. F. (1963). A theoretico-historical review of the threshold concept. *Psychological Bulletin*, *60*, 356–370. http://dx.doi.org/10.1037/h0040633.

Criss, A. H. (2006). The consequences of differentiation in episodic memory: similarity and the strength based mirror effect. *Journal of Memory and Language*, *55*(4), http://dx.doi.org/10.1016/j.jml.2006.08.003.

Criss, A. H. (2009). The distribution of subjective memory strength: List strength and response bias. *Cognitive Psychology*, *59*, 297–319. http://dx.doi.org/10.1016/j.cogpsych.2009.07.003.

Criss, A. H. (2010). Differentiation and response bias in episodic memory: Evidence from reaction time distributions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 484–499. http://dx.doi.org/10.1037/a0018435.

Davis-Stober, C. P., & Brown, N. (2013). Evaluating decision maker "type" under p-additive utility representations. *Journal of Mathematical Psychology*, *57*, 320–328. http://dx.doi.org/10.1016/j.jmp.2013.08.002.

Dube, 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. http://dx.doi.org/10.1037/a0024957.

Dube, C., Rotello, C., & Pazzaglia, A. (2013). The statistical accuracy and theoretical status of discrete-state MPT models: Reply to Batchelder and Alexander (2013). *Psychological Bulletin*, *139*, 1213–1220. http://dx.doi.org/10.1037/a0034453.

Dube, C., Starns, J. J., Rotello, C. M., & Ratliff, 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., Schulman, A. I., & Greenberg, G. Z. (1959). Operating characteristics determined by binary decisions and by ratings. *Journal of Acoustical Society of America*, *31*, 768. http://dx.doi.org/10.1121/1.1907783.

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.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Healy, A. F., & Kubovy, M. (1981). Probability matching and the formation of conservative decision rules in a numerical analog of signal detection. *Journal of Experimental Psychology: Human Learning and Memory*, *7*, 344–354. http://dx.doi.org/10.1037/0278-7393.7.5.344.

Hsu, Y.-F., & Doble, C. W. (2015). A threshold theory account of psychometric functions with response confidence under the balance condition. *British Journal of Mathematical and Statistical Psychology*, *68*, 158–177. http://dx.doi.org/10.1111/bmsp.12040.

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/10.1037/0278-7393.30.6.1147.

Kellen, D., & Klauer, K. C. (2016). Elementary signal detection and threshold theory. In E.-J. Wagenmakers (Ed.), *Stevens' handbook of experimental psychology and cognitive neuroscience. Vol. V* (4th ed.). New York: John Wiley & Sons, Inc., (in press).

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.

Kinchla, R., Townsend, J., Vellott, J., & Atkinson, R. (1966). Influence of correlated visual cues on auditory signal detection. *Perception & Psychophysics*, *1*, 67–73. http://dx.doi.org/10.3758/BF03207824.

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

Krantz, D. H. (1969). Threshold theories of signal detection. *Psychological Review*, *76*, 308–324. http://dx.doi.org/10.1037/h0027238.

Larkin, W. D. (1965). Rating scales in detection experiments. *Journal of Acoustical Society of America*, *37*, 748–749.

Lindner, W. A. (1968). Recognition performance as a function of detection criterion in a simultaneous detection-recognition task. *The Journal of the Acoustical Society of America*, *44*(1), 204–211. http://dx.doi.org/10.1121/1.1911056.

Luce, R. D. (1959). *Individual choice behavior*. New York: Wiley.

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. (1977). Thurstones's discriminal processes fifty years later. *Psychometrika*, *42*, 461–489. http://dx.doi.org/10.1007/bf02295975.

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.

Luce, R. D. (2000). *Utility of gains and losses: measurement-theoretical and experimental approaches*. New York: Psychology Press.

Luce, R. D. (2010). Behavioral assumptions for a class of utility models: a program of experiments. *Journal of Risk and Uncertainty*, *41*, 19–37. http://dx.doi.org/10.1007/s11166-010-9098-5.

Luce, R. D., & Green, D. M. (1974). Detection, discrimination, and recognition. In E. C. Carterette, & M. P. Friedman (Eds.), *Handbook of perception. Vol. II* (pp. 299–342). New York: Academic Press.

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. *Cognitive Psychology*, *57*, 335–384. http://dx.doi.org/10.1016/j.cogpsych.2008.02.004.

Mandler, G., Pearlstone, Z., & Koopmans, H. S. (1969). Effects of organization and semantic similarity on recall and recognition. *Journal of Verbal Learning and Verbal Behavior*, *8*, 410–423. http://dx.doi.org/10.1016/S0022-5371(69)80134-9.

Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology*, *44*, 190–204.

Norman, D. A. (1964). Sensory thresholds, response biases, and the neural quantum theory. *Journal of Mathematical Psychology*, *1*, 88–120. http://dx.doi.org/10.1016/0022-2496(64)90018-5.

Pazzaglia, A., Dube, C., & Rotello, C. (2013). A critical comparison of discrete-state and continuous models of recognition memory: Implications for recognition and beyond. *Psychological Bulletin*, *139*, 1173–1203. http://dx.doi.org/10.1037/a0033044.

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.

Raaijmakers, J. G. W., & Shiffrin, R. M. (2002). Models of memory. In H. Pashler, & D. Medin (Eds.), *Stevens' handbook of experimental psychology. Vol. II* (3rd ed.). New York: John Wiley & Sons, Inc.

Rouder, J., & Morey, R. D. (2009). The nature of psychological thresholds. *Psychological Review*, *116*, 655–660. http://dx.doi.org/10.1037/a0016413.

Rouder, J.N., Province, J.M., Swagman, A.R., & Thiele, J.E. (2013). From ROC curves to psychological theory. Manuscript submitted for publication.

Singmann, H., & Kellen, D. 2013. MPTinR: Analysis of Multinomial Processing Tree models with R. Behavior Research Methods, 45, 560–575.

Steingrimsson, R. (2011). Evaluating a model of global psychophysical judgments for brightness: II. Behavioral properties linking summations and productions. *Attention, Perception, & Psychophysics*, *73*, 872–885. http://dx.doi.org/10.3758/s13414-010-0067-5.

Stevens, S. S., Morgan, C. T., & Volkmann, J. (1941). Theory of the neural quantum in the discrimination of loudness and pitch. *The American Journal of Psychology*, *54*, 315–335. http://dx.doi.org/10.2307/1417678.

Swets, J., Tanner, W. P., Jr., & Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review*, *68*, 301–340. http://dx.doi.org/10.1037/0033-295X.68.5.301.

Van Zandt, T., & Maldonado-Molina, M. M. (2004). Response reversals in recognition memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *30*, 1147–1166. http://dx.doi.org/10.1037/0278-7393.30.6.1147.

Wickelgren, W. A. (1968). Testing two-state theories with operating characteristics and a posteriori probabilities. *Psychological Bulletin*, *69*, 126–131. http://dx.doi.org/10.1037/h0025264.

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.

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.