Likelihood ratio sequential sampling models of recognition memory

Adam F. Osth [a,*], Simon Dennis [b], Andrew Heathcote [c]

[a] University of Melbourne, Australia
[b] University of Newcastle, Australia
[c] University of Tasmania, Australia

ARTICLE INFO

ABSTRACT

The mirror effect – a phenomenon whereby a manipulation produces opposite effects on hit and false alarm rates – is benchmark regularity of recognition memory. A likelihood ratio decision process, basing recognition on the relative likelihood that a stimulus is a target or a lure, naturally predicts the mirror effect, and so has been widely adopted in quantitative models of recognition memory. Glanzer, Hilford, and Maloney (2009) demonstrated that likelihood ratio models, assuming Gaussian memory strength, are also capable of explaining regularities observed in receiver-operating characteristics (ROCs), such as greater target than lure variance. Despite its central place in theorising about recognition memory, however, this class of models has not been tested using response time (RT) distributions. In this article, we develop a linear approximation to the likelihood ratio transformation, which we show predicts the same regularities as the exact transformation. This development enabled us to develop a tractable model of recognition-memory RT based on the diffusion decision model (DDM), with inputs (drift rates) provided by an approximate likelihood ratio transformation. We compared this "LR-DDM" to a standard DDM where all targets and lures receive their own drift rate parameters. Both were implemented as hierarchical Bayesian models and applied to four datasets. Model selection taking into account parsimony favored the LR-DDM, which requires fewer parameters than the standard DDM but still fits the data well. These results support log-likelihood based models as providing an elegant explanation of the regularities of recognition memory, not only in terms of choices made but also in terms of the times it takes to make them.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

In recognition memory, participants study a list of items and during a test phase and are asked to discriminate between studied items (targets) and unstudied items (lures). Two of the most successful modeling frameworks for decision making in recognition memory are signal detection theory (SDT) and sequential sampling. In signal detection theory, different stimulus conditions are represented as continuous evidence distributions (usually Gaussian in shape), with the observer placing a criterion on the evidence axis. Models in the SDT framework are successful for accounting for the shape of the receiver operating characteristic (ROC). To construct an ROC, participants undergo recognition memory testing across a range of different bias conditions; hit rates (HR) are plotted against false alarm rates (FAR) for each bias condition. SDT models were successful in predicting the curvilinear shape of the ROC, a nearly universal finding in recognition memory (Egan, 1958; Wixted, 2007).

* Corresponding author.
  E-mail address: adamosth@gmail.com (A.F. Osth).

Nonetheless, a major weakness of SDT models is their inability to predict the shape of response time (RT) distributions: if RT is determined by the distance from the response criterion, SDT models are not able to correctly predict right skewed RT distributions for both correct and error responses (Ratcliff & McKoon, 2008).

Sequential sampling models are the most successful framework for predicting the shapes of choice RT distributions. In sequential sampling models, evidence is sampled from a stimulus until it reaches one of two response boundaries corresponding to the decision alternatives; the response boundary is the choice and the time taken to reach the boundary is the RT. We focus on the *diffusion model*, in which evidence begins to accumulate at the starting point z that is placed between two response boundaries, the upper boundary a, and the lower boundary at 0. The rate of evidence accumulation is called the *drift rate* (denoted by v): positive drift rates tend toward the upper boundary and negative drift rates tend toward the lower boundary. As the absolute value of the drift rate increases, the rate of correct responses increases and RT decreases. Errors are made because evidence accumulation is noisy; on each timestep Gaussian noise is added to the accrued evidence. To account for perceptual encoding and response output processes that are outside of the scope of evidence accumulation, a fixed constant t_{er} is added to the RT distribution to reflect nondecision time.

There were two major successes of classical diffusion models. First, in contrast to distance-from-criterion SDT models, they naturally predicted right skewed RT distributions for both correct and error responses. Second, they are able to account for the speed-accuracy tradeoff through changes in the response boundary. A decreased boundary produce faster RTs as the diffusion process has less distance to travel, but more errors result because closer response boundaries make it more likely that the diffusion process will reach the incorrect boundary by accident. Nonetheless, the classical diffusion model also has a number of weaknesses. As response boundaries increase errors disappear entirely, whereas experiments examining speed-accuracy tradeoffs find that asymptotic accuracy is usually far from perfect, especially in recognition memory (e.g., Reed, 1976). Additionally, classical diffusion models have difficulty with the relative speeds of correct and error responses. Under speeded conditions, errors are often faster than correct responses, whereas with more cautious responding errors are often slower than correct responses. Classical diffusion models, in contrast, predict equivalent RT distributions for correct and error responses under unbiased responding (Laming, 1968).

The problems with both of these modeling frameworks were solved by marrying them into a single framework, originally by Laming (1968) in discrete time, and later in continuous time in Ratcliff's diffusion decision model (DDM: Ratcliff, 1978; Ratcliff & McKoon, 2008). The DDM uses an SDT front-end for a diffusion process: drift rates for each trial are sampled relative to a drift criterion d_c from Gaussian evidence distributions with standard deviation η. Trial-to-trial variability in the drift rate ensures that there is an *asymptotic d'*; as response boundaries are increased, performance can never exceed the limit imposed by the overlap of target and lure drift rate distributions. Additionally, drift rate variability allows the model to predict error responses that are slower than correct responses (Ratcliff, 1978; Ratcliff & McKoon, 2008). A diagram of the DDM can be seen in Fig. 1.

Ironically, in the years that followed the publication of the seminal Ratcliff (1978) article, both SDT models and sequential sampling models developed largely independently of each other. A major development in SDT models of recognition memory was the rejection of the equal variance signal detection model based on investigations of the z-transformed ROC (zROC). Equal variance signal detection models predict linear zROCs with a slope of 1. However, many investigations have revealed zROC slopes of around 0.8 (Egan, 1958; Glanzer & Adams, 1990; Heathcote, 2003; Ratcliff, McKoon, & Tindall, 1994; Ratcliff, Sheu, & Gronlund, 1992), or even less when random item variability is taken into account (Averell, Prince, & Heathcote, 2016; Pratte & Rouder, 2012; Pratte, Rouder, & Morey, 2010). As a consequence, theorists have adopted the *unequal variance signal detection* (UVSD) model, which allows greater variability for targets than for lures, potentially due to the contribution of encoding variability (Wixted, 2007). Another development in SDT models, which will be described in more detail below, is the usage of *log-likelihood ratio signal detection theory* models to capture the mirror effect (Glanzer & Adams, 1985; Glanzer, Hilford, & Maloney, 2009).

The DDM was also updated with the adoption from Laming (1968) of cross-trial variability in the starting point of evidence accumulation (Ratcliff & Rouder, 1998; Ratcliff, Van Zandt, & McKoon, 1999). Rather than have the starting point fixed across trials, the starting point was sampled from a uniform distribution with range s_z. The inclusion of this variability parameter allowed for a complete account of the speed-accuracy tradeoff: the DDM with cross-trial variability in both starting point and drift rates can predict faster errors than correct responses in speeded conditions while predicting slower correct than error responses in conditions that emphasise accuracy (Ratcliff & Smith, 2004). The model was further updated with the inclusion of cross-trial variability in nondecision time (t_{er}), where nondecision time is sampled from a uniform distribution with range s_t, to allow for better predictions of the leading edge of the RT distributions across different levels of performance (Ratcliff, Gomez, & McKoon, 2004). However, the majority of DDM applications to recognition memory continued to use equal variance for targets and lures (Arnold, Broder, & Bayen, 2015; Bowen, Spaniol, Patel, & Voss, 2016; Criss, 2010; Ratcliff & Smith, 2004; Ratcliff, Thapar, & McKoon, 2004; Ratcliff, Thapar, & McKoon, 2010; Ratcliff, Thapar, & McKoon, 2011; White & Poldrack, 2014).

A re-introduction of contemporary SDT influences to the DDM came from Starns, Ratcliff, and McKoon (2012), who tested whether or estimates from the DDM were consistent with unequal variance signal detection models. They applied the DDM to binary ROC data, where an ROC was created by giving participants yes/no decisions with bias manipulated via changes in response proportions. This procedure allows for application of the DDM by manipulating the starting point along with the drift criterion across the bias conditions. Starns et al. manipulated bias using five different levels of target proportions, yield-
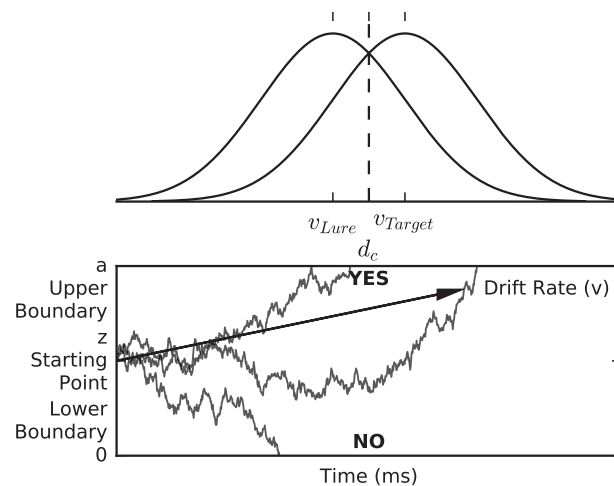
Fig. 1. Basic description of the Ratcliff diffusion model (DDM). The drift rate, v, for a given trial is a sample from a normal distribution with standard deviation η relative to a drift criterion d_c. This drift rate is then used to drive a diffusion process (bottom panel). During the diffusion process, evidence accumulates through the trial, beginning at the starting point z, and continues until either the upper response boundary (a) or the lower response boundary (0) is reached. The boundary that is reached is the choice whereas the time taken to reach the boundary is the response time (RT). Evidence accumulation is noisy, such that diffusion processes with the same drift rate will often reach different boundaries and produce different RTs. Depicted are three sample trajectories with the same drift rate.

ing a five point ROC. Simulations of the standard equal variance DDM predicted zROC slopes of 1.0 in all conditions, consistent with equal variance SDT models. When different η parameters were estimated for targets and lures, targets were found to have much higher cross-trial drift rate variability than lures and the model was able to account for the zROC slopes less than one found in the data.

Although Starns et al. (2012) used a binary ROC paradigm, there are sufficient constraints within a response time distribution to estimate drift rate variability without a response bias manipulation (Ratcliff & Tuerlinckx, 2002). Specifically, the η parameter allows for the prediction of slower error responses relative to incorrect responses. Starns and Ratcliff (2014) report DDM fits to a large number of recognition memory datasets that lacked bias manipulations, and which yielded the same conclusion as the binary ROC paradigm: greater drift rate variability for targets than for lures. Although there are deviations between the ratio of variability estimated from diffusion models fits and the slope of the zROC (Ratcliff & Starns, 2009; Starns et al., 2012), the work of Starns and Ratcliff reveals that there are qualitative consistencies between the conclusions of SDT models and the DDM.

2. Log-likelihood ratio signal detection theory

The present work aims to test whether response time distributions are consistent with log-likelihood ratio signal detection models. Log-likelihood ratio signal detection models were popularized by Glanzer and colleagues (Glanzer & Adams, 1990; Glanzer, Adams, Iverson, & Kim, 1993; Glanzer et al., 2009) due to their natural ability to capture the mirror effect. The mirror effect occurs when a manipulation produces opposite effects on the hit rate (HR) and false alarm rate (FAR). Although Glanzer and Adams (1985, 1990) identified many manipulations that produce the mirror effect, perhaps the most studied is the word frequency effect, where words of low natural language frequency exhibit higher hit rates and lower false alarm rates than high frequency words. Other factors, such as study time or repetitions, also produce a mirror effect when they are manipulated across different study lists: lists with items that were studied for longer durations, or that contain repetitions, exhibit both higher hit rates and lower false alarm rates than lists with items that were presented for shorter durations or only once (Criss, 2006; Hirshman, 1995; Osth & Dennis, 2014; Osth, Dennis, & Kinnell, 2014; Stretch & Wixted, 1998b).

The mirror effect challenged certain "fixed-strength" SDT models (e.g., Wickelgren & Norman, 1966), wherein the location of the lure distribution is fixed for all experimental conditions, with only the location of the target distribution and decision criterion varying across conditions. With these assumptions, the mirror effect could be produced if the location of the LF target distribution is higher than that of the HF target distribution, producing the elevated HR, while also exhibiting a higher decision criterion, producing the lower FAR. However, the mirror effect has been found even with two alternative forced choice (2AFC) testing, which is presumed not to involve a response criterion (Glanzer & Adams, 1990; Glanzer & Bowles, 1976).

Log-likelihood ratio signal detection (henceforth referred to as LR-SDT) models naturally predict the mirror effect. A fixed strength SDT model, as depicted in panel A of Fig. 2, contains a lure distribution that is fixed across both conditions, along with two target distributions, one corresponding to a weak condition and another corresponding to a strong condition (here and throughout the article we use the terms "weak" and "strong" agnostically - they can correspond to any manipulation
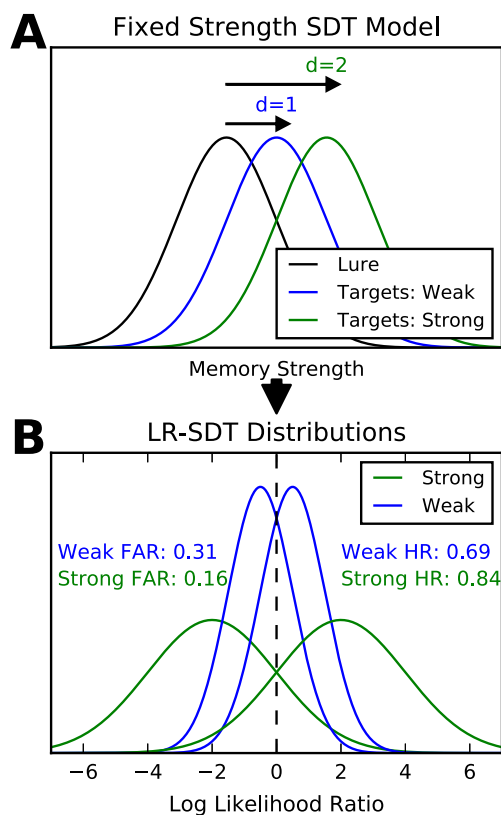
Fig. 2. (A) A fixed strength SDT model with a single lure distribution and two target distributions, one corresponding to a weak condition and one corresponding to a strong distribution. (B) Log likelihood ratio distributions for both weak and strong conditions after transforming the distributions in panel A. Predicted hit and false alarm rates are depicted for a criterion placed at 0 on the decision axis.

that affects performance). The fixed strength SDT model can be transformed to a log-likelihood ratio decision axis by taking a strength *x* and calculating the logarithm of the likelihood that the item is a target divided by the likelihood that the item is a lure; the log-likelihood ratio λ. When this is done for the entire strength distributions for all conditions, it results in mirror ordered log-likelihood ratio distributions as seen in panel B of Fig. 2. Specifically, the mean of the lure distribution for the strong condition is lower than the mean of the lure distribution for the weak condition, resulting in a lower FAR for the strong condition, and the strong target distribution is above the weak target distribution, resulting in a higher HR. Panel B shows predicted HR and FAR from each distribution after being compared to a decision criterion at 0 on the decision axis, which is the point where there is equal probability that an item is a target or a lure.

To understand how this transformation results in the mirror effect, consider testing a lure from a weak condition. The numerator of the log-likelihood ratio is relatively large given the low mean of the weak target distribution. However, if the lure is from a strong category, the numerator of the log-likelihood ratio is small because there is very low likelihood that the item came from the strong target distribution. Psychologically, the log-likelihood ratio transformation represents the idea that the strength of an item alone does not drive a decision; instead, an item's strength is compared to its expected memorability before a decision is made (e.g., Brown, Lewis, & Monk, 1977). Nearly all current quantitative models of recognition memory use likelihood ratios; these include Attention Likelihood Theory (ALT: Glanzer et al., 1993), the Retrieving Effectively from Memory (REM: Shiffrin & Steyvers, 1997) model, the Subjective Likelihood in Memory (SLiM: McClelland & Chappell, 1998) model, the Bind Cue Decide Model of Episodic Memory (BCDMEM: Dennis & Humphreys, 2001), and the model of Osth and Dennis (2015).

LR-SDT predicts more than just the mirror effect. Recently, Glanzer et al. (2009) discussed other regularities from ROC procedure that are consistent with LR-SDT models. Inspection of panel A of Fig. 2 reveals that the strong distributions have higher variability than the weaker distributions, a prediction which can be tested using ROCs. Specifically, when one plots the FAR to weak lures on the *x*-axis and the strong lures on the *y*-axis, this should yield zROC slopes less than one, revealing higher variability for the strong lures (an example can be seen in panel B of Fig. 2). Glanzer et al. (2009) reviewed dozens of ROC studies and found that this prediction was confirmed across a large range of manipulations, including study time, repetitions, study-test delay, and word frequency. This effect has been referred to as the *variance effect*.

Another prediction is that the length of a zROC should be shorter for conditions of higher performance, a prediction referred to as the *zROC length effect* (an example can be seen in panel C of Fig. 3). This prediction was first tested and con-

firmed by Stretch and Wixted (1998a). Glanzer et al. (2009) confirmed that it held for a wide range of performance manipulations. Additionally, through the 1990's a number of predictions for 2AFC recognition, such as the concentering of the underlying distributions in response to performance manipulations, were tested and confirmed (Glanzer, Adams, & Iverson, 1991; Hilford, Glanzer, & Kim, 1997; Kim & Glanzer, 1993; Kim & Glanzer, 1995, more discussion of the 2AFC paradigm can be found in the General Discussion). We know of no other modeling framework that is able to simultaneously account for all of these results.

Despite the successes of the LR-SDT modeling framework in explaining regularities in ROC functions, yes/no testing, and 2AFC testing, LR-SDT models have yet to be tested with response times, despite the strong constraints that RT distributions have been shown to impose on cognitive models (e.g., Donkin & Nosofsky, 2012; Donkin, Nosofsky, Gold, & Shiffrin, 2013; Fific, Little, & Nosofsky, 2010; Ratcliff & Murdock, 1976; Ratcliff & Smith, 2004). To this end the present investigation uses the LR-SDT framework, rather than the standard SDT framework, as a front-end for the DDM. This approach affords a considerable simplification relative to past applications of the DDM to recognition memory, where mean drift rates for targets and lures in all conditions have been freely estimated (Criss, 2010; Ratcliff et al., 2004; Ratcliff et al., 2010; Starns et al., 2012; Starns & Ratcliff, 2014).

For example, consider a dataset with a two-level word-frequency manipulation, high frequency (HF) and low frequency (LF), requiring drift rate estimates for HF lures, LF lures, HF targets, and LF targets. If all conditions are assumed to have the same drift rate standard deviation, a total of five parameters must be estimated. An equal variance LR-SDT model, in contrast, employs only two d parameters, one for the HF targets and one for the LF targets. The subsequent likelihood-ratio transformation then yields four drift rate means and four standard deviations, corresponding to the HF lure, HF target, LF lure, and LF target conditions. However, as we now discuss, the finding that lures and targets have unequal variance presents a barrier to developing this approach.

3. The unequal variance problem and an approximate linear solution

As mentioned previously, there is strong evidence for unequal variance between targets and lures from both SDT modeling (Glanzer, Kim, Hilford, & Adams, 1999; Heathcote, 2003; Ratcliff et al., 1992; Wixted, 2007) and RT models (Ratcliff & Starns, 2009; Starns & Ratcliff, 2014; Starns et al., 2012). This makes it necessary to employ unequal variance between targets and lures before performing the likelihood ratio transformation. However, with Gaussian distributions, this presents a problem, as there are two points of intersection between the target and lure distributions, meaning that there are two values of x with the same likelihood ratio. Glanzer et al. (2009) demonstrated this mathematically by showing that the transfer function between the strength axis and the log-likelihood ratio axis is both non-linear and non-monotonic in the unequal variance case. The equal variance case, in contrast, has a linear transfer function. Given that a linear transformation of a Gaussian distribution is still a Gaussian, the log-likelihood transformation for the equal variance case preserves the assumed Gaussian shape of memory strength distribution.
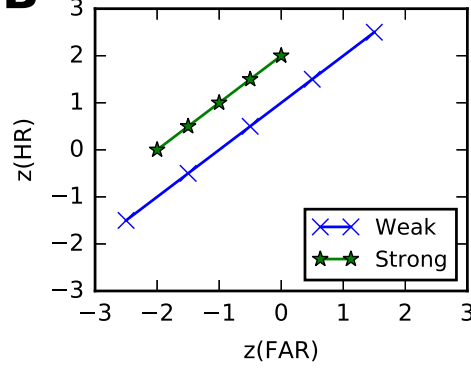
A demonstration of the likelihood transformation for three different standard deviations of the target distribution (which we will denote as S; without loss of generality we set the standard deviation of the lure distribution to 1) can be seen in Fig. 4 for $d = 1.5$. The top panels show the strength distributions, whereas the middle panels show the resulting log-likelihood ratio distributions. The left panels show the case where S is 1.1. One can see that for this value, the transformation still results in log-likelihood ratio distributions that are approximately normal in shape. However, as S is increased to 1.25 and 1.5, the distributions become more distorted and do not resemble Gaussian distributions. Specifically, these distributions are non-central chi square in shape. The bottom panel in Fig. 4 shows the transfer functions for each value of S, including the equal variance case ($S = 1$). One can see that as S is increased, the transfer functions increasingly deviate from linearity.

The nonlinear transfer function is problematic because the DDM assumes that drift rate distributions are Gaussian. First passage time expressions for the DDM require integration over the cross-trial variability in drift rate (Ratcliff, 1978). Integration can be quite time-intensive for the non-central chisquare distribution, which lacks an analytic solution for its cumulative distribution function (CDF), and to our knowledge, does not possess fast approximations like with the normal distribution. In contrast, an analytic solution is available to the DDM integral when drift rate variability is Gaussian (we used the fast-dm-30 C code in our computations, see Voss, Voss, & Lerche, 2015). For this reason, we developed a linear approximation to the transfer function via a first-order Taylor series expansion, which results in approximate log-likelihood distributions that are Gaussian in shape.

When constructing the linear approximation, we must choose which point the approximation intersects with the quadratic transfer function of the unequal variance model. There were two plausible points we considered: the point at which the target and lure distributions intersect, which we will refer to as the *ideal observer* approximation model, and a simpler model which intersects at the point $d/2$, which is the point at which the two distributions intersect when there is equal variance. Simulations and fits of the two models produced nearly identical results. We have chosen to use the simpler $d/2$ approximation in this article, as it reduces to the equal variance likelihood ratio model when $S = 1$. The equation for the ideal observer approximation model, in contrast, has a singularity when $S = 1$. We were also able to prove for the $d/2$ case that the variance and zROC length effect hold for all values of d and S. The mirror effect, as we will demonstrate, is somewhat more complex. Mirror ordering of the means of the distributions ($\mu_{SL} < \mu_{WL} < \mu_{WT} < \mu_{ST}$) is guaranteed for all values of d and S, whereas the mirror effect as observed in the HR and FAR is not.
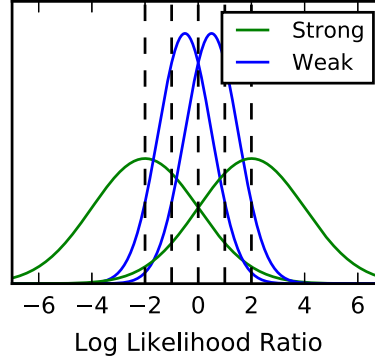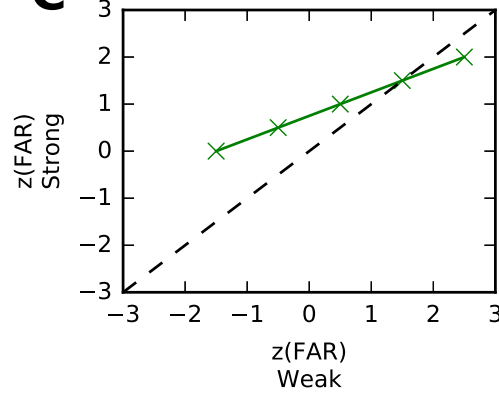
Fig. 3. Demonstrations of the ROC regularities from the log-likelihood ratio transformation. (A) Log-likelihood ratio distributions for weak and strong conditions along with five criteria. (B) Old-new zROCs for the weak and strong conditions; note that the zROC for the strong condition is shorter. (C) New-new zROC with the weak condition on the *x* axis and the strong condition on the *y* axis. The dashed line has a slope of one for reference.

Full derivations of the Taylor series expansion can be seen in the Appendix. The approximation results in log-likelihood ratio distributions with expectation and standard deviation for targets (T) and lures (L) are as follows:

$$\mu_{\lambda L} = -\left(\left(\frac{d^2}{2}\right)\left(\frac{S^2+3}{4S^2}\right) + \log(S)\right) \tag{1}$$

$$\mu_{\lambda T} = d^2 \frac{S^2+1}{2S^2} + \mu_{\lambda L} \tag{2}$$

$$\sigma_{\lambda L} = d \frac{S^2+1}{2S^2} \tag{3}$$

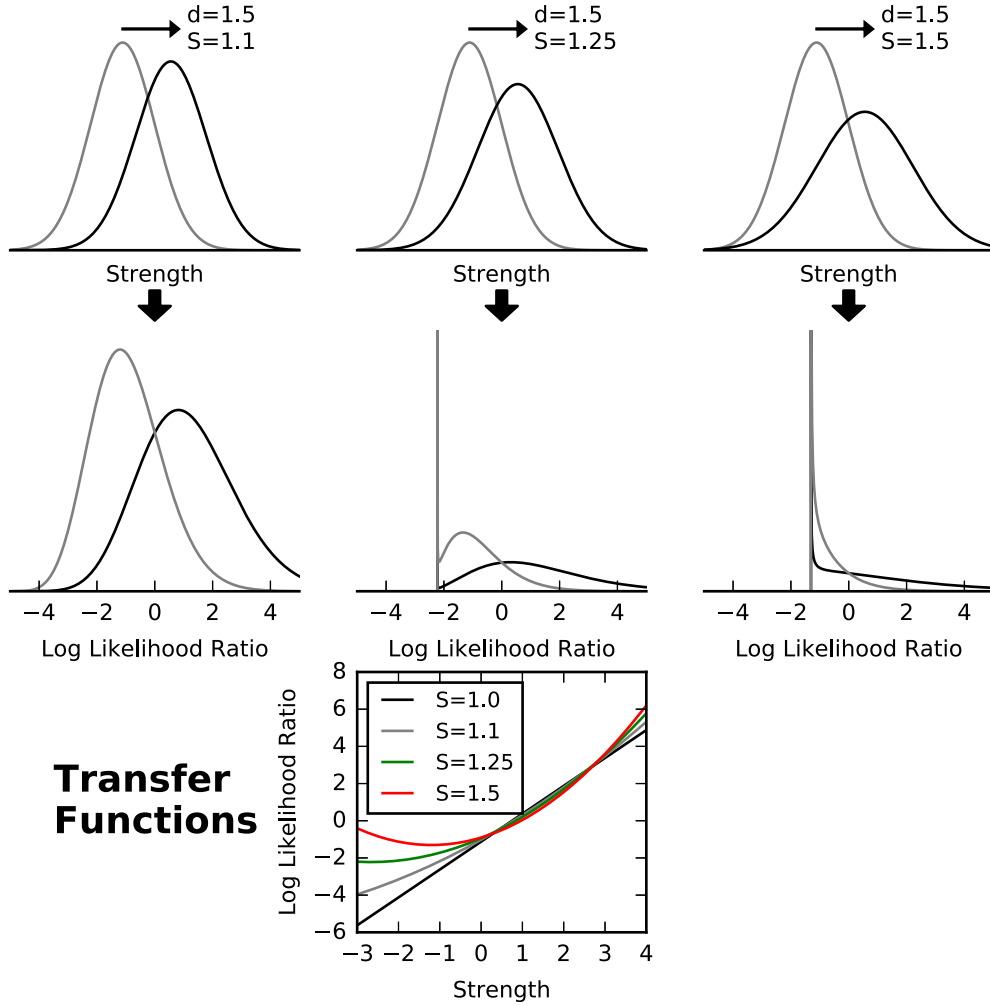$$\sigma_{\lambda T} = S\sigma_{\lambda L} \tag{4}$$

Fig. 4. Top panels: Strength distributions with $d = 1.5$ for $S = 1.1$ (left), $S = 1.25$ (middle), and $S = 1.5$ (right). Middle panels: log-likelihood ratio distributions after the transformation from the strength distributions in the top panels. Bottom panel: Transfer functions between strength and log-likelihood ratio axes for $S = 1.0, 1.1, 1.25, 1.50$.

The variance effect follows directly from Eqs. (3) and (4). Variances increase monotonically with d. Mirroring of the means of the LR distributions can be seen by taking the derivative of (2) and (1) with respect to d:

$$\frac{\partial \mu_{\lambda T}}{\partial d} = \frac{3S^2}{4} \tag{5}$$

$$\frac{\partial \mu_{\lambda L}}{\partial d} = -\frac{d(S^2 + 3)}{4} \tag{6}$$

Eq. (5) is positive and Eq. (6) is negative. This means that as d is increased, it is accompanied by increases in $\mu_{\lambda T}$ and decreases in $\mu_{\lambda L}$.

Note that this guarantees a mirror ordering of the means of the log LR distributions, but does not guarantee a mirror effect in the HR and FAR. Given that the variances increase with the means, one can imagine situations where the increase in variance is larger than the increase in the means of the distribution, which would not produce a mirror effect in the HR and FAR. We consider the HR and FAR for a criterion of zero below:

$$HR(\mu_{\lambda T}, \sigma_{\lambda T}, 0) = 1 - \Phi(\mu_{\lambda T}, \sigma_{\lambda T}, 0) \tag{7}$$

$$FAR(\mu_{\lambda L}, \sigma_{\lambda L}, 0) = 1 - \Phi(\mu_{\lambda L}, \sigma_{\lambda L}, 0) \tag{8}$$

Where Φ is the cumulative normal distribution function. We can simplify the above by rescaling by $1/\sigma$:

$$HR(\mu_{\lambda T}, \sigma_{\lambda T}, 0) = 1 - \Phi(\mu_{\lambda T}/\sigma_{\lambda T}, 1, 0) \tag{9}$$

$$FAR(\mu_{\lambda L}, \sigma_{\lambda L}, 0) = 1 - \Phi(\mu_{\lambda L}/\sigma_{\lambda L}, 1, 0) \tag{10}$$

Because Φ is monotonic decreasing with respect to the mean, when $\mu_{\lambda T}/\sigma_{\lambda T}$ rises the HR rises. Similarly, when $\mu_{\lambda L}/\sigma_{\lambda L}$ falls, the FAR decreases. It is sufficient then to examine $\mu_{\lambda T}/\sigma_{\lambda T}$ and $\mu_{\lambda L}/\sigma_{\lambda L}$ to assess if the mirror effect is occurring.

Considering the FAR portion first we are interested in the derivative:

$$\frac{\partial \mu_{\lambda L}/\sigma_{\lambda L}}{\partial d} = -(S^2 + 3)/(4(S^2 + 1)) + 2\log(S)S^2/(S^2 + 1)d^2 \tag{11}$$

For a mirror effect to occur this must be less than 0:

$$\frac{\partial \mu_{\lambda L}/\sigma_{\lambda L}}{\partial d} < 0 \tag{12}$$

Which occurs when

$$\frac{\sqrt{8\log(S)S^2}}{S^2 + 3} < d \tag{13}$$

For the equal variance case $S = 1$, this is true for all d. When $S > 1$, the mirror effect can be broken for some small values of d ($d < 1$). Fig. 6 shows when this occurs.

Considering the hit rate portion we are interested in the derivative:

$$\frac{\partial \mu_{\lambda T}/\sigma_{\lambda T}}{\partial d} = 1/S - (S^2 + 3)/(4S(S^2 + 1)) + 2S\log(S)S/((S^2 + 1)d^2) \tag{14}$$

For a mirror effect to occur this must be greater than 0:

$$\frac{\partial \mu_{\lambda T}/\sigma_{\lambda T}}{\partial d} > 0 \tag{15}$$

Which occurs when

$$\frac{8S^2\log(S)}{-4S^3 + S^2 - 4S + 3} < d^2 \tag{16}$$

Because the numerator is always positive and the denominator is always negative when $S >= 1$ this will be true in all cases where $S >= 1$. Note that this is not the case when $S < 1$, and Fig. 6 shows that for those cases the HR portion of the mirror effect can be broken.

Following Glanzer et al. (2009), the zROC length effect can be demonstrated by constructing an inverse transfer function from Eq. (44):

$$x(\lambda) = \frac{\lambda + \frac{d^2}{2} - \log(S)}{d\alpha} \tag{17}$$

Using Eq. (17), we can then compare two criteria, λ_{max} and λ_{min} to refer to the strictest and most lenient criteria, respectively. The length of the zROC, Λ, is:

$$\Lambda = \frac{\lambda_{max} - \lambda_{min}}{d\alpha} \tag{18}$$

Eq. (18) reveals that Λ is inversely related to d.

Fig. 5 compares the approximation against the true likelihood ratio transformation where the log-likelihood ratio transformation distributions are non-central chisquare (top panels of Fig. 5). The transfer functions are shown in the middle panels. One can see that as S is increased, the approximation of the true log-likelihood ratio distributions is poorer. Nonetheless, inspection of the ROC predictions in the bottom panels of Fig. 5 reveals that both models fall on the same ROC function, with the difference being that the approximation's ROC points are shifted to the left from the true distribution.

Inspection of Eqs. (3) and (4) also reveals that the expressions for σ for both distributions is the same with the exception that $\sigma_{\lambda target}$ is scaled by S. This means that the relative variability between the target and lure distributions is the same for all values of d and exactly equals the relative variability of the underlying memory strength distributions.

Overall, these results indicate that the approximate likelihood transformation preserves the properties of the exact likelihood transformation in the unequal variance case that explain many benchmark recognition memory phenomena. We now explore whether its combination with the DDM can also accommodate RT distribution effects in recognition memory.

4. Model selection

We compare two variants of the DDM: the conventional parameterization where each target and lure condition receives its own drift rate parameter (which will henceforth be referred to as the "standard DDM"), along with a version of the DDM where the drift rates are specified by the $d/2$ approximation of the log-likelihood ratio transformation (which will henceforth be referred to as the "LR-DDM"). The standard DDM model employs four parameters to capture the mirror effect: mean
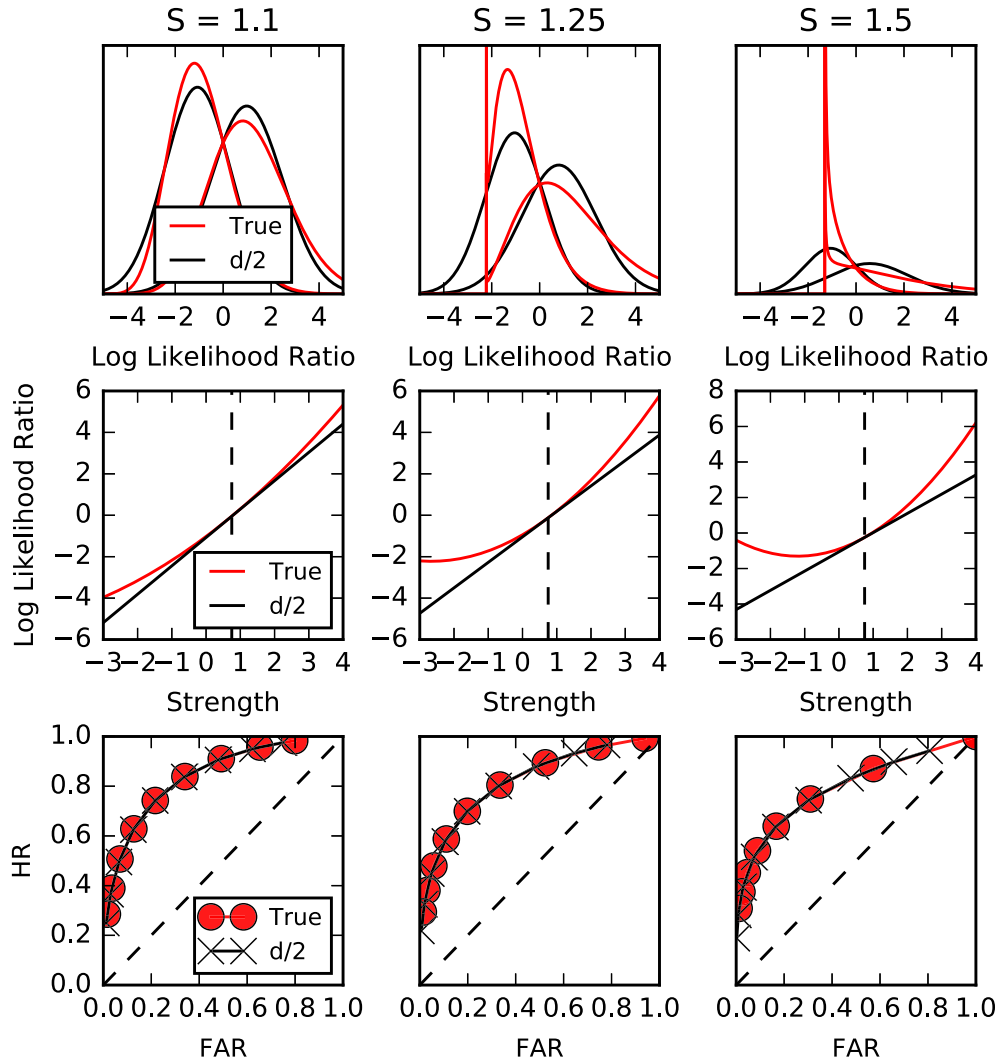
Fig. 5. Top panels: True log-likelihood ratio distributions along with the $d/2$ approximation with $d = 1.5$ and $S = 1.1$ (left), $S = 1.25$ (middle), and $S = 1.5$ (right). Middle panels: Transfer functions for both the true model along with the $d/2$ approximation for each value of S. Vertical dashed lines are placed at $d/2$. Bottom panel: ROC predictions for the true and $d/2$ approximation model for each value of S.

drift rates, v_{WL}, v_{WT}, v_{SL}, and v_{ST}, where W and S refer to the weak and strong conditions and T and L refer to targets and lures, in addition to two drift rate standard deviation (η) parameters for targets and lures if unequal variance is allowed, or a single η parameter for both distributions if the traditional equal variance version is employed. The unequal variance model is the one that was endorsed by Starns and Ratcliff (2014). We did not allow for the η parameters to vary across factors other than whether an item is a target or lure because Starns and Ratcliff (2014) found that η did not vary across different strength conditions.

In the LR-DDM, three parameters corresponding to the fixed strength SDT model are estimated, namely, the means of the target distributions (d_{weak} and d_{strong}) along with the standard deviation of the target distributions, S. These parameters are transformed by the $d/2$ approximation to the log-likelihood ratio into drift rate distribution parameters v and η for targets and lures in the weak and strong conditions according to Eqs. (1)–(4). We also compare the performance of this model to an equal variance version where S is fixed to 1, in which the approximation is exact. A diagram of the unequal variance version of the LR-DDM can be seen in Fig. 7.

There are several things to note about the standard DDM. First, the LR-DDM is a special case of the standard DDM, so the standard DDM always fits data at least as well as the LR-DDM. Second, it is the model that is conventionally used in fits to recognition memory data (e.g.: Criss, 2010; Starns, 2014; Starns & Ratcliff, 2014; Starns et al., 2012). Third, although the standard DDM can account for the mirror effect by allowing mirror ordered drift rates ($v_{SL} < v_{WL} < v_{WT} < v_{ST}$), it does not *predict* the mirror effect a priori the way the log-likelihood ratio model does. It is for this reason that the standard DDM is more flexible than the likelihood ratio variant: violations of the mirror effect can be easily accommodated. Consider a "broken"
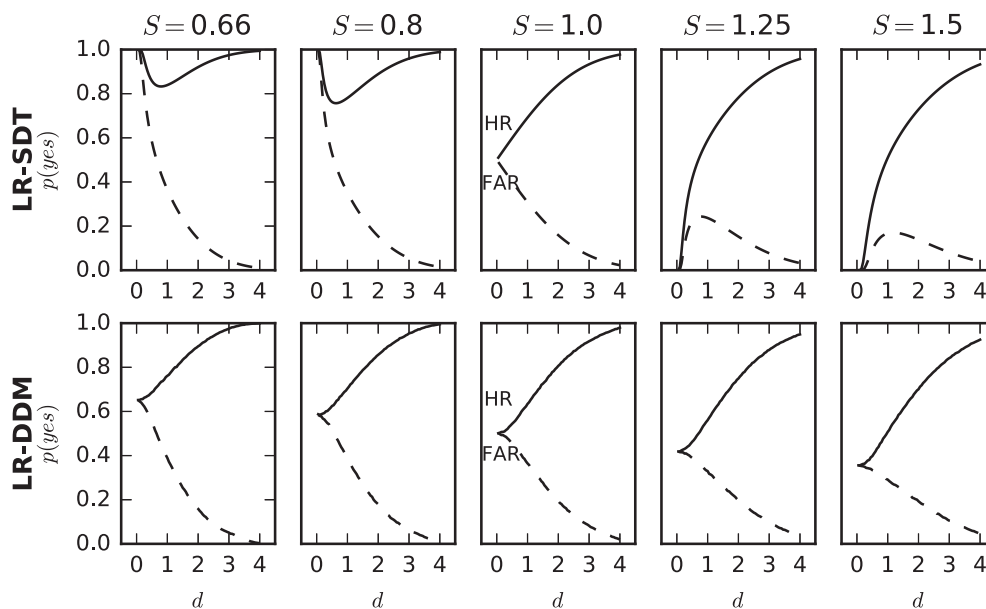
Fig. 6. Effects of the S parameter on HR (solid lines) and FAR (dashed lines) for the LR-SDT (top row) and LR-DDM (bottom row) models for various values of d. DDM parameters: a = 1.5, s_z/a = .05, z/a = .5.

mirror effect where LF words perform better than HF words, but HF words have lower FAR, which is the opposite of the usual finding. The standard DDM can accommodate this pattern by allowing a higher drift rate for LF lures, whereas the likelihood ratio model would be unable to account for the data with a single value of S.

4.1. Effects of the S parameter: symmetry of the mirror effect

Historically, one of the principle reasons for inclusion of drift rate variability was that it allowed the diffusion model to predict slower errors than correct responses. Somewhat counter-intuitively, for a fixed drift rate and with a starting point that is equidistant from both response boundaries errors and correct responses are predicted to have the same RT distribution (although they may occur with different probabilities). Ratcliff and McKoon (2008) illustrated the mechanism by which slow errors are produced by considering the weighted mean response time: when drift rate variability is included, most of the error responses come from low drift rates, which are slow. This over representation of slow responses among errors means that correct responses are faster. This tendency is especially obvious for conditions encouraging wide boundaries (i.e., cautious responding), because then errors are most likely to stem from drift rate variability rather than other sources, such as within-trial noise or starting point variability. Under such conditions, at least when accuracy away from ceiling, the difference between error and correct RT can provide sufficient constraint to estimate the level of rate variability. However, rate variability is intrinsically hard to estimate, and so getting a precise measurement of the difference between target and lure rate variability can be challenging.

For the LR-DDM, drift rate variability, and hence the S parameter, can be identified using another aspect of the data, the symmetry in the mirror effect. Predictions for HR and FAR from the LR-DDM model can be seen with various values of d in the bottom row of Fig. 6. The equal variance model ($S = 1.0$) produces a perfectly symmetrical mirror effect: as d is increased, both the HR and FAR diverge from .5 at the same rate. However, when there is unequal variance, HR and FAR change at different rates. When targets have greater variability ($S > 1.0$), the HRs increases at a much greater rate than the FAR decreases. Conversely, when lures have greater variability than targets ($S < 1.0$), the FAR changes at a much greater rate than the HR. One may also note that in contrast to LR-SDT, the LR-DDM does not exhibit any deviations from the mirror effect for conventional values of S. This is likely because the variability from the within-trial noise of the DDM is sufficient to wash out some of the effect of the increasing variability with increases in d.

However, it should be noted that these predictions only hold for when S is constant across the different values of d. With regard to the word frequency mirror effect, there have been cases where the mirror effect is "broken" and HRs are equivalent for LF and HF words while the FAR portion is intact (higher FAR for HF words: Criss & Shiffrin, 2004; Hirshman & Arndt, 1997; Joordens & Hockley, 2000). Although this might suggest a model where $S < 1.0$, that would fail to predict the slope of the old-new zROC; to our knowledge there has never been an observed zROC with a slope suggesting higher variability for lures than for targets in long-term recognition. However, several studies have found results suggesting higher variability for LF words than HF words (Glanzer & Adams, 1990; Glanzer et al., 1999; Ratcliff et al., 1994). We conducted simulations and confirmed that under some parameter combinations, increasing S with d can produce large differences in the FAR with little change or
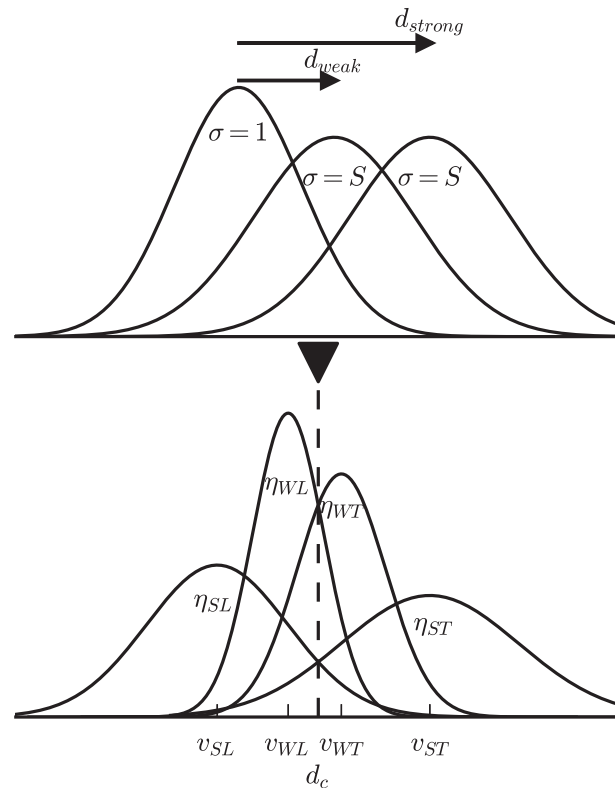
Fig. 7. Diagram depicting the drift rate parameterization for the LR-DDM, which has three free parameters corresponding to the fixed strength SDT model: the means of the target distributions d_{weak} and d_{strong}, along with the σ of the target distribution S. These parameters undergo the approximate $d/2$ likelihood ratio transformation to arrive at v and η parameters corresponding to the SL, WL, WT, and ST conditions, where W and S refer to the weak and strong conditions and T and L refer to targets and lures. There are also variants which include a drift criterion (d_c), which shifts the locations of the means.

even a reversal in the HR pattern. In this article, we only pursue models where S is constant across the values of d. This was sufficient for the data considered, which did not display broken mirror effects. The issue of whether a positively correlated change in S with d is sufficient to enable the LR-DDM to fit broken mirror effects is left to future research.

4.2. The drift criterion

The DDM has two sources of bias: the start point of accumulation z and the drift criterion d_c. The drift criterion is analogous to the decision criterion in SDT. It is subtracted from the input to the DDM so that the mean drift rates are positive for conditions with a correct response on the upper boundary and negative for conditions with a correct response on the lower boundary. This parameter cannot be identified if mean drift rates for both targets and lures are estimated without the addition of a selective influence assumption. One could alternatively follow conventional SDT parameterizations and fix the mean of the lure distribution and parameterize the drift criterion across different conditions instead; this is equivalent to the standard DDM model, in which means of the lure distribution are allowed to vary and the drift criterion is fixed. In recognition memory there have been attempts to selectively influence the drift criterion across different bias conditions, where bias is manipulated by varying the proportions of targets and lures. The underlying input is assumed not vary with across the bias conditions, while both the drift criterion and start point are free to vary. These investigations have found that the drift criterion largely invariant across manipulations such as target proportion, with most of the bias manipulation being addressed by variation in the start point (Criss, 2010; Ratcliff & McKoon, 2008; Starns et al., 2012).

Shifts in the drift criterion have been found to be associated with manipulations of a standard relative to which decisions are made. For instance, in a two choice numerosity detection task, where participants are asked whether a presented number of asterisks is greater or lower than a given standard number, Leite and Ratcliff (2011) found that manipulation of the standard produced changes in the drift criterion. However, just as in recognition memory studies, manipulations of response proportions produced changes in only the start point. White and Poldrack (2014) generalized these results to both recognition memory and line length judgments, where in both tasks manipulations of the criterial amount of evidence for responding produced changes in the drift criterion. They noted that the empirical signature for these two different types of bias lies in different effects on the RT distributions: changes in z produce relatively large changes in the leading edge (i.e., the fastest

responses), whereas changes in the drift criterion produce little change in the leading edge, but large changes in the tail of the RT distributions (i.e., skewing, see also Ratcliff & McKoon, 2008).

The necessity of assuming selective influence to estimate the drift criterion is restricted to the conventional parameterizations of the DDM, and do not apply to the LR-DDM. For the LR-DDM, target and lure drift rates are not independently estimated, but are instead derived from a transformation of the d and S parameters. For an LR-DDM model that lacks a drift criterion, bias can only contribute through the start point z. To understand how that can be an overly restrictive assumption, consider a participant that is very liberal in their responding, producing high HR and FAR. If only the z parameter is allowed to vary, it captures this bias with a high value of z that is biased toward the "YES" boundary, which produces leading edges for YES responses that are much faster than those for NO responses. However, this bias can also be implemented with a low value of the drift criterion, which increases drift rates for both targets and lures, leaving the leading edges relatively equivalent across YES and NO responses, and instead produces differences in the tails of the RT distributions. Because large individual differences in response bias have been identified using SDT analyses (Aminoff et al., 2012; Kantner & Lindsay, 2012), we included in our model comparison variants of the LR-DDM that include a drift criterion parameter.

4.3. Benchmark datasets

To test the LR-DDM against the standard DDM we sought recognition memory datasets which included multiple conditions for lures. Datasets where items were presented for different durations within the same study list would not suffice as there would only be a single condition for lures at test. In addition, we sought datasets where there was a reasonable amount of data per participant (around 50 trials per condition or greater), as the DDM is notorious for requiring a lot of data per participant to constrain model parameters (Wagenmakers, 2009). To provide a comprehensive test we fit datasets that displayed both the item mirror effect, through a manipulation of word frequency (Rae, Heathcote, Donkin, Averell, & Brown, 2014), and a list-strength mirror effect, through varying the number of study item presentations (five experiments from Starns, Ratcliff, & White, 2012). We also fit a dataset from Criss (2010) that included both a repetition and word frequency manipulation. In both the Starns et al. (2012) and Criss (2010) datasets, the number of item presentations was manipulated across different study lists, such that each list yielded a different lure condition. Next, we describe the experiments in more detail, along with the associated parameterization for each model. Table 1 summarizes the datasets and Table 2 the parameterizaitons.

4.4. Dataset 1: Rae et al. (2014)

The first dataset, reported by Rae et al. (2014), manipulated of both word frequency and speed-accuracy emphasis. The word frequency manipulation produced a mirror effect: low frequency words exhibited greater hit rates and lower false alarm rates than high frequency words. It is for this reason that we denote LF words as the "strong" condition and HF words as the "weak" condition. This dataset was also ideal for our purposes because it had a relatively large number of participants ($N = 47$) performing a relatively large number of trials (768 per participant).

Traditional accounts of the speed/accuracy tradeoff in the DDM only manipulate the response caution parameter a (e.g.: Ratcliff et al., 1999; Wagenmakers, Ratcliff, Gomez, & McKoon, 2008). Rae et al. (2014) found that assumption was overly restrictive in accounting for the speed/accuracy emphasis; the models also required changes in drift rate to properly capture the accuracy difference between the two conditions. Additionally, their analysis favored models where several other parameters changed across speed/accuracy emphasis conditions, including t_{er}, s_t, z, and s_z. We followed their preferred model and allowed all of these parameters to vary across the speed and accuracy emphasis conditions. The only parameters that were kept constant across conditions were those that reflected drift rate variability (η and S).

4.5. Datasets 2 and 3: Starns et al. (2012)

We additionally employed five datasets from Starns et al. (2012); these datasets were all of the within-subjects datasets in their article. In their experiments, item strength was manipulated by repetitions across lists. In two of these experiments, participants either studied lists where all items were presented once (1X: pure weak) or all items were presented five times (5X: pure strong). In three of the experiments, participants studied mixed lists of 1X and 5X items but were only tested on one category of items (1X or 5X, referred to as pure weak test lists or pure strong test lists), where participants were informed which strength category of item they were tested on. Given that the results of the experiments were quite similar to each other, we pooled the participants from each of these experiments into two datasets: one corresponding to the pure study lists (Dataset 2) and another corresponding to the mixed study lists (Dataset 3).

Starns et al. (2012) found their manipulation was inconsequential, in that both pure and mixed study lists produced the same results: tests on 5X items produced higher HR and lower FAR. They applied the standard DDM and found lower drift rates for lures in the strong conditions relative to the weak conditions. Consistent with the analyses of Criss (2010), the repetition manipulation was most evident in the drift rates. In our fits to these datasets, only drift rates were allowed to vary across the repetition conditions and all other parameters were kept constant.

Table 1
Datasets fit in this article.

Dataset	Exp.	N	Trials	Manip.
1	Rae et al. (2014)	47	768	WF (2 levels), SA (2 levels)
2	SRW E1&3 – Pure	38	216.6	Rep. (2 levels)
3	SRW E1,2&3 – Mixed	60	206.6	Rep. (2 levels)
4	Criss E2	16	1524.9	Rep. (2 levels), WF (2 levels)

Notes: N = number of participants, Trials = mean number of trials per participant in each design cell, Manip. = manipulations used in the experiment, WF = word frequency, SA = speed/accuracy emphasis, E = experiment, Pure = pure list composition, Mixed = mixed list composition, Rep. = repetitions., SRW = Starns, Ratcliff, & White (2012).

Table 2
Model parameters.

Model	UVSD	DC	Parameters	N
			Dataset 1	
Standard DDM			$z/a \sim$ E; $sz/a \sim$ E; $a \sim$ E; $t_{er} \sim$ E; $s_t \sim$ E; $v \sim$ E, F, T; η	19
Standard DDM	✔		$z/a \sim$ E; $sz/a \sim$ E; $a \sim$ E; $t_{er} \sim$ E; $s_t \sim$ E; $v \sim$ E, F, T; η; S	20
LR-DDM			$z/a \sim$ E; $sz/a \sim$ E; $a \sim$ E; $t_{er} \sim$ E; $s_t \sim$ E; $d \sim$ E, F	14
LR-DDM		✔	$z/a \sim$ E; $sz/a \sim$ E; $a \sim$ E; $t_{er} \sim$ E; $s_t \sim$ E; $d \sim$ E, F; d_c	15
LR-DDM	✔		$z/a \sim$ E; $sz/a \sim$ E; $a \sim$ E; $t_{er} \sim$ E; $s_t \sim$ E; $d \sim$ E, F; S	15
LR-DDM	✔	✔	$z/a \sim$ E; $sz/a \sim$ E; $a \sim$ E; $t_{er} \sim$ E; $s_t \sim$ E; $d \sim$ E, F; S; d_c	16
			Datasets 2 and 3	
Standard DDM			z/a; sz/a; a; t_{er}; s_t; $v \sim$ R,T; η	10
Standard DDM	✔		z/a; sz/a; a; t_{er}; s_t; $v \sim$ R,T; η; S	11
LR-DDM			z/a; sz/a; a; t_{er}; s_t; $d \sim$ R	7
LR-DDM		✔	z/a; sz/a; a; t_{er}; s_t; $d \sim$ R; d_c	8
LR-DDM	✔		z/a; sz/a; a; t_{er}; s_t; $d \sim$ R; S	8
LR-DDM	✔	✔	z/a; sz/a; a; t_{er}; s_t; $d \sim$ R; S; d_c	9
			Dataset 4	
Standard DDM			z/a; sz/a; a; t_{er}; s_t; $v \sim$ F, R, T; η	14
Standard DDM	✔		z/a; sz/a; a; t_{er}; s_t; $v \sim$ F, R, T; η; S	15
LR-DDM			z/a; sz/a; a; t_{er}; s_t; $d \sim$ F, R	9
LR-DDM		✔	z/a; sz/a; a; t_{er}; s_t; $d \sim$ F, R; d_c	10
LR-DDM	✔		z/a; sz/a; a; t_{er}; s_t; $d \sim$ F, R; S	10
LR-DDM	✔	✔	z/a; sz/a; a; t_{er}; s_t; $d \sim$ F, R; S; d_c	11

Notes: UVSD = variant with unequal variance, DC = variant with a drift criterion, *N* = number of parameters per participant, "$\sim$" = varies as a function of, E = speed/accuracy emphasis, F = word frequency, R = repetitions, T = item type (target vs. lure).

4.6. Dataset 4: Criss (2010)

We fit data from Experiment 2 of Criss (2010), which presented participants with four types of study lists: HF words presented once, LF words presented once, HF words presented 5X, and LF words presented 5X. Criss (2010) found reduced FAR for lures in 5X presented lists. Model selection revealed that this effect was best captured by a change only in drift rates and not in the starting point *z*. Following her example, only drift rates were allowed to vary across each of the four strength conditions, while all other parameters were kept constant.

4.7. Hierarchical Bayesian models

Model parameters for both models were estimated using hierarchical Bayesian analysis. Although space constraints do not afford a thorough introduction to hierarchical Bayesian analysis (but see Lee, 2011; Lee & Wagenmakers, 2014; Shiffrin, Lee, Kim, & Wagenmakers, 2008), we would like to highlight three distinct advantages over conventional methods. First, it allows for estimates of both group and participant level parameters, which is advantageous because fitting to group data via averaging and fitting individual participant data can produce different results in cognitive models (Brown & Heathcote, 2003; Estes & Maddox, 2005; Heathcote, Brown, & Mewhort, 2000). Second, Bayesian methods allow for the quantification of uncertainty in the parameter estimates as posterior distributions. Third, hierarchical methods are advantageous in estimating data from individual participants when there are not a large number of trials per participant; this is because in hierarchical models estimates of the participant level parameters are influenced by the group level parameters. When there is a large degree of uncertainty in the individual parameter estimates, these parameters get pulled toward the group estimate, a phenomenon referred to as "shrinkage". This is advantageous in fitting the DDM, which canonically requires large amounts of data to reliably estimate the parameters of the model (Wagenmakers, 2009). The advantages of hierarchical modeling of the DDM were also detailed by Vandekerckhove, Tuerlinckx, and Lee (2011).

Bayesian estimation for complex models requires Markov chain Monte Carlo (MCMC) algorithms to estimate the posterior distributions. However, in models such as the DDM, parameter estimates are correlated with each other (Ratcliff & Tuerlinckx, 2002), which is problematic for conventional sampling algorithms. For this reason, we used differential evolution Markov chain Monte Carlo (DE-MCMC: Turner, Sederberg, Brown, & Steyvers, 2013), a method of posterior sampling that is robust to parameter correlations. With DE-MCMC, on each MCMC iteration a proposal for a vector of parameters is generated by randomly sampling two other parameter vectors from other MCMC chains, taking a scaled difference between those parameter vectors, and then adding that scaled difference to the current parameter vector. This solves the problem of parameter correlation because proposals are generated by sampling from a correlated surface of accepted proposals. We encourage interested readers to consult the Turner et al. (2013) article for a detailed and technical description of this procedure.

A challenge in implementing fully hierarchical models is that variance parameters can get stuck near zero during MCMC sampling, preventing convergence from occurring. This can be especially problematic for the sz/a and η parameters of the DDM, which only weakly influence the data and tend to be quite similar across participants in a hierarchical fit. We were able to implement fully hierarchical models that circumvented this issue by using a technique where the group level parameters were randomly shuffled across chains when the subject level likelihoods are calculated, making it such that participant level parameters were evaluated under group level parameters from different chains on each iteration.[1]

As the LR-DDM is a special case of the standard DDM we would expect it to fit the data better than the relatively constrained LR-DDM. For this reason that we use a model selection technique, the deviance information criterion (DIC: Spiegelhalter, Best, Carlin, & van der Linde, 2002), that balances a model's ability to capture the data against its complexity; the winning model provides the most parsimonious yet accurate account of the data (Myung, 2000). As we wished to take into account complexity associated with not only the DDM parameters for each participant (θ), but also the parameters of the population model (ϕ) from which they were drawn, the version of DIC we used included both the probability of the hyper-parameters – $p(\theta|\phi)$ – as well as the probability of the DDM parameters – $p(data|\theta)$ – in the calculation of both the deviance and penalty term.

A preference for the LR-DDM in model selection is not a foregone conclusion. Three factors could potentially produce advantages for the standard DDM. Although mirror effects were reported in each of the datasets, the original papers did not report analyses of individual differences. Consider if half of participants showed a pattern where weak HR < strong HR but weak FAR = strong FAR and the other half showed pattern where weak HR = strong HR and weak FAR > strong FAR; averaging together these two groups of participants would reveal a mirror effect even though individual participants did not show a mirror pattern. Because our hierarchical fitting method takes account of individual variation, it will be sensitive to participants that do not exhibit the mirror effect, and so will indicate a better fit by the standard DDM.

Second, the S parameter in the LR-DDM can produce asymmetries in the mirror effect when S is different from 1.0. When $S > 1.0$, it produces a pattern where HR changes at a greater rate than the FAR, whereas when $S < 1.0$ the opposite pattern arises. The S parameter is also constrained by the response time distributions in the same way that the standard DDM is constrained. Thus, if there was an asymmetry in the mirror effect where FAR changed at a much greater rate than the HR as strength was increased, the LR-DDM could accommodate this with a value of $S < 1.0$, but this would compromise its ability to capture RT distributions relative to the standard DDM, which would likely show a signature of greater drift rate variability for targets. Finally, Eqs. (3) and (4) show the LR-DDM produces increasing drift rate variability as d is increased. However, Starns and Ratcliff (2014) did not find any differences in η between strength conditions for targets. If high values of η in strong conditions are detrimental to the model's ability to fit the data, the standard DDM will be preferred.

DDM likelihoods were estimated using the partial differential equation method of Voss and Voss (2008) and a combination of analytic and numerical integration methods of accounting for parameter variability as implemented in fast-dm-30 (Voss et al., 2015). Models were fit to all of the response times rather than quantile summaries of the response time distributions. Very fast (< 200 ms) and very slow (> 2500 ms) outlier response times which were removed from the analyses. This resulted in no more than 1% of the data being omitted from each dataset.

Model parameters z and s_z were treated as proportions of the boundary a for stability in parameter estimation. We refer to these parameters as z/a and sz/a. We additionally estimated η_{target} in the standard DDM model as the relative variability of the target distribution S, which is $\eta_{target}/\eta_{lure}$. This was done for congruence with the LR-DDM model. Parameters for each dataset can be seen in Table 2.

Each of the models share the parameters $z/a, s_z/a, a, t_{er}$, and s_t. These parameters were treated as samples from group level distributions with mean M and standard deviation ς. Because these parameters are also bounded, we treat samples for each participant as coming from truncated normal distributions (denoted as TN):

$$z/a \sim TN(M_z, \varsigma_z, 0, 1) \tag{19}$$
$$s_z/a \sim TN(M_{sz}, \varsigma_{sz}, 0, 1) \tag{20}$$
$$a \sim TN(M_a, \varsigma_a, 0, \infty) \tag{21}$$
$$t_{er} \sim TN(M_{ter}, \varsigma_{ter}, 0, \infty) \tag{22}$$
$$s_t \sim TN(M_{st}, \varsigma_{st}, 0, \infty) \tag{23}$$

---

[1] We would like to thank Brandon Turner for this very helpful suggestion.

Because the z/a and s_z/a parameters are proportions and their constituents are positive these parameters fall between zero and one, so were truncated to the $(0, 1)$ interval. The remaining parameters are bounded below at 0 but unbounded on the right.

For the standard versions of the DDM, there are drift rates for target and lure items in each strength condition, the η_{lure} parameter, and the S parameter. Participants drift rate variability parameters are treated as samples from zero truncated normal distributions, and v parameters are sampled from normal distributions because drift rates can be negative or positive.

$$v_{old} \sim Normal(M_{vold}, \varsigma_{vold}) \tag{24}$$
$$v_{new} \sim Normal(M_{vnew}, \varsigma_{vnew}) \tag{25}$$
$$\eta_{lure} \sim TN(M_{\eta lure}, \varsigma_{\eta lure}, 0, \infty) \tag{26}$$
$$S \sim TN(M_S, \varsigma_S, 0, \infty) \tag{27}$$

The EVSD variants of the conventional DDM lack an S parameter and instead have only a single η parameter that corresponds to both targets and lures.

The LR-DDM models have d parameters for each strength condition. They are sampled from truncated normal distributions because their values cannot go below zero, as the transfer functions are only applicable for values of d greater than zero (Glanzer et al., 2009).

$$d \sim TN(M_d, \varsigma_d, 0, \infty) \tag{28}$$

For the group level mean (M) parameters, we used mildly informative priors:

$$M_{z,sz} \sim TN(.5, .5, 0, 1) \tag{29}$$
$$M_{st} \sim TN(.25, .25, 0, \infty) \tag{30}$$
$$M_{ter} \sim TN(.5, .5, 0, \infty) \tag{31}$$
$$M_a \sim TN(2, 2, 0, \infty) \tag{32}$$
$$M_{S,\eta lure} \sim TN(1, 1, 0, \infty) \tag{33}$$
$$M_{vold} \sim Normal(2, 2) \tag{34}$$
$$M_{vnew} \sim Normal(-2, 2) \tag{35}$$
$$M_d \sim TN(4, 4, 0, \infty) \tag{36}$$
$$M_{dc} \sim Normal(0, .5) \tag{37}$$

For the group level standard deviation (ς) parameters we used the following mildly informative priors:

$$\varsigma_{a,S,\eta lure,vold,vnew,d,dc} \sim \Gamma(1, 1) \tag{38}$$
$$\varsigma_{z,sz,st,ter} \sim \Gamma(1, 3) \tag{39}$$

With Dataset 1, a total of 82 MCMC chains were used due to the large number of parameters in each of the models, while Datasets 2–3 used 48 chains and Dataset 4 used 60 chains. After 2500 burn-in iterations were discarded, the MCMC chains were thinned by only accepting 1 sample every 10th iteration. The process continued until 1500 MCMC samples were accepted for each chain. Convergence was assessed using the Gelman-Rubin statistic; a model was considered converged if this statistic was less than 1.1 for all parameters. This criterion was satisfied for all models. Chains were also visually assessed for convergence.

4.8. Model selection results

Like the Akaike Information Criterion commonly used to correct for model complexity in maximum likelihood estimation, DIC is the sum of two terms, a measure of the minimum misfit and a penalty term for model complexity. We used the deviance of the mean of the posterior parameter estimates as a minimum misfit measure. The penalty term is twice the "effective" number of parameters (pD). In hierarchical models shrinkage means that the actual number of parameters overestimates model complexity; pD attempts to take the effect of shrinkage into account, although it is well known that DIC still tends to prefer overly complex models.

Table 3 compiles the model selection results for all models in each dataset. For each of the LR models, pD is considerably lower than for their standard DDM counterparts, indicating that the greater simplicity of the LR-DDM in terms of the nominal number of parameters translates into greater simplicity when shrinkage is also taken into account.

With respect to model selection (i.e., preferring the models with a smaller DIC) there are four regularities that can be seen for every dataset. First, unequal variance variants of both the standard and LR-DDM models outperform the equal variance versions, which replicates the analyses of Starns and Ratcliff (2014). Additionally, LR-DDM variants with the drift criterion win over the variants that lack the drift criterion in each dataset; this applies to both the equal variance and unequal variance models. Third, for Datasets 1–3, the LR-DDM with unequal variance wins over the unequal variance standard DDM whereas the unequal variance DDM wins in Dataset 4.

Table 3
Model selection results.

Model	UVSD	DC	pD	DIC
Dataset 1				
Standard DDM			903.01	−9843.14
Standard DDM	✔		943.54	−9896.91
LR-DDM			611.26	−10087.99
LR-DDM		✔	671.57	−10216.19
LR-DDM	✔		682.66	−10105.62
LR-DDM	✔	✔	738.08	**−10236.92**
Dataset 2				
Standard DDM			348.15	5265.32
Standard DDM	✔		396.74	5240.93
LR-DDM			232.15	5196.16
LR-DDM		✔	279.59	5052.34
LR-DDM	✔		272.55	5106.54
LR-DDM	✔	✔	314.65	**5047.40**
Dataset 3				
Standard DDM			588.49	9711.52
Standard DDM	✔		658.78	9671.58
LR-DDM			397.71	9717.25
LR-DDM		✔	458.67	9322.09
LR-DDM	✔		465.86	9497.09
LR-DDM	✔	✔	526.91	**9300.86**
Dataset 4				
Standard DDM			230.34	−3162.63
Standard DDM	✔		266.95	−3307.83
LR-DDM			157.69	−2510.17
LR-DDM		✔	166.15	−3184.78
LR-DDM	✔		175.57	−3105.47
LR-DDM	✔	✔	196.43	**−3355.84**

Notes: UVSD = model variant with unequal variance, DC = model variant that includes a drift criterion parameter, pD = effective number of parameters, DIC = Deviance Information Criterion, Unequal variance = unequal variance between targets and lures is allowed in the model.
Winning model is highlighted in bold.

Most importantly, the LR variants with the drift criterion as well as unequal variance perform best out of all of the models in every dataset. As DIC measures model preference on a logarithmic scale, a difference or 10 or greater is usually considered as indicating a strong preference (Pratte & Rouder, 2012). Thus, in every data set there is strong evidence for the LR-DDM over the standard DDM. Confidence in a strong preference for the LR-DDM is increased by the fact that DIC's bias favoring complex models should have advantaged the standard DDM.

4.9. Posterior predictives

To evaluate how well each model is capturing the data, posterior predictives from the winning models in each class (namely the unequal variance standard DDM and the unequal variance LR-DDM with the drift criterion) were examined. Posterior predictives are obtained by using posterior parameters to generate data, and comparing these predictions to the data. For each model, 3.33% of posterior samples from each participant's posterior distributions were selected (1 in every 30th sample) and predictions from the model were simulated. To summarize the response time distributions, the .1, .5, and .9 quantile of each participant's correct and error response times were calculated for data and predictions. HR and FAR along with the quantile summaries were then averaged across participants. Three notable trends emerge in the data. First, each dataset exhibits a mirror effect. Consistent with the predictions of an LR-SDT model with greater variability for targets, the mirror effect was asymmetric, with larger changes in the HR than the FAR from weak to strong (Dataset 1: $\Delta HR = .078$, $\Delta FAR = −.06$, Dataset 2: $\Delta HR = .19$, $\Delta FAR = −.09$, Dataset 3: $\Delta HR = .22$, $\Delta FAR = −.13$, Dataset 4: $\Delta HR = .25$, $\Delta FAR = −.12$). Second, correct response times decrease as performance is increased for both targets and lures, a trend which is most evident in the tail of the RT distribution (the .9 quantile). Third, error RTs are not reliably affected by the performance manipulations.

Posterior predictive distributions for each dataset can be seen in Fig. 8, depicted as violin plots, where the density of the posterior distribution is indicated by the width of the plot. The figure indicates that both models closely match the data and reproduce the three qualitative trends described above. The standard DDM captures the choice probabilities better than the LR-DDM, as expected. However, the average account given by the LR-DDM captures all of the important trends in the data. Inspection of the fits for some of the individual participants revealed that a few deviated from the mirror pattern in their HR and FAR. The standard DDM was able to accommodate this pattern while the LR-DDM was not able to fit these participants. Nonetheless, these cases were sufficiently rare such that there was little impact on the fit to average HR and FAR. Further,
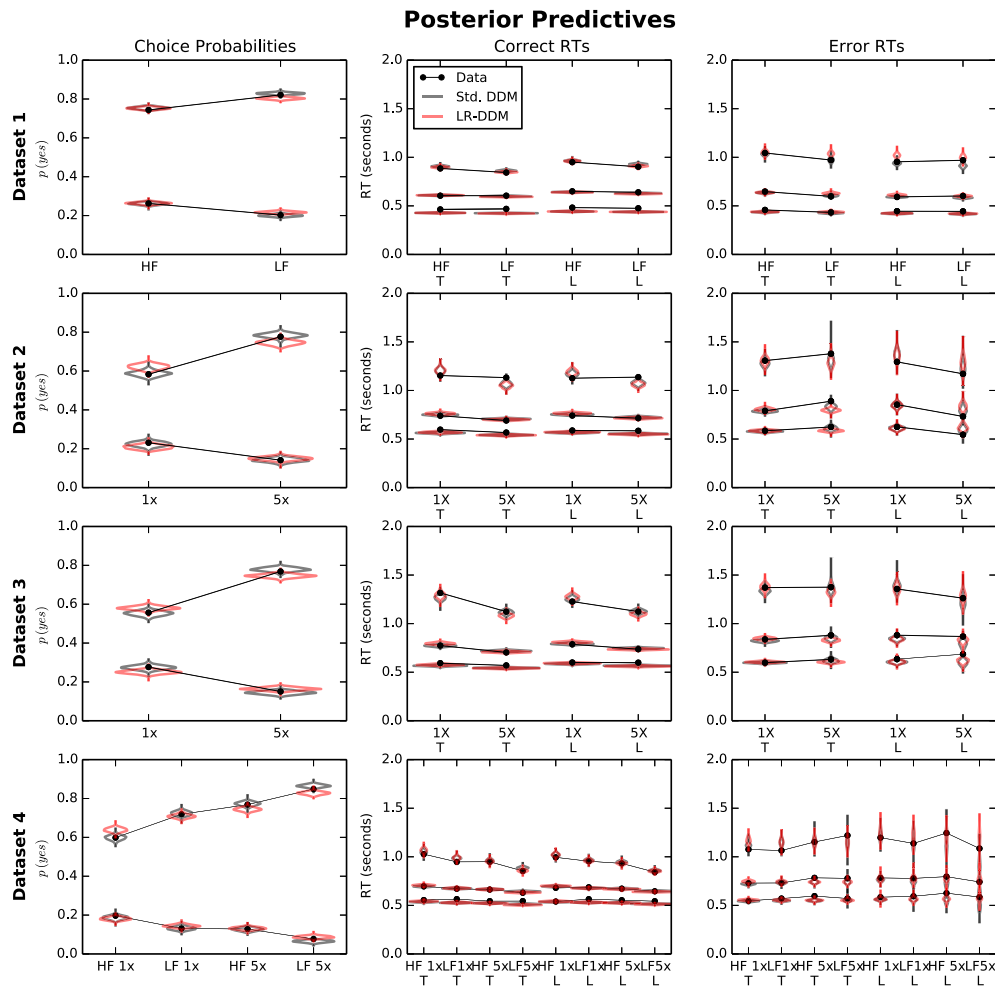
Posterior Predictives

Fig. 8. Posterior predictive distributions for each dataset for both the standard DDM with unequal variance and the unequal variance LR-DDM with the drift criterion along with the data. Predictions and data were averaged across participants. Depicted are the HR and FAR (left), correct RT .1, .5, and .9 quantiles (middle), along with the error .1, .5, and .9 quantiles (right). T = target, L = lure.

inspection of the correct and error RT distributions at both the individual and group levels revealed that both models appear to capture the data almost equally well.

4.10. Parameter estimates

In order to better understand the model selection results, we compare parameter estimates for the winning model (the unequal variance LR-DDM variant with a drift criterion) and the best standard DDM model (the unequal variance DDM). Quantiative comparisons were carried out using Bayesian p values, which are calculated by taking the difference between two posterior distributions and evaluating the proportion of samples above zero. Two-tailed tests for an $\alpha = .05$ were constructed by considering a difference to be significant if the proportion of samples was <.025 or >.975 (e.g., Pratte & Rouder, 2012).

Fig. 9 compares the group mean relative variability of the target distribution, M_S, across the two models. The posterior distributions are depicted as violin plots. Both models support the conclusion that there is greater variability in the target distribution than the lure distribution (all S values > 1.0). The LR-DDM produced significantly different estimates for Dataset 1 ($p = .997$) and Dataset 3 ($p = .015$). The cases where the estimates differ likely occur because the S parameter plays an additional role in the LR-DDM model relative to the standard DDM; breaking the symmetry of the mirror effect for values of S that deviate from 1, with values of S above 1.0 producing bigger changes in the HR portion than the FAR portion. For the standard DDM, S is mostly constrained by the difference in the speed of error and correct responses, with the degree of asymmetry mostly determined by v estimates. For Datasets 1 and 3 it appears that the relative error/correct speed and mirror asymmetry constraints traded off to some degree, resulting in lower S estimates.
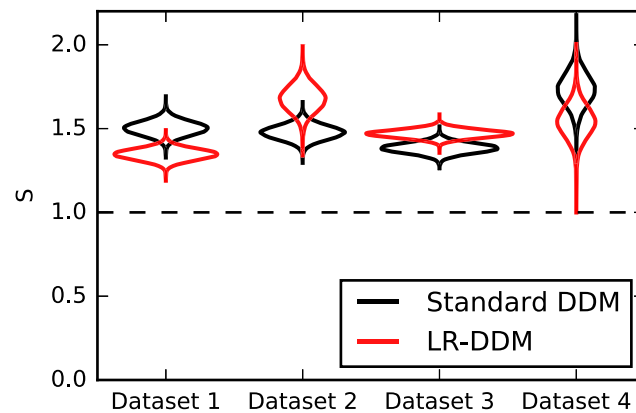
Fig. 9. Posterior distribution of the group mean M_S for both the Standard DDM (unequal variance) and LR-DDM (unequal variance with a drift criterion parameter) for Datasets 1, 2, 3, and 4.
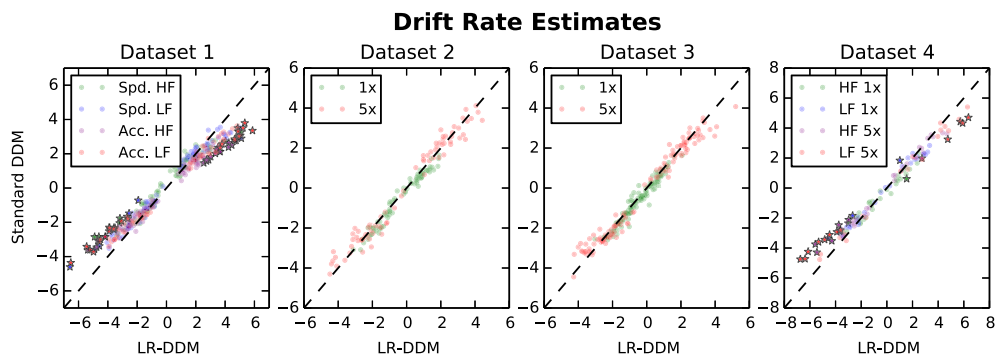

Fig. 10. Median estimates of each participant's drift rate for each condition in each dataset. For the LR-DDM model, this was calculated by transforming d and S according to Eqs. (1) and (2). Stars indicate significant differences between the standard and LR-DDM.

The LR-DDM models can outperform the standard DDM in model selection if they predict similar drift rates (i.e., $v_{SL}, v_{WL}, v_{WT}, v_{ST}$). This is because each must be separately estimated for the standard DDM. In the LR-DDM, in contrast, they are produced by transforming only two estimated parameters, (d, S) according to Eqs. (1) and (2), and so it suffers less of a complexity penalty for the same effective outcome.

Fig. 10 shows a scatterplot of the median posterior drift rate estimates for each participant and model in each dataset; significant differences according to the Bayesian p-value analysis are depicted as stars. Evidently there is a strong correspondence between models, with correlations of .987, .983, .984, and .988 for Datasets 1–4 respectively. The correspondence was closest for Datasets 2 and 3, with linear regression slopes between each of the model's drift rates of .956 and .936, where a slope of 1 would indicate the strongest correspondence. For Datasets 1 and 4 the slopes were .760 and .822 respectively. Comparisons between the two models' drift rates using Bayesian p-values converge with the slope analysis: there are no significant differences between the two models' drift rates in Datasets 2 and 3, whereas in Datasets 1 and 4 15.7% and 18.75% of the drift rate comparisons significantly differed between the two models. Inspection of Fig. 10 reveals that the weaker correspondence for Datasets 1 and 4 occurred because drift rates at the extreme end of the scale were larger in absolute magnitude in the LR-DDM model than the standard DDM model; these regions are also where the bulk of the significant differences between the two models arose. The divergence could be due to the fact that, in the LR-DDM, drift rate variability η increases linearly with d, so better performance due to larger values of d is counteracted to some degree by higher drift rate variability. To compensate, the LR-DDM may require larger values of d to predict the same performance as the standard DDM model, where η is fixed across all conditions.

Fig. 11 shows estimates of the drift criterion in the unequal variance LR-DDM for each dataset. To conveniently summarize the variability across participants, the median of each participant's posterior distribution was calculated and these estimates were summarized in histograms. One can see that there is substantial variability across participants in their drift criterion, suggesting that bias is not merely in the starting point of accumulation but in drift rates as well. This is congruent with the analyses of Bowen et al. (2016), who found that individual differences in bias were a function of both drift rate and starting point differences across participants.
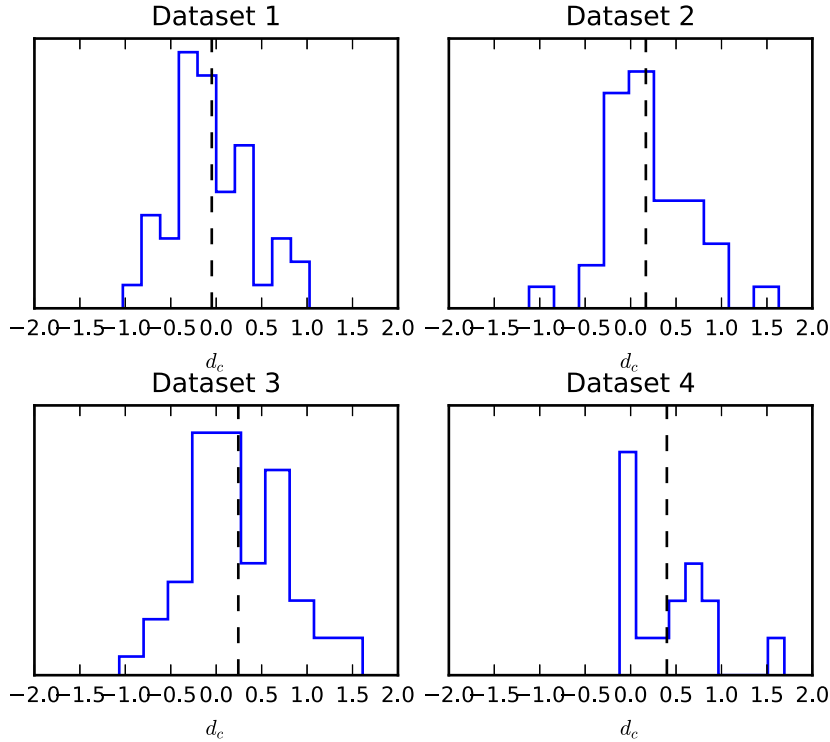
Drift Criterion Estimates

Fig. 11. Median estimates of each participant's drift criterion (d_c) from the unequal variance LR-DDM for each dataset.

Group mean estimates for the other DDM parameters ($M_{ter}, M_{st}, M_a, M_z$, and M_{sz}) can be seen in Fig. 12. There is a very strong consistency between the parameters of the two models, with the exception of M_{sz}, which is significantly larger in the LR-DDM in both the speed and accuracy conditions of Dataset 1 ($p \sim 0$ for both comparisons). No other parameter comparisons were significant. As mentioned previously, the primary purpose of starting point variability in the DDM is to be able to account for errors that are faster than correct responses. However, inspection of Fig. 8 reveals that both models appear equally able to capture the speeds of both correct and error responses. One possibility is that the LR-DDM requires larger values of sz/a to be able to overcome the relatively high drift rate variability in conditions of higher performance, as higher drift variability slows errors.

5. General discussion

Log-likelihood Ratio Signal Detection Theory (LR-SDT) provide a parsimonious explanation of several benchmark recognition memory phenomena, such as the mirror effect, the variance effect, and zROC length effect. LR-SDT models accomplish this by making a transformation of memory strengths into a log-likelihood ratio. The transformation effectively scales the observed memory strength for a given item by its expected memory strength for each state (studied and unstudied), and compares the two scaled values. In this article, we evaluate this framework's ability to capture RT distributions by using LR-SDT as a way to construct drift rate distributions for the Diffusion Decision Model (DDM).

When, as is commonly observed, target and lure memory strength distributions have unequal variance, the log-likelihood ratio transformation is nonlinear. We developed a linear approximation to the log-likelihood ratio transformation in the unequal variance case and showed that it preserved the benchmark mirror, variance and zROC length effects. Because the approximation is linear, a Gaussian memory strength distribution remains Gaussian when transformed. This provides a substantial computational advantage when the transform specifies the drift rate distribution for the DDM, because the DDM has an analytic solution for a Gaussian rate distribution which makes its use in computational modeling substantially more tractable. We called the resulting model the LR-DDM.

In order to test the LR-DDM, we compared it to the standard approach to applying the DDM to recognition memory data, where rates are estimated separately for each target and lure condition. We evaluated several variants of the LR-DDM and standard DDM by fitting them to four datasets using hierarchical Bayesian estimation. Our Bayesian approach allowed us to identify the variants that provided the best tradeoff between goodness-of-fit and simplicity. The four datasets came from experiments that produced both item mirror effects, by manipulating word frequency with study-test lists, and
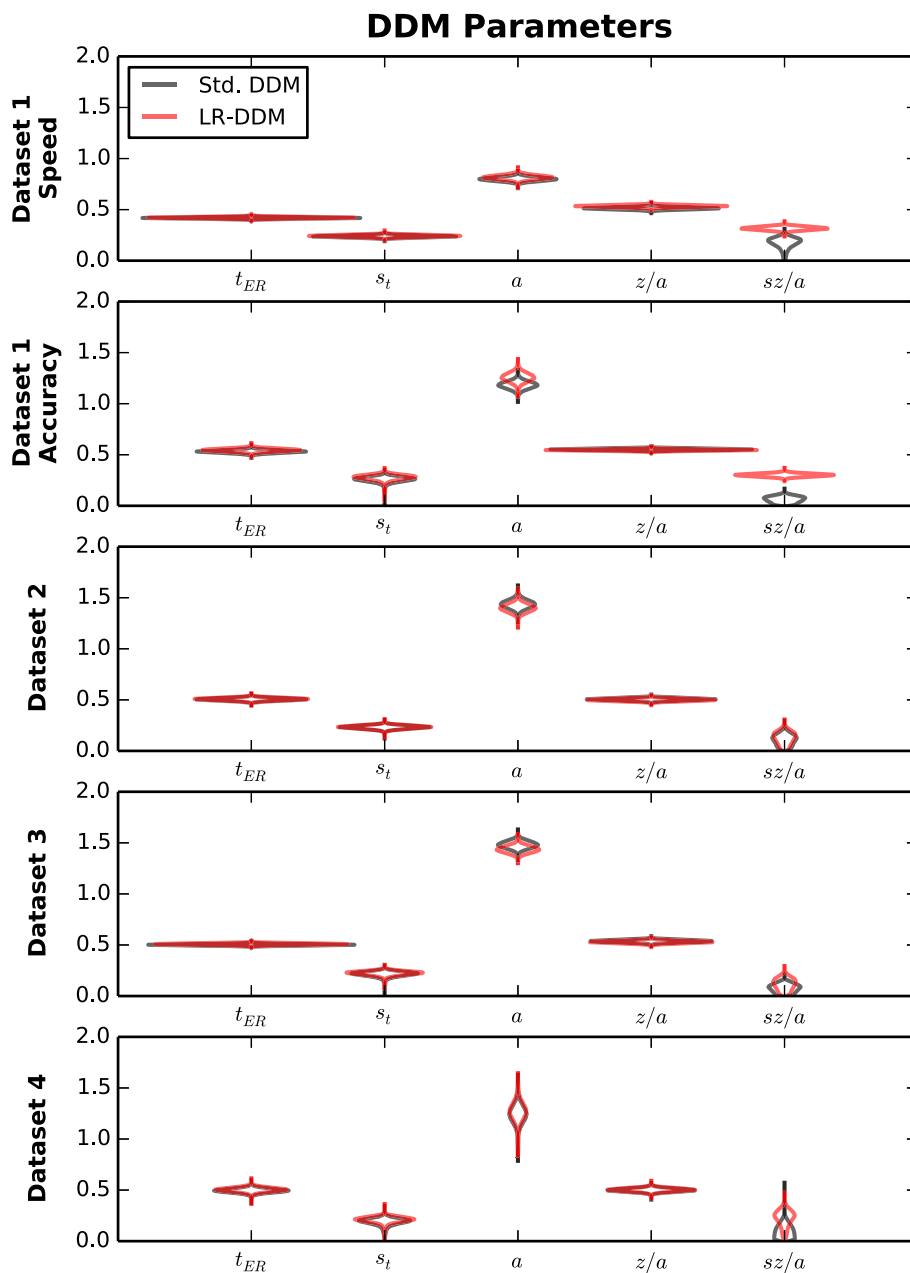
Fig. 12. Posterior distributions for the group mean parameters for the DDM parameters unrelated to drift rates ($t_{ER}, s_t, a, z/a$, and sz/a) for each of the datasets.

list-strength mirror effects, by manipulating the number of study item presentations between study-test lists. The variants included both equal and unequal variance models. For the LR-DDM we also examined variants with an estimated drift criterion and with a drift criterion fixed at zero. For the standard DDM, which allows drift rates for both targets and lures, a drift criterion could not be estimated as the drift criterion cannot be identified in the designs of the datasets we examined. Although one could follow the conventional parameterization in SDT and fix the drift rates for lures and parameterize a drift criterion instead, this model would be identical to the standard DDM used in the present investigation, and so there would be no change in our conclusions.

Consistent with the results of Starns and Ratcliff (2014), in every dataset unequal variance variants of each model were preferred over the equal variance variants in model selection for both the standard and LR-DDM models. In addition, the LR-DDM was found to be able to accommodate the shapes of RT distributions, providing fits to RTs that appeared ot be of the same quality as the fits of the standard DDM. The LR-DDM was also able to accommodate the mirror effects in the data.

Consistent with the predictions of an unequal variance LR model with greater variability for targets, each dataset employed exhibited a mirror effect that was asymmetric, with larger changes in the HR than the FAR, although the LR-DDM model did not fit the choice probabilities quite as well as the standard DDM. The LR-DDM variants that employ a drift criterion fared best in model selection in every dataset, with follow up analyses indicating that were was substantial individual variation in the bias associated with the drift criterion.

Investigation of parameter estimates provided some insight as to why the LR-DDM model was preferred: its predicted drift rates were very similar to the drift rates that were estimated in the standard DDM. This occurred even though the LR-DDM model employs fewer parameters and predicts the mirror effect a priori. These results suggest that the standard DDM, which is the model most frequently applied to recognition memory data, produces drift rates that approximate those generated from the log-likelihood ratio transformation. In the following sections, we discuss evidence for the LR-SDT framework in other paradigms, how the model can be extended to deal with broken mirror effects, and the relation to current process models of recognition memory.

5.1. Two Alternative Forced Choice (2AFC) recognition

In this article we have restricted to discussion of findings from yes/no recognition. Nonetheless, converging evidence for LR-SDT has been obtained using the 2AFC paradigm. Glanzer and Bowles (1976) conducted a relatively strong test of the mirror effect in 2AFC by using null trials in their design, which are trials that lack a correct answer. The null trials consisted of HF and LF word pairs that were either both targets or both lures. A priori, one might expect participants to merely guess on these trials, producing choice probabilities of .5 for either response option. Instead, in null target trials, participants chose the LF word more often ($p(LF_{target}, HF_{target}) > .5$) whereas in the null lure trials the reverse occurred and participants more often chose the HF word ($p(HF_{lure}, LF_{lure}) > .5$). Such a finding specifically necessitates a model that has a mirror ordering in the means of the distributions. An alternative account of the mirror effect is one where HF and LF words have identical means but HF words have greater variance than LF words, as this would predict higher HR and lower FAR for LF words. However, the variances account incorrectly predicts $p(LF_{target}, HF_{target}) = .5$ and $p(HF_{lure}, LF_{lure}) = .5$ for the null choice trials (Osth & Dennis, 2015).

In LR-SDT models, the means of the target and lure distributions approach zero as performance is decreased. Glanzer et al. (1991) demonstrated that, as a consequence, the discriminability between the HF and LF target and HF and LF lure distributions, which can be measured using $p(LF_{target}, HF_{target})$ and $p(HF_{lure}, LF_{lure})$, is monotonically related to performance, a phenomenon referred to as concentering. Glanzer and colleagues conducted numerous studies testing the predictions of concentering: manipulations of study time (Kim & Glanzer, 1993), repetitions (Hilford et al., 1997), study-test delay (Glanzer et al., 1991), and test position (Kim & Glanzer, 1995), produced the concentering pattern; specifically, the aforementioned null choices approached .5 in conditions of poorer performance. This is interesting when one considers the null lure trials: a decision between two unstudied items is affected by manipulations of the study episode, such as study time and study-test delay. Thus, data from the 2AFC paradigm offer converging results in support of the predictions of LR-SDT.

We did not consider data from the 2AFC paradigm in this article because the DDM has not yet been extended to 2AFC recognition. The DDM we consider here is a *relative evidence model*: evidence for one response counts against evidence for the other response. Relative evidence models do, however, make some interesting predictions about response times for null trials in the 2AFC paradigm. Drift rates for null target trials, for instance, could be calculated as $v_{LFold} - v_{HFold}$, which would be expected to be much lower than drift rates for a valid trial (e.g.: $v_{LFold} - v_{HFnew}$), making the prediction that null trials should overall be slower than valid trials. Nonetheless, one could alternatively construct a race model, where both response options have their own diffusion processes and whichever process terminates first determines the decision(e.g.: Ratcliff & Smith, 2004; Usher & McClelland, 2001). A race model predicts fastest decisions for null target trials due to both options having a high drift rate, intermediate decision times for valid trials, and the slowest decision times for null lure trials.

Very few investigations have considered response times in 2AFC recognition with the exception of a recent study by Jou, Flores, Cortes, and Leka (2016). Jou et al. found much slower RTs on null lure trials than valid trials, which is consistent with the predictions of both the race and relative evidence models. However, response times for null target trials were somewhat slower than valid trials. On the surface, these results appear contrary to both models. Future work that applies computational models to RTs in the 2AFC paradigm may be required to decide between the two classes of models, or to develop an alternative account.

5.2. Relation to process models and partially informed likelihood ratio models of recognition memory

The prevalence of the mirror effect and the relative ease in which log-likelihood ratio models can account for it has established likelihood ratio computation as standard decision mechanism in quantitative models of recognition memory, including the REM (Shiffrin & Steyvers, 1997), SLiM (McClelland & Chappell, 1998), and BCDMEM (Dennis & Humphreys, 2001) models along with the models of Cox and Shiffrin (2012) and Osth and Dennis (2015). Nonetheless, LR-SDT models have been criticized for assuming a great deal of knowledge on the part of the participant. Specifically, Hintzman (1994) questioned how participants could know the relevant categories in an experiment (such as weak and strong items), the locations of the distributions for each of the categories, and also which category an item belongs to at test.

Criss and McClelland (2006) introduced an important distinction between several of these models and the exact likelihood-ratio transformation. Specifically, they referred to this class of models as "partially informed" likelihood-ratio models. Partially informed likelihood ratio models, including including REM, SLiM, and BCDMEM, make a distinction between the memory strength parameters (d and S in our notation) and the participant's estimates of these parameters that are employed in the likelihood ratio transformation. In the partially informed models matches to the representations in memory are used to estimate memory strength parameters on a trial by trial basis. Thus, these models provide a mechanistic explanation of how likelihood ratio transformations could occur. This was a critical development, as the LR-SDT framework in and of itself is agnostic as to how such an estimation occurs.

Under conditions of performance normally seen in episodic recognition memory experiments these models are still subject to the same predictions as LR-SDT models. Log likelihood ratio distributions in REM, SLiM, and BCDMEM produce mirror effects as performance is increased and the variability of the log-likelihood ratio distributions increases as performance is increased. In this sense, testing the LR-SDT framework provides a test of the whole family of likelihood ratio models, including models that are partially informed. The results of the model selection in the present article suggest that such process models of recognition memory, including REM, SLiM, BCDMEM, and the Osth and Dennis (2015) model, show promise in their extension to accommodate RT distributions by applying a back-end diffusion model.[2]

An additional account for how participants might learn the necessary information to calculate likelihood ratio transformations separate from the aforementioned process models was proposed by Wixted and Gaitan (2002), who reasoned that expected strength distributions may be constructed from experience with the task. Adult human participants have extensive history with performing recognition before entering the laboratory. However, in animal learning experiments, researchers can understand the learning history and how this affects performance. Wixted and Gaitan (2002) reviewed evidence from pigeons performing recognition memory and found that they exhibit mirror effects with different retention intervals in the delayed match to sample paradigm, but only after sufficient experience with the task. Wixted and Gaitan (2002) presented a model where expected memory strengths for a given condition are built from expertise with the task with the goal of predicting rewards. When the model was sufficiently experienced, it was isomorphic to the predictions of LR-SDT.

Our introduction of the linear approximation to the unequal variance signal detection model was motivated by mathematical elegance, computational efficiency and its ability to capture concisely fundamental patterns in recognition data. However, it also sheds light on the nature of the problem that the decision system must solve in order to perform close to optimally, as it requires learning of only two simple factors, a difference in shift and a difference in scaling when conditioning likelihood estimates on the test item being previously studied or not. This is a considerably simpler burden on the participant than the true unequal variance likelihood ratio transformation, in which the participant has to estimate a quadratic transfer function between memory strengths and log likelihood ratios.

5.3. Broken mirror effects and future directions

As was previously mentioned, "broken" mirror effects have been observed where the HR portion is equivalent or even reversed across two strength classes despite a change in the FAR (Balota, Burgess, Cortese, & Adams, 2002; Criss & Shiffrin, 2004; Hirshman & Arndt, 1997; Joordens & Hockley, 2000). Often these reversals are associated with conditions that likely cause a poor initial encoding, such as the influence of Midazalam during study (Hirshman, Fisher, Henthorn, Arndt, & Passannante, 2002) or Dementia of the Alzheimers Type (Balota et al., 2002). Within LR-SDT such violations could be explained by nominal targets acting like lures, because they were not encoded at during study (e.g.: DeCarlo, 2002, 2007). Depending on the proportion of items for which encoding failed, the greater recognition rate for LF over HF words that were properly encoded (i.e., the usual HR portion of the mirror effect) may be nullified or even reversed by the greater recognition rate for HF than LF words for which encoding failed (i.e., the usual greater recognition rate for HF than LF lures). Although the FAR difference due to word frequency is more stable, it too has been found to reverse under at least one circumstance, when lure items are very similar to targets both semantically and orthographically (Malmberg, Holden, & Shiffren, 2004). In this case similar lures may be acting like targets, because they inherit much of the memory strength of the targets to which they are similar, and so have a higher recognition rate when they are LF than HF words.

Another example of a deviation from the standard pattern was reported by Hemmer and Criss (2013), who measured the word frequency effect on a continuous scale, rather than using the traditional two level (HF vs. LF) design. They found a U-shaped pattern of HRs, where the HRs decrease as WF is increased but begin to increase again at the very high end of the word frequency scale. As previously mentioned, broken mirror effects may emerge in LR-SDT if different WF classes have different values of S, perhaps explaining these results if very high frequency words have higher values of S. Another possible explanation within the LR-SDT framework is that for extreme ends of the WF scale, participants under-estimate the frequency of very high frequency words, producing deviations between the true values of d and the estimated value of d used in the likelihood ratio transformation. In particular, it may be important to make a distinction between the true and subjective (i.e., estimated by the participant) strengths (d) in LR-SDT. If true and subjective ds are allowed to differ arbitrarily for every condition in an experiment, the model is equivalent to a traditional free parameter SDT model, and is no longer com-

---

[2] Cox and Shiffrin (2012) presented a process model that is a partially informed likelihood ratio model. However, their model has only been used to predict mean RT and has not yet been extended to complete RT distributions.

mitted to predicting a mirror effect. Nonetheless, one could preserve some of the constraint of the LR-SDT approach by enforcing functional relationship between true and subjective values of *d*. For example, a sigmoidal relationship would predict accurate subjective estimate in the middle WF ranges and misestimations at the extremes. A consequence of the misestimation is that as WF increases to the upper end of the scale, very high frequency words are treated in the LR transform as if they are weaker than they actually are, which will slow the deceleration of the HR and potentially reverse it under some parameterizations.

Alternately, very high frequency words may differ in other item characteristics that could affect recognition memory, such as having shorter word lengths, which would increase the HRs. Shorter word lengths could manifest as changes in either *d* or *S*. Confounding from co-variaton in item characteristics may be the reason why a separate investigation of word frequency using a continuous scale by Pratte et al. (2010) that incorporated estimation of individual item effects found a preserved mirror effect across the entire range they examined. These examples illustrate that, although mirror effects provide a pervasive regularity for which LR-SDT (and hence the LR-DDM) provide a parsimonious explanation, there also exist mechanisms by which these models can provide plausible explanations of exceptions. We believe that it is a virtue if a theory makes strong predictions that are correct in the majority of cases, even though they might not always be right, because the exceptions highlight interesting areas for further investigation. Such investigation can then focus on obtaining converging evidence to develop and test auxiliary mechanisms, or, if to many such mechanisms end up being required, they may overturn the theory in favor of a more parsimonious explanation. In contrast, fairly unconstrained models, such as the standard DDM, serve mainly as descriptions that are not theoretically productive because they lack testable predictions.

In light of these considerations, it will likely be useful for the future development of the LR-DDM to focus on phenomena that it does not naturally predict, such as broken mirror effects. Even in cases where a mirror effect occurs on average, the existence and influence of plausible mechanisms that tend to break the mirror effect – such as failures of encoding or subjective estimation of memory strength, or item confounds and similarity effects – make it likely that there will be quantitative deviations from the predictions of the LR-DDM, for at least in some participants and items. This in turn means that reliance on formal model selection using complexity penalties to fairly compare the LR-DDM to more purely descriptive approaches is likely not to be sufficient by itself. Instead, developing a coherent account of the causes of exceptions, and comparisons with alternative models that also make theoretical commitments, and hence testable predictions, are more likely to be fruitful.

6. Conclusion

We provided a quantitative test of the log-likelihood ratio signal-detection theory (LR-SDT) of recognition memory by developing the LR-DDM, a model constituted of an approximate log-likelihood ratio transformation that specifies the input to a diffusion decision model (DDM). We compared this model to the DDM as standardly applied to recognition memory data, where each condition receives its own independent drift rate estimate. Model selection favored the more parsimonious LR-DDM, and evaluation of the LR-DDM's posterior predictive estimates indicated that, despite its simplicity, the LR-DDM consistently provided a good account of the data for most participants in several experiments displaying word frequency and list-strength mirror effects. These results suggest that the LR-SDT framework is not restricted to ROCs, but can also be extended to provide a comprehensive account of both choice probabilities and RT distributions.

Acknowledgements

Appendix A. Derivations of the *d*/2 Approximation of the log likelihood ratio transformation

In this section, we introduce the mathematics behind the *d*/2 approximation to the log-likelihood ratio transformation.

Following Glanzer et al. (2009), we assume new items are sampled from a $N(0, 1)$ distribution (i.e., a Gaussian distribution with mean zero and standard deviation 1) and old items are sampled from a $N(d, S)$ distribution. They show that when $S = 1$, the likelihood transformation for a strength value *x* is:

$$\lambda(x) = dx - dx^2/2 \tag{40}$$

When $S > 1$, the likelihood transformation is quadratic:

$$\lambda(x) = \frac{S^2 - 1}{2S^2}x^2 + \frac{d}{S^2}x - \left(\frac{d^2}{2S^2} + \log(S)\right). \tag{41}$$

To obtain a linear approximation of the above equation via a first-order Taylor series, we take the first derivative of Eq. (41):

$$\lambda(x)' = \frac{x(S^2 - 1)}{S^2} + d/S^2 \tag{42}$$

A linear approximation around the point $x = a$ is then given by:

$$\lambda_a(x) = \lambda(a) + (c - a)\left(\frac{S^2 - 1}{S^2} a + \frac{d}{S^2}\right) \tag{43}$$

When $a = d/2$ (the intersection between the two distributions when there is equal variance) the approximation has an analogous form to Eq. (40):

$$\lambda_{d/2}(x) = dc\frac{S^2 + 1}{2S^2} - \left(\frac{d^2(S^2 + 3)}{8S^2} + \log(S)\right) \tag{44}$$

This results in the following log-likelihood ratio distributions corresponding to targets (T) and lures (L):

$$\mu_{\lambda T} = d^2 \frac{S^2 + 1}{2S^2} - \left(\left(\frac{d^2}{2}\right)\left(\frac{S^2 + 3}{4S^2}\right) + \log(S)\right) \tag{45}$$

$$\mu_{\lambda L} = -\left(\left(\frac{d^2}{2}\right)\left(\frac{S^2 + 3}{4S^2}\right) + \log(S)\right) \tag{46}$$

$$\sigma_{\lambda T} = dS\frac{S^2 + 1}{2S^2} \tag{47}$$

$$\sigma_{\lambda L} = d\frac{S^2 + 1}{2S^2} \tag{48}$$

References

Aminoff, E. M., Clewett, D., Freeman, S., Frithsen, A., Tipper, C., Johnson, A., ... Miller, M. B. (2012). Individual differences in shifting decision criterion: A recognition memory study. *Memory & Cognition, 40*(7), 1016–1030.

Arnold, N. R., Broder, A., & Bayen, U. J. (2015). Empirical validation of the diffusion model for recognition memory and a comparison of parameter-estimation methods. *Psychological Research, 79*, 882–898.

Averell, L., Prince, M., & Heathcote, A. (2016). *Journal of Memory and Language, 88*(C), 51–69.

Balota, D. A., Burgess, G. C., Cortese, M. J., & Adams, D. R. (2002). The word-frequency mirror effect in young, old, and early-stage Alzheimer's disease: Evidence for two processes in episodic recognition performance. *Journal of Memory and Language, 46*(1), 199–226.

Bowen, H. J., Spaniol, J., Patel, R., & Voss, A. (2016). A diffusion model analysis of decision biases affecting delayed recognition of emotional stimuli. *PLoS ONE, 11*(1).

Brown, S. D., & Heathcote, A. (2003). Averaging learning curves across and within participants. *Behavior Research Methods, Instruments, & Computers, 35*(1), 11–21.

Brown, J., Lewis, V. J., & Monk, A. F. (1977). Memorability, word frequency and negative recognition. *The Quarterly Journal of Experimental Psychology, 29*(3), 461–473.

Cox, G. E., & Shiffrin, R. M. (2012). Criterion setting and the dynamics of recognition memory. *Topics in Cognitive Science, 4*, 135–150.

Criss, A. H. (2006). The consequences of differentiation in episodic memory: Similarity and the strength based mirror effect. *Journal of Memory and Language, 55*, 461–478.

Criss, A. H. (2010). Differentiation and response bias in episodic memory: Evidence from reaction time distributions. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 36*(2), 484–499.

Criss, A. H., & McClelland, J. L. (2006). Differentiating the differentiation models: A comparison of the retrieving effectively from memory model (REM) and the subjective likelihood model (SLiM). *Journal of Memory and Language, 55*, 447–460.

Criss, A. H., & Shiffrin, R. M. (2004). Interactions between study task, study time, and the low-frequency hit rate advantage in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 30*(4), 778–786.

DeCarlo, L. T. (2002). Signal detection theory with finite mixture distributions: Theoretical developments with applications to recognition memory. *Psychological Review, 109*(4), 710–721.

DeCarlo, L. T. (2007). The mirror effect and mixture signal detection theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(1), 18–33.

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review, 108*(2), 452–478.

Donkin, C., & Nosofsky, R. M. (2012). The structure of short-term memory scanning: An investigation using response time distribution models. *Psychonomic Bulletin & Review, 19*, 363–394.

Donkin, C., Nosofsky, R. M., Gold, J., & Shiffrin, R. M. (2013). Discrete-slots models of visual working-memory response times. *Psychological Review, 120*, 873–902.

Egan, J. P. (1958). *Signal detection theory and ROC analysis*. Hearing and Communication Laboratory.

Estes, W. K., & Maddox, W. T. (2005). Risks of drawing inferences about cognitive processes from model fits to individual versus average performance. *Psychonomic Bulletin & Review, 12*(3), 403–408.

Fific, M., Little, D. R., & Nosofsky, R. M. (2010). Logical-rule models of classification response times: A synthesis of mental-architecture, random-walk, and decision-bound approaches. *Psychological Review, 117*(2), 309–348.

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory and Cognition, 13*(1), 8–20.

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(1), 5–16.

Glanzer, M., Adams, J. K., & Iverson, G. (1991). Forgetting and the mirror effect in recognition memory: Concentering of underlying distributions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(1), 81–93.

Glanzer, M., Adams, J. K., Iverson, G. J., & Kim, K. (1993). The regularities of recognition memory. *Psychological Review, 100*(3), 546–567.

Glanzer, M., & Bowles, N. (1976). Analysis of the word-frequency effect in recognition memory. *Journal of Experimental Psychology: Human Learning and Memory, 2*(1), 21–31.

Glanzer, M., Hilford, A., & Maloney, L. T. (2009). Likelihood ratio decisions in memory: Three implied regularities. *Psychonomic Bulletin & Review, 16*(3), 431–455.

Glanzer, M., Kim, K., Hilford, A., & Adams, J. K. (1999). Slope of the receiver-operating characteristic in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(2), 500–513.

Heathcote, A. (2003). Item recognition memory and the receiver operating characteristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(6), 1210–1230.

Heathcote, A., Brown, S. D., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review, 7*(2), 185–207.

Hemmer, P., & Criss, A. H. (2013). The shape of things to come: Evaluating word frequency as a continuous variable in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 39*, 1947–1952.

Hilford, A., Glanzer, M., & Kim, K. (1997). Encoding, repetition, and the mirror effect in recognition memory: Symmetry in motion. *Memory & Cognition, 25*(5), 593–605.

Hintzman, D. L. (1994). On Explaining the mirror effect. *Journal of Experimental Psychology-Learning Memory and Cognition, 20*(1), 201–205.

Hirshman, E. (1995). Decision processes in recognition memory: Criterion shifts and the list-strength paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(2), 302–313.

Hirshman, E., & Arndt, J. (1997). Discriminating alternative conceptions of false recognition: The cases of word concreteness and word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(6), 1306–1323.

Hirshman, E., Fisher, J., Henthorn, T., Arndt, J., & Passannante, A. (2002). Midazolam amnesia and dual-process models of the word-frequency mirror effect. *Journal of Memory and Language, 47*(4), 499–516.

Joordens, S., & Hockley, W. E. (2000). Recollection and familiarity through the looking glass: When old does not mirror new. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1534–1555.

Jou, J., Flores, S., Cortes, H. M., & Leka, B. G. (2016). The effects of weak versus strong relational judgments on response bias in Two-Alternative-Forced-Choice recognition: Is the test criterion-free? *Acta Psychologica, 167*, 30–44.

Kantner, J., & Lindsay, D. S. (2012). Response bias in recognition memory as a cognitive trait. *Memory & Cognition, 40*, 1163–1177.

Kim, K., & Glanzer, M. (1993). Speed versus accuracy instructions, study time, and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(3), 638–652.

Kim, K., & Glanzer, M. (1995). Intralist interference in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(5), 1096–1107.

Laming, D. (1968). *Information theory of choice reaction time*. New York: Wiley.

Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical Bayesian models. *Journal of Mathematical Psychology, 55*, 1–7.

Lee, M. D., & Wagenmakers, E. J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press. Unpublished lecture notes, University of California, Irvine ~mdlee/CourseBook_v1.pdf" xlink:type="simple"><http://www.socsci.uci.edu/~mdlee/CourseBook_v1.pdf>.

Leite, F. P., & Ratcliff, R. (2011). What cognitive processes drive response biases? A diffusion model analysis. *Judgment and Decision Making, 6*(7), 651–687.

Malmberg, K. J., Holden, J. E., & Shiffrin, R. M. (2004). Modeling the effects of repetitions, similarity, and normative word frequency on old-new recognition and judgments of frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(2), 319–331.

McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review, 105*(4), 724–760.

Myung, J. I. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology, 44*(1), 190–204.

Osth, A. F., & Dennis, S. (2014). Associative recognition and the list strength paradigm. *Memory & Cognition, 42*(4), 583–594.

Osth, A. F., & Dennis, S. (2015). Sources of interference in item and associative recognition memory. *Psychological Review, 122*(2), 260–311.

Osth, A. F., Dennis, S., & Kinnell, A. (2014). Stimulus type and the list strength paradigm. *Quarterly Journal of Experimental Psychology, 67*(9), 1826–1841.

Pratte, M. S., & Rouder, J. N. (2012). Assessing the dissociability of recollection and familiarity in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(6), 1591–1607.

Pratte, M. S., Rouder, J. N., & Morey, R. D. (2010). Separating mnemonic process from participant and item effects in the assessment of ROC asymmetries. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(1), 224–232.

Rae, B., Heathcote, A., Donkin, C., Averell, L., & Brown, S. (2014). The hare and the tortoise: Emphasizing speed can change the evidence used to make decisions. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 40*(5), 1226–1243.

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

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review, 111*, 159–182.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation, 20*, 873–922.

Ratcliff, R., McKoon, G., & Tindall, M. (1994). Empirical generality of data from recognition memory: Receiver-operating characteristic functions and implications for the global memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(4), 763–785.

Ratcliff, R., & Murdock, B. B. (1976). Retrieval Processes in Recognition Memory. *Psychological Review, 83*(3), 190–214.

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science, 9*(5), 347–356.

Ratcliff, R., Sheu, C. F., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review, 99*(3), 518–535.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential-sampling models for two choice reaction time. *Psychological Review, 111*, 333–367.

Ratcliff, R., & Starns, J. J. (2009). Modeling confidence and response time in recognition memory. *Psychological Review, 116*(1), 59–83.

Ratcliff, R., Thapar, A., & McKoon, G. (2004). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory and Language, 50*, 408–424.

Ratcliff, R., Thapar, A., & McKoon, G. (2010). Individual differences, aging, and IQ in two-choice tasks. *Cognitive Psychology, 60*, 127–157.

Ratcliff, R., Thapar, A., & McKoon, G. (2011). Effects of aging and IQ on item and associative memory. *Journal of Experimental Psychology: General, 140*(3), 464–487.

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review, 9*, 438–481.

Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review, 106*(2), 261–300.

Reed, A. V. (1976). List length and the time course of recognition in immediate memory. *Memory & Cognition, 4*(1), 16–30.

Shiffrin, R. M., Lee, M. D., Kim, W., & Wagenmakers, E. J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science, 32*, 1248–1284.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM – Retrieving effectively from memory. *Psychonomic Bulletin & Review, 4*(2), 145–166.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B, 64*, 583–639.

Starns, J. J. (2014). Using response time modeling to distinguish memory and decision processes in recognition and source tasks. *Memory & Cognition, 42*, 1357–1372.

Starns, J. J., & Ratcliff, R. (2014). Validating the unequal-variance assumption in recognition memory using response time distributions instead of ROC functions: A diffusion model analysis. *Journal of Memory and Language, 70*, 36–52.

Starns, J. J., Ratcliff, R., & McKoon, G. (2012). Evaluating the unequal-variance and dual-process explanations of the zROC slopes with response time data and the diffusion model. *Cognitive Psychology, 64*, 1–34.

Starns, J. J., Ratcliff, R., & White, C. N. (2012). Diffusion model drift rates can be influenced by decision processes: An analysis of the strength-based mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(5), 1137–1151.

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

Stretch, V., & Wixted, J. T. (1998b). On the differences between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(6), 1379–1396.

Turner, B. M., Sederberg, P. B., Brown, S. D., & Steyvers, M. (2013). A method for efficiently sampling from distributions with correlated dimensions. *Psychological Methods, 18*(3), 368–384.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review, 108*, 550–592.

Vandekerckhove, J., Tuerlinckx, F., & Lee, M. D. (2011). Hierarchical diffusion models for two-choice response times. *Psychological Methods, 16*(1), 44–62.

Voss, A., & Voss, J. (2008). A fast numerical algorithm for the estimation of diffusion model parameters. *Journal of Mathematical Psychology, 52*(1), 1–9.

Voss, A., Voss, J., & Lerche, V. (2015). Assessing cognitive processes with diffusion model analyses: A tutorial based on fast-dm-30. *Frontiers in Psychology, 6*(14), 470.

Wagenmakers, E. J. (2009). Methodological and empirical developments for the Ratcliff diffusion model of response times and accuracy. *European Journal of Cognitive Psychology, 21*, 641–671.

Wagenmakers, E. J., Ratcliff, R., Gomez, P., & McKoon, G. (2008). A diffusion model account of criterion shifts in the lexical decision task. *Journal of Memory and Language, 58*(1), 140–159.

White, C. N., & Poldrack, R. A. (2014). Decomposing bias in different types of simple decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(2), 385–398.

Wickelgren, W. A., & Norman, D. A. (1966). Strength models and serial position in short-term recognition memory. *Journal of Mathematical Psychology, 3*, 316–347.

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review, 114*(1), 152–176.

Wixted, J. T., & Gaitan, S. C. (2002). Cognitive theories as reinforcement history surrogates: The case of likelihood ratio models of human recognition memory. *Animal Learning & Behavior, 30*(4), 289–305.