**Thresholded Source Memory in Continuous Report: A Circular Diffusion Model Account**

Jason Zhou a, Adam F. Osth a, Simon D. Lilburn a, and Philip L. Smitha

a Melbourne School of Psychological Sciences, The University of Melbourne

Corresponding Author:

Jason Zhou

Melbourne School of Psychological Sciences

The University of Melbourne

Parkville, VIC 3052, AUSTRALIA

jasonz1@student.unimelb.edu.au

Tel: LAB NUMBER

Declaration of Interest: none

Funding sources:

Abstract

**Insert Abstract**

*Keywords:* **Insert Keywords**

Episodic memory is memory of a particular event or occasion. A key part of episodic memory is the context in which an event occurred. This contextual component of episodic memory is known as *source memory*, and its importance to everyday memory functioning is easily illustrated. For instance, when identifying the suspect of a crime, a memory within the context of a crime scene will have very different implications than a memory of a chance encounter. To successfully complete such an identification task, the witness needs to: 1) encode information about the suspect and the context within his or her memory, 2) bind this information into an overall representation, 3) recognize the suspect at a later time, 4) retrieve the context in which the event occurred, and 5) mapping the retrieved information onto a response.

Our aim in this article is to compare competing models of source memory using a continuous outcome decision task. We present a novel method of analysis of the results of this task using the circular diffusion model of Smith (2016). This model provides a characterization of the decision processes that are involved in retrieving items from memory in continuous outcomes tasks and predicts both distributions of decision times and decision outcomes in such tasks. Using this model allows us to distinguish the contributions of memory processes and retrieval processes to source memory performance in a precise way.

Several classes of models in the source memory literature have been developed to formalize the nature of source memory retrieval. Continuous models of source memory predict that retrieved information may be inaccurate but not absent, allowing for a gradual decline in the quality of information retrieved (Banks, 2000; Mickes, Wais & Wixted, 2009). In contrast, threshold or discrete-state models holds that retrieval fails discretely, and so performance is made up of either precise responses, or guesses when the memory is subthreshold (Batchelder & Riefer, 1990; Klauer & Kellen, 2010). A third class of models can be regarded as hybrids of continuous and threshold models, and are known as dual-process models, in which different kinds of retrieval mechanism support different kinds of memory (Bowers & Schacter, 1990). In the case of source memory, dual-process models and the high threshold model make identical predictions.

The recognition memory literature is well-developed and the continuing debate between single-process and dual-process models in recognition memory can be applied to the source memory task. Although source memory is distinct from recognition memory, the two types of memory are often tested simultaneously, and models have been developed to account for both types of tasks (Starns, Hicks, Brown & Martin, 2008).

An influential example dual-process models in recognition memory is that of Yonelinas (1994), in which episodic memory is assumed to involve a mixture of two processes: a fast familiarity-based process and a slower recollection process. Familiarity, defined in this framework, is a quick judgment about whether or not an item has been encountered before based on the strength of its representation in memory (Yonelinas, 1994). Recollection is defined as a slower search process for qualitative information about the item (Atkinson & Juola, 1974). Recollection is assumed to be a thresholded retrieval process while familiarity is a continuous signal-detection process (Yonelinas, 1999). The basis for this assumption is the fact that in recollection people either succeed or fail to retrieve information. While the amount of information above the threshold may vary, Yonelinas (1999) argues that there is a threshold below which there is no information retrieved.

Within a dual-process framework, performance on an item recognition task is assumed to comprise a mixture of both familiarity and recollection processes, in which targets that were studied will be more familiar than lures which were not (Yonelinas, 1994). If successful, recollection enhances familiarity by providing details of the study event. Critically, both targets and lures have some level of familiarity, and differ continuously in the degree to which they are familiar, but only targets can be recollected because they are associated with a study event. Under the assumptions of the model, familiarity cannot be used to determine the source of items as it only consists of the strength at which items can be recalled, not the details of their associated study event. Consequently, Yonelinas (1994) conceptualized source memory as a pure recollection process, in that a correct response depends only on successful recollection of contextual details of the study event.

Given that source memory is dependent on recollection in this way, and that recollection according to the dual-process model is thresholded, the model implies that performance in source memory tasks should exhibit a retrieval threshold. When forced to make a source memory judgement in the case that recollection fails, participants can only guess in the absence of information. The dual-process model therefore predicts that overall performance on source memory tasks should be a mixture of two discrete components: informed responses and uninformed guesses, depending on whether recollection succeeds or fails respectively. Evidence for thresholded responding has been obtained almost exclusively from two-choice tasks.

# Two-choice Tasks

Traditionally, evidence both for and against a threshold in recollection has come from the examination of Receiver Operating Characteristic (ROC) curves (Yonelinas & Parks, 2007; Yonelinas, 1999; Slotnick & Dodson, 2005). In a two-choice paradigm with two possible sources of information, continuous and threshold models make divergent predictions about the shape of ROC curves. . The continuous model predicts a curvilinear ROC. This is because each of the two sources is associated with a normally distributed memory strength, which overlap with each other. As the response criterion is varied, the ratio of hit rates to false alarms will be such that the resultant shape of the plot is curvilinear (Slotnick & Dodson, 2005). In contrast, in a threshold model, each source is associated with a memory strength threshold, and where the strength of the memory representation fails to meet either response threshold, the ratio of false alarms to hit rates across criterion points is constant, producing a linear ROC (Rouder, Morey, Cowan, Zwilling, Morey & Pratte, 2008). The dual process model, in which source memory is dependent on a thresholded recollection process, also predicts linear source ROCs when familiarity is the same for both sources (Yonelinas, 1999).

The premise that recollection is thresholded was challenged by a reanalysis of the Yonelinas (1999) data by Slotnick and Dodson (2005), in which they conditioned source performance on recognition confidence ratings for each item. This reanalysis demonstrated that if source ROCs were plotted separately for different levels of confidence reported in the item recognition task, the highest confidence source ROCs were in fact curvilinear, contrary to the predictions of the dual-process model. Performance for unrecognized items was at chance and these items were on the diagonal of the ROC. As items rated with lower recognition confidence were included in the original data, the shape of the overall source ROC were apparently increasingly linear, and more consistent with the predictions of the threshold model. The authors argued that only the items that were recognised with high confidence contained diagnostic source information, and that the linearity of source ROCs observed by Yonelinas (1999) was an artifact of collapsing across all recognition confidence ratings, and was thus not evidence for a recollection threshold.

Yonelinas and Parks (2007) responded to the Slotnick and Dodson (2005) analysis by proposing that source ROCs are typically linear, but become more curvilinear under a number of conditions. One such condition is when an item and a source are treated holistically as one item, known as *unitised familiarity*, which is continuous. We will return to this point in the Discussion. While this proposal represented a concession towards a continuous contribution under certain circumstances, Klauer and Kellen (2010) were later able to account for curvilinear ROCs using only discrete states by allowing for a variable mapping between recognition confidence ratings and source memory thresholds. At present, then, there is a lack of consensus about whether apparently linear or curvilinear ROCs reflect thresholded or continuous retrieval processes.

# Continuous Report Tasks

Harlow and Donaldson (2013) addressed the need for more diagnostic data by replacing binary decision outcomes with a continuous report paradigm, which yielded a continuous measure of response accuracy. In the Harlow and Donaldson (2013) continuous report paradigm, source information was provided by a point located on the circumference of a circle, which represented the context, and which was paired with a word item. When later cued with that word, participants were required to reproduce the associated location. This procedure allowed for a continuous measure of the error in the angle between the reported and true source locations. The researchers’ use of a continuous measure of source memory performance allowed them not only to measure the accuracy of source memory judgments, but also the distribution of response errors. Instead of categorizing responses as either correct or incorrect as in a two-choice task, their task, which captures an entire distributions of response accuracy, provides a more detailed picture of trial-to-trial variability in retrieval performance. The additional information in such distributions may be more diagnostic than ROC curves of the underlying retrieval processes. Critically, the threshold and continuous models of source memory make divergent predictions about the distributions of response errors in continuous report tasks.

According to the threshold model, items that fall below the recollection threshold will result in guesses, which will be distributed uniformly across all possible response options. Items that exceed the threshold and are successfully retrieved will cluster, with some error, around the true value of the item source. This two-process account of continuous report performance parallels similar proposals in the visual working memory literature, where Zhang and Luck (2008) used a two-component mixture model comprised of a von Mises distribution and a uniform distribution to argue for an item-capacity-limited model visual working memory. The von Mises distribution is a circular analogue of the Gaussian distribution and, like the Gaussian distribution, has a bell-shaped density function. Items in memory are represented with high accuracy and responses to them follow a von Mises distribution; items not in memory lead to guessing and responses to them follow a uniform distribution.

Harlow and Donaldson (2013) took a similar approach in modelling performance in their source memory task, using a wrapped Cauchy distribution to characterize the shape of the marginal distribution of response errors when items exceeded the retrieval threshold. The wrapped Cauchy distribution differs from the von Mises distribution in that its shape is more leptokurtic, with a higher peak and heavier tails. mixture of a wrapped Cauchy distribution and a uniform distribution produces a high-peaked, heavy-tailed distribution (Harlow & Donaldson, 2013). Harlow and Donaldson (2013) found that source accuracy data was better fit by the threshold model better than by its continuous counterpart, which assumes that all responses follow a single wrapped Cauchy distribution, predicts that responses made with moderate memory strength would result in a wider spread of responses around the true location without a uniformly distributed guessing component.

## Source Memory for Unrecognized Items

Although Harlow and Donaldson’s (2013) method represents an innovative way to charactertize the retrieval processes in source memory tasks, a limitation of their experimental design was that participants’ source judgments for items were not conditioned on their confidence in a separate item recognition decision. . Slotnick and Dodson (2005) showed how source memory ROC shapes depend on recognition confidence in the two-choice paradigm, in that unrecognized items have no source discriminability, and it is possible that continuous source memory judgments are affected in a similar way. Hautus, MacMillan, and Rotello (2008) modeled performance in two-choice source memory tasks using a multivariate signal detection model in the style of Banks (2002), and showed that unrecognized items have no source discriminability. Their findings mirrors those of Slotnick and Dodson (2005) who found that source performance was at chance for items recognised with low confidence. Additionally, Hautus et al. (2008) found that the decrease in performance with decreasing confidence was too abrupt to be captured by a continuous model. The abrupt shift to chance performance was, however, able to be captured by a model with a guessing process.

A lack of source discriminability for unrecognized items has been replicated numerous times (Bell, Mieth, & Buchner, 2017; Malejka & Broder, 2016; Onyper, Zhang, & Howard, 2010), although these studies often employed a procedure where item and source ratings were obtained in the same test trials. When item recognition and source memory tests were in separate blocks, Osth, Fox, McKague, Heathcote, and Dennis (2018) observed reliable source memory for unrecognized items, but discriminability was still low (*d’* ~ .1).

If the lack of source memory for unrecognized items generalizes to continuous report tasks, then guesses would result in a heavy-tailed error distribution, which would not necessarily reflect a threshold in memory retrieval but might simply reflect a state in which source retrieval was not attempted. In the context of the findings of Harlow and Donaldson (2013), this account of apparent guessing behavior predicts that if unrecognized items are excluded, the heavy tails in the error distribution will disappear, and that a continuous model will be preferred in account for source performance. An aim of our study was therefore to investigate a continuous-report measure of source-memory performance conditional on the accuracy of previous recognition judgments. In order to do this, we must consider the mapping between retrieved information and an observed response through the lens of a decision model.

# Insights from Models of Decision-Making

In completing a source or recognition memory task, not only do participants need to retrieve information from memory, they must also make a decision on how to respond based on the information retrieved (Ratcliff, 1978). Critically, both the decision outcome and the latency of the decision offer insight into the decision-making process. Much of the existing body of source memory research, particularly in the continuous report paradigm, lacks an explicit account of properties of the decision process. Past research in the recognition memory literature has shown that when the properties of decision processes as well as response latencies are taken into account, the kinds of conclusions that can be made about recognition memory differ from those made when decision-making is not explicitly considered (Dube, Starns, Rotello & Ratcliff, 2012; Osth, Bora, Dennis, & Heathcote, 2017; Osth & Farrell, 2019; Ratcliff & Starns, 2013; Starns, Ratcliff, & McKoon, 2012).

Diffusion models have emerged as increasingly influential accounts of decision processes which predict both response time (RT) and response accuracy data, and which naturally explain well-documented phenomena like the speed-accuracy trade-off (Ratcliff, Smith, Brown & McKoon, 2016). Diffusion models have also been used extensively in the past to model memory retrieval, and more recent research has proposed a general theory of memory and decision-making in which decisions about stimuli within recognition memory and visual working memory are made using a diffusion process (Osth, Jansson, Dennis, & Heathcote, 2018; Smith & Ratcliff, 2009). In the most common form of the diffusion model (Ratcliff, 1978), the decision process is modeled as noisy evidence accumulation between a pair of absorbing boundaries that represent the decision criteria for the task. Evidence is accumulated until the process reaches one or other boundary: The first boundary reached determines the response and the time to first reach a boundary is the decision time component of RT. The diffusion decision model in shown in Figure 1.

A close up of a map

Description automatically generated

Figure 1. Diffusion decision model. Evidence is accumulated by a Wiener diffusion between a pair of absorbing boundaries that represent the decision criteria for responses Ra or Rb. The starting point is z and the boundaries are located at 0 and a. The first boundary reached determines the response and the time taken to reach it determines the decision time. The rate at which evidence accumulates is the drift rate, which is normally distributed across trials with standard deviationη.

The diffusion model assumes that multiple sources of variability affect the decision process, including moment-to-moment variability in the accumulation of evidence and trial-to-trial variability in the quality of evidence entering the decision process. The moment-to-moment variability reflects the noisiness of the evidence provided by the retrieval process, while the trial-to-trial variability reflects differences in the stimulus information on which the decision is based. The rate of evidence accumulation on any trial is known as the drift rate. Drift rates can vary across trials, with high drift rate trials resulting in high accuracy and fast RTs, while trials with lower drift rates result in slower and less accurate responses (Ratcliff, Smith & McKoon, 2015).

The relationship between accuracy and RT is intuitive: when the evidence for the correct response is strong, it will accumulate more rapidly, leading to faster responses, and be more likely to reach the correct boundary before the error boundary, leading to higher accuracy. In contrast, when the evidence is weak, it will accumulate more slowly, leading to slower responses, and be less likely to reach the correct boundary before the error boundary, leading to lower accuracy. When the drift rate varies between trials, leading to a mixture of strong and weak evidence, then the mean RT for correct responses will be shorter than mean RT for errors. This is because most error responses come from trials with low drift rates, which have slow RTs, while most correct responses come from trials with high drift rates, which have fast RTs. Without this variability, correct and error RT distributions will be the same; it is only when drift rates vary between trials that the model predicts this slow error pattern. This phenomenon, known as a *slow error* pattern,has been reliably observed when decision making is difficult (Ratcliff et al., 2016).

The circular diffusion model, of Smith (2016) extends the two-choice diffusion model of Ratcliff (1978), which represents decision-making as a one-dimensional evidence accumulation process (diffusion on a line), to account for continuous report tasks. In the circular model, the drift rate is defined as a vector in a two-dimensional (2D) space having both an identity component for the position of the stimulus on the circle, represented as the phase angle of the drift vector, and a magnitude component, representing the quality of the evidence for any particular response (Figure 2). When a response is made, the magnitude of the drift vector determines RT in the same way as does the scalar drift rate does in the standard Ratcliff model, while the point at which the evidence accumulation process exits the circle determines the response outcome.

A close up of a device

Description automatically generated Xθ

Figure 2. Circular diffusion model of continuous report. Evidence is accumulated by a two-dimensional Wiener diffusion on the interior of a disk, whose bounding circle, of radius a, represents the decision criterion. Evidence is accumulated starting at the origin until the process hits the bounding circle. The hitting point, Xθ, is the decision outcome and the hitting time, Ta,is the decision time. The drift rate is vector-valued and consists of two components, (μ1, μ2), which jointly specify its magnitude and direction. In polar coordinates the magnitude is represented by the drift norm ||μ|| and direction is represented by the phase angle θμ The noisy sample path represents evidence accumulation on a single experimental trial.From P. L. Smith (2016). “Diffusion theory of decision making in continuous report’ Psychological review, 123, 425-451. Figure 2. Copyright American Psychological Association.

The properties of the circular diffusion model closely parallel those of the two-choice diffusion model. When the only source of variability in the model is moment-to-moment variability in the evidence accumulation process the model predicts that decision times will be the same for all decision outcomes. When there is across-trial variability in drift rates, the model predicts that accurate responses will be faster than inaccurate responses. When there is across-trial variability in decision criterion, represented by variability in the diameter of the bounding circle, the model predicts that accurate responses will be slower than inaccurate responses. These properties are continuous counterparts of the slow error and fast error properties predicted by the two-choice diffusion model with across-trial variability in drift rate and starting point, respectively.

Mathematically, the circular diffusion model predicts that, for a fixed drift rate and decision criterion, the distribution of decision outcomes will follow a von Mises distribution. The spread of decision outcomes predicted by the von Mises distribution depends on a precision parameter, κ. Precision is the inverse of variance: high precision represents low variance and vice versa. The von Mises precision predicted by the circular diffusion model is jointly a function of the drift norm, ||μ||, the decision criterion, *a*, and the noise in the evidence accumulation process, σ2. Specifically,

In words, this equation says that precision is equal to the quality of the information in the stimulus, represented by the drift norm, multiplied by the amount of evidence required for a response, represented by the decision criterion, divided by the noisiness of the evidence accumulation process. This relationship between precision, strength of evidence and the decision criterion is a key feature of the circular diffusion model which motivates our application of the model to the source memory task. Analysis of precision without consideration of RTs miss a critical decision effect in this relationship

An important property of the circular diffusion model for this study is that, while a fixed drift rate predicts a von Mises error distribution, the introduction of trial-to-trial variability in drift rate leads to peaked, high-tailed distributions, similar to those found by Harlow and Donaldson (2013) and in the visual working memory literature (Zhang & Luck, 2008). The peaked high-tailed distributions are the result of mixing trials that have high and low drift rate norms. High and low drift rates lead to error distributions with high and low precision, respectively. Mixtures of high and low precision von Mises distributions lead to peaked, heavy-tailed distributions like those found experimentally (van den Berg, Awh, & Ma, 2014). Harlow and Donaldson interpreted these kinds of distributions as evidence of an underlying memory retrieval threshold. We propose that instead, it is possible that heavy-tails reflect a continuous mixture of trials with high and low drift rate.

# The Current Study

There are two aims in the current study. The first is to investigate whether the heavy-tailed distributions found by Harlow and Donaldson (2013) could have been the result of source guessing for unrecognized items. We did this by investigating source memory performance conditional on confidence in the recognition task. If heavy-tailed distributions of errors are due to source guessing on unrecognized items, then they should be eliminated on trials on which recognition confidence is high.

Our second aim is to model with circular diffusion model to determine the extent to which RTs constrain the results. Using the circular diffusion model provides an elaborated account of the decision-making process, and accounts for performance in continuous report source memory tasks without requiring a retrieval threshold mechanism. In doing so, we aim to provide an account of both response times and response error distributions in such a task.

Our experimental task also included a manipulation of the imageability and concreteness of the stimulus words, as rated on the MRC Psycholinguistic Database. Harlow and Donaldson (2013) selected words for low ratings on both metrics to prevent participants from visualizing a concrete object in a source location. In our study, we drew stimuli from pools of low and high imageability and concreteness words which allowed us to quantify and compare the effect of these attributes. This experimental manipulation was intended to allow us to examine the effect of these stimuli attributes on source judgements, however, we did not observe a difference between these conditions. We fit models to each condition separately but in the interests of presentability, data and model predictions are averaged in the figures displayed in this paper.

# Method

## Stimuli and apparatus

Stimuli were presented on a 20’’ Dell 2009W LDC Monitor, set with a screen refresh rate of 60 Hz. Software written in MATLAB using PsychToolbox controlled stimulus presentation and recorded responses. Stimuli consisted of words generated from the MRC Psycholinguistic Database, selected/ for low concreteness (minimum 100, maximum 456) and imageability (minimum 100, maximum 481) in the low stimulus set, and high concreteness (minimum 543, maximum 611) and high imageability (minimum 545, maximum 609) in the high stimulus set. Words were displayed in size 24 point “Courier New” white font positioned in the center of a uniform mean luminance field.

## Participants

Twenty participants were recruited online through the University of Melbourne SONA system. Each participant was expected to complete four 60-minute sessions, for which they were paid $12 at the completion of each session. One participant who did not complete all four sessions was excluded from analysis (*N* = 19). All participants were provided with plain language statements and consent forms, and gave informed consent prior to data collection.

## Procedure

Participants completed the experimental tasks over four sessions, Each of the four sessions consisted of 180 trials, which was broken up into 18 blocks of 10 items each. Blocks were comprised of a study phase, followed by a test phase. In the study phase, participants were presented with a black cross on a dark gray outline at the start of each trial for 600 ms, which was followed by the display of a word in the centre of the screen for 1500 ms. To ensure that participants attended to the source information, they were instructed to indicate the previous location of the cross on the blank target circle using a computer mouse. Responses made within 6 degrees of the true target location were classified as attended and advanced participants to the next item. Responses further away were deemed unattended and the words “TRY AGAIN” was displayed for 1000 ms, then the location was then re-presented for 250ms, and the verification task was repeated. Participants were then instructed to complete a distractor task, which involved 30 seconds of arithmetic problems. Following this, participants were shown a scrambled list of 10 previously studied items and 10 foils, and asked to rate each item on a six-point confidence Old/New scale. Finally, in the source memory retrieval task, participants were cued with the words for 1500 ms, and then indicated the recalled location by a clicking a mouse on the circumference of a grey response circle. There was no time limit on the decision task. A schematic for one trial in each of the phases is shown in Figure 3.

A picture containing screenshot

Description automatically generated

Figure 3. Schematic of display presented to the participant in one trial in each phase of the experiment. There was also an arithmetic distractor task between the encoding and recognition phases of the block, which is not depicted in the figure. The mouse cursor is shown in the center of the “Source Task” panel to illustrate the procedure. In the actual experiment, cursor was hidden from the participant and replaced with a red dot with a diameter of four pixels.

# 

# Results

The results are presented in five parts. First, we ascertain whether individual participants’ responses in the source retrieval task were above chance. As responses were made on a continuous scale, above-chance performance translates into a deviation from uniformity in responding. We did this in order to distinguish participants who were responding at chance from those who showed better-than-chance source memory performance. Second, we investigate source memory judgments conditioned on the prior recognition response and show that conditioning source responding on successful recognition does not fully account for the heavy tails in the distribution of source memory accuracy. Third, we fit the Zhang and Luck (2008) mixture model to the marginal distributions of response accuracy, conditioned on recognition performance, and show that recognition affects the precision of the source information that is retrieved and not the proportion of guessing responses. Finally, we present fits of versions the circular diffusion model to the joint distributions of RT and accuracy. This analysis generalizes the Zhang and Luck (2008) mixture model analysis to account for both speed and accuracy.

**Data Screening**

Preliminary inspection of the data suggested that some participants performed the source retrieval task with very low accuracy. A Rayleigh test for uniformity identified two participants whose data showed no evidence for a departure from uniformity in at least one condition, which can be interpreted as completely random responding (Table X; Fisher, 1993). These participants will be referred to as a *low response accuracy* subgroup. We identify this subgroup with the expectation that the data from the remaining *high response accuracy* group will be more diagnostic for the purposes of distinguishing between the models.

|  |  |
| --- | --- |
| Table X  Rayleigh Test for Uniformity for Source Memory Response Error | |
| Participant | *p* |
| 1 | .87\* |
| 2 | <.01 |
| 3 | <.01 |
| 4 | <.01 |
| 5 | <.01 |
| 6 | <.01 |
| 7 | <.01 |
| 8 | <.01 |
| 9 | .01 |
| 10 | <.01 |
| 11 | <.01 |
| 12 | <.01 |
| 13 | .24\* |
| 15 | .04 |
| 16 | <.01 |
| 17 | <.01 |
| 18 | <.01 |
| 19 | <.01 |
| 20 | <.01 |

\* *p* values greater than 0.05, indicating no evidence of a departure from uniformity for participants 1 and 13.

## Source Memory for Unrecognized Items

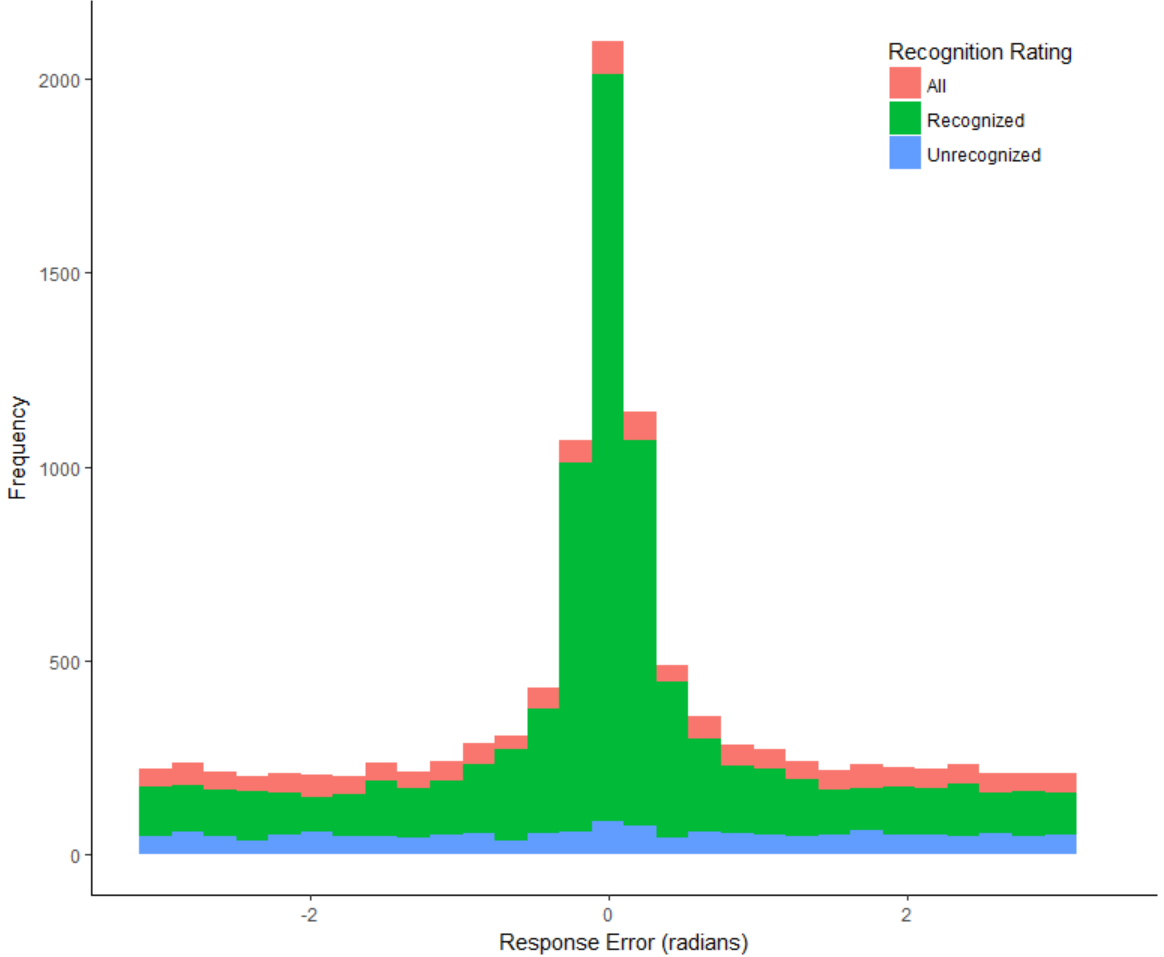
The data for each participant were split into three categories on the basis of participants’ confidence in the recognition phase of the experiment. Items which were rated three and below were deemed unrecognized by the participants, while successful recognition was defined by a rating of four and above. Of the recognized items, ratings of the maximum value of six are further specified as highly recognized items. Figure 4 shows the frequency of response errors across all participants grouped according to these categories of confidence in the recognition phase.

Figure 4. Frequency of angular response error in the source retrieval task, collapsed across participants. The subset of unrecognized items (rated three or below on the six-point confidence scale) yields source responses which are uniform, indicating no source memory for unrecognized items. However, the frequency of unrecognized items was relatively low and the exclusion of these items from the distribution of responses (in red), does not eliminate the heavy tails from the distribution of responses for recognized items (in green).

The results of Rayleigh tests of uniformity of source responses for unrecognized items are displayed in Table 4. The distributions of these responses were uniform for all participants, indicating that no source memory was present when recognition confidence was low. Participants

|  |  |
| --- | --- |
| Table 4  Rayleigh Test on Source Memory for Unrecognized Items | |
| Participant | *p* |
| 1 | 0.46 |
| 2 | 0.38 |
| 3 | 0.16 |
| 4 | 0.48 |
| 5 | 0.96 |
| 6 | 0.30 |
| 7 | 0.75 |
| 8 | 0.17 |
| 9 | 0.80 |
| 10 | 0.07 |
| 11 | 0.01\* |
| 12 | 0.40 |
| 13 | 0.62 |
| 15 | 0.02\* |
| 16 | 0.06 |
| 17 | 0.76 |
| 18 | 0.02\* |
| 19 | 0.17 |
| 20 | 0.44 |

**Item Recognition Performance**

An independent samples *t*-test applied to individual-level hit rates for high and low imageability conditions (Table 5) indicated that there was no significant difference in hit rates across the two conditions *t*(35.85) = .68, *p* = .503. Coupled with the Rayleigh test on unrecognized items, this suggests that although source performance for unrecognized items was uniform, the majority of items were successfully recognized, and so guessing due to recognition failure does not fully account for the heavy tails of the error distributions.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table 5  Item Recognition Hit Rates and False Alarms | | | | |
| Condition | Hit Rate | | False Alarms | |
|  | *M* | *SD* | *M* | *SD* |
| High Imageability | .87 | .14 | .14 | .10 |
| Low Imageability | .84 | .13 | .14 | .11 |

**Mixture Model**

To attempt to replicate the Harlow and Donaldson (2013) finding, we first used the Zhang and Luck (2008) mixture model to fit the marginal distribution of response error. The model had two free parameters, one for the von Mises precision, which described the spread of responses around the true location, and a mixing parameter *π*,which described the proportion of trials which were driven by information in a von Mises distribution, as opposed to guesses in a uniform distribution. The mean best fitting parameters of the mixture model to the response accuracy data, excluding the low response accuracy group, are shown in Table X. The parameter estimates at an individual level are in Appendix X.

If heavy-tailed distributions of source errors were attributable to guessing for unrecognized items, then we expect that any difference between conditions to be reflected in the mixing parameter. Instead, we observe that conditioning on recognition largely affects the precision parameter and not the memory parameter, which further suggests that source guessing for unrecognized items does not sufficiently account for the heavy-tailed properties in the source error data.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table X | | | | |
| *Parameter Values for Best Fits of the Simple Mixture Model to Highly Recognized Individual Data.* | | | | |
| Recognition Rating | Low Imageability | | High Imageability | |
|  | Precision | *π* | Precision | *π* |
| All | 19.22 | 0.51 | 23.89 | 0.51 |
| Recognized | 18.24 | 0.50 | 23.79 | 0.54 |
| Highly Recognized | 18.03 | 0.50 | 19.81 | 0.56 |

*Note.*  Precision refers to the precision of the information-driven retrieval process. *π* represents proportion of responses driven by information.

**Circular Diffusion Models**

While the mixture model was fit only to response error, we now also take latency into account by jointly modelling error and RT with the circular diffusion model. This allows us to decompose precision into drift norm and decision criterion. With trial-to-trial variance in drift norm, the circular diffusion model can produce heavy tailed distributions, which may allow an alternate account for the observed patterns in the data without a discrete guessing process. As we observed no difference between the summary statistics for the high and low imageability conditions, we pooled data between these conditions together and fit the model to this unified dataset.

We tested three alternative versions of the circular diffusion model that expressed different hypotheses about the process of memory retrieval. The first of these was designed to be analogous to the continuous model of source memory presented in Harlow and Donaldson (2013), and was implemented as a circular diffusion model with across-trial variability in drift rates. We refer to this variant as the continuous diffusion model, in which drift rate variability was set to be equal in both dimensions of the two-dimensional (2D) space. Mean drift rate was described by the parameter *μ,* which varied along a Gaussian distribution with standard deviation *η.*

The decision criterion was represented by *a*. Finally, there was a non-decision time parameter, *Ter­* , and non-decision time variability *st*. Like the standard diffusion model, the circular model assumes that RT is the sum of the decision time and a time for other (encoding and response) processes. We used the onset of the response circle to begin timing in the source retrieval task, however, participants may have started to retrieve information prior to the onset of timing, during the display of the cue word immediately prior to the response circle. For this reason, we allowed *Ter­* to be negative to account for this premature engagement with the source retrieval task.

The second model variant embodied the thresholded retrieval property preferred by Harlow and Donaldson (2013), which we call the threshold diffusion model. This was implemented as a mixture of two diffusion processes: one with positive drift rate and no between-trial drift variability, and a second that was modeled as a diffusion process with zero drift rate. The zero-drift process provides a diffusion process implementation of a guessing process, in which the decision process is driven only by noise (Smith, Saber, Corbett, & Lilburn, in press). Unlike guessing paramteres in the models mentioned in the introduction which account for accuracy but not RT, the zero-drift process is able to predict both accuracy and RT. Mixing proportions for the two processes were allowed to vary between the imageability conditions. This model had six free parameters. The mean drift rate parameter was shared with the continuous model (*μ*), with the same interpretation, as well as *Ter­,* the non-decision time parameter and non-decision time variability *st*. The mixing proportion between information-driven and guessing processes was represented by *π.* The decision criterion was estimated separately for the information-driven component (*a1*) and the guessing component (*a2*).

The third model was a combination of the continuous and threshold diffusion models. It assumed a mixture of zero-drift and nonzero-drift processes, like the threshold diffusion model, but also allowed for across-trial variability in drift rates, similar to the some-or-none model of Onyper, Zhang, and Howard (2010). This model, which we name the hybrid diffusion model, incorporates both the continuous and threshold diffusion models in that it had a mixture of information-driven and guessing processes as well as trial-to-trial drift rate variability. Thismodel had seven free parameters, all of which are shown in Table X.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table X  Symbols and definitions of free parameters estimated in diffusion model variants | |  |  |  |
| Symbol | Parameter | Inclusion in Model | | |
| Continuous | Threshold | Hybrid |
| *μ* | Mean drift | Y | Y | Y |
| *η1* | Drift variability | Y | N | Y |
| *a1* | Decision criteria, information-driven component | Y | Y | Y |
| *a2* | Decision criteria, guessing component | N | Y | Y |
| *π* | Mixing proportion | N | Y | Y |
| *Ter* | Non-decision time | Y | Y | Y |
| *st* | Non-decision time variability | Y | Y | Y |

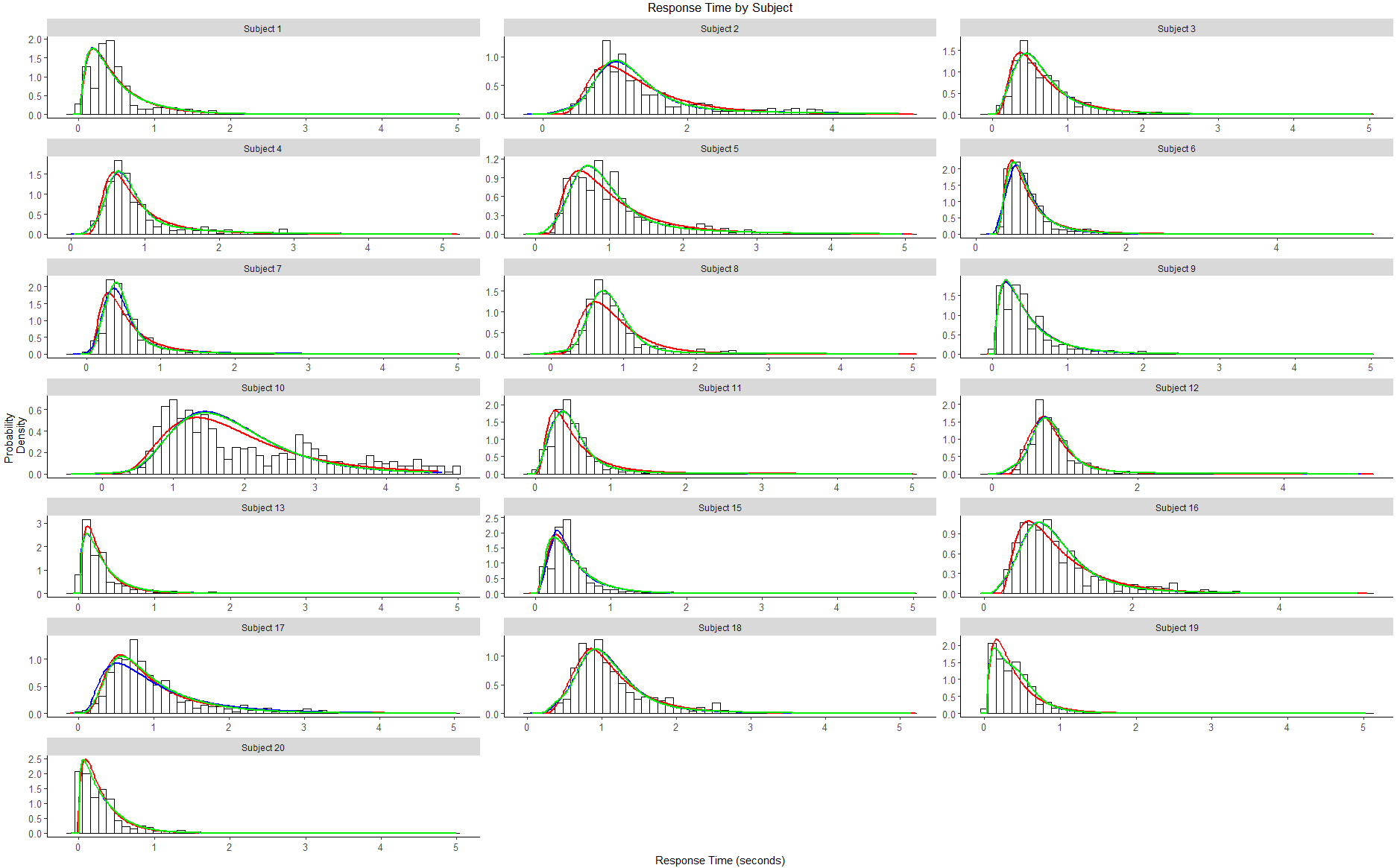
*Note.* Not all parameters were estimated for all three models. The continuous diffusion model did not include a mixed guessing process, and therefore lacked *a2*, *π1* and *π2.* The threshold diffusion model did not have drift variability and lacked *η1* and *η2*. The hybrid diffusion model included all parameters listed.

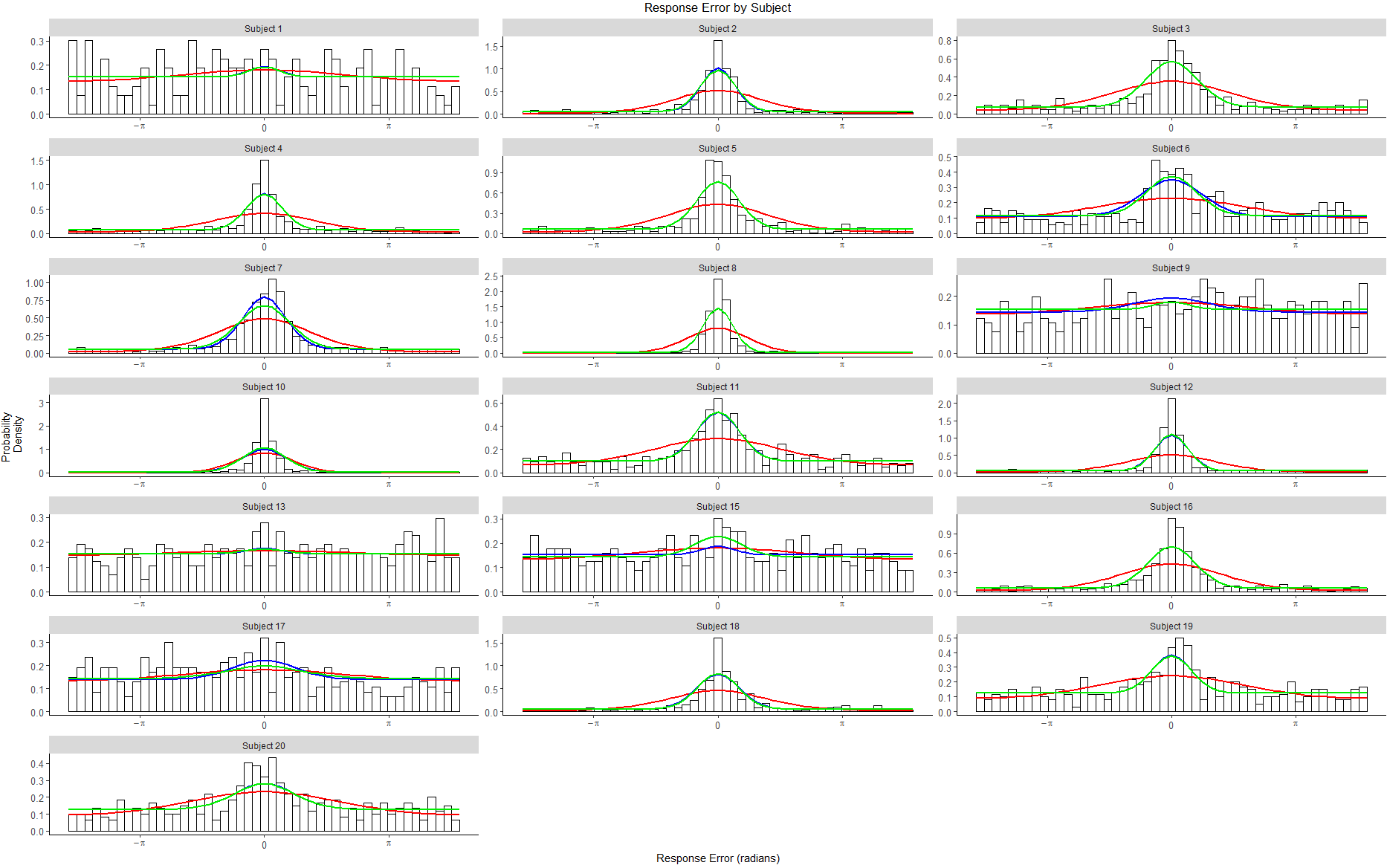
The three variants of the circular diffusion model were each fit using maximum likelihood estimation to data on trials that were highly recognized (rated four or higher in the item recognition phase) at an individual level. We excluded data for each participant that were deemed too fast or slow, which was defined as being beyond three standard deviations of the median RT for that participant. The Bayesian Information Criterion (BIC) and Log Likelihood (LL) for the three models’ fits to each participant is shown in Table X.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table X | | | | |
| Bayesian Information Criterion (BIC) values for Fits of the Models to Individual Data | | | | |
| Participant | | Continuous | Threshold | Hybrid |
| High Precision | 2 | 2211.62 | 1974.04 | **1973.43** |
|  | 3 | 1875.35 | **1743.92** | 1749.93 |
|  | 4 | 3724.85 | **3158.48** | 3162.36 |
|  | 5 | 2180.82 | **1985.66** | 1991.72 |
|  | 6 | 1581.47 | **1529.58** | 1530.76 |
|  | 7 | 1448.65 | **1275.30** | 1278.28 |
|  | 8 | 1222.85 | **606.70** | 620.79 |
|  | 9 | **2060.22** | 2065.52 | 2068.87 |
|  | 10 | 1974.49 | 1857.13 | **1832.28** |
|  | 11 | 1821.20 | **1638.65** | 1648.34 |
|  | 12 | 1479.85 | **1020.29** | 1024.03 |
|  | 15 | **1880.61** | 1884.47 | 1889.55 |
|  | 16 | 1965.88 | 1824.08 | **1823.12** |
|  | 17 | **1978.34** | 2001.04 | 2035.88 |
|  | 18 | 2062.39 | **1832.53** | 1834.55 |
|  | 19 | 1620.87 | **1545.20** | 1551.37 |
|  | 20 | 1602.36 | 1570.42 | **1566.27** |
|  |  |  |  |  |
| Low Precision | 1 | **885.60** | 886.72 | 892.04 |
|  | 13 | **1566.15** | 1573.21 | 1579.12 |

Lowest BIC for each participant is indicated in boldface

Both the threshold and the hybrid models consistently outperformed the continuous model without guessing, particularly for those participants in the high precision subgroup. Of those participants whose data was best fit by the continuous model, the margin of this advantage over the threshold and hybrid models was small, relative to the margin when the threshold and hybrid models perform better than the continuous model. Fits of the models to response error (Figure X) and RT (Figure X) data show where the continuous model misses the data. For participants that respond successfully to the task with a reasonable level of consistency (with a response error distribution that does not appear uniform), the continuous-strength model underestimates the proportion of responses with very high precision and overestimates the proportion of responses with a moderate level of precision. In contrast, the threshold and hybrid variants capture the general structure of both response error and RT data, although both models struggle to predict the extreme peaks in the error distribution and the leading edges of the RT distributions.





The reason for this behavior can be seen in the joint density plot (Figure X), which shows the RT quantile predictions for each model over response error quantiles. The continuous model is able to produce a leptokurtic shape similar to the threshold models through trial-to-trial variability in drift norm, producing a slow error pattern. This slow error pattern would be reflected in the joint quantile plot with RT quantiles becoming gradually slower as response error increases. However, the data exhibits 

Overall, these results suggest that participants sometimes do respond in a no-information guessing state, which is mixed with a distribution of responses driven by information which is centered on the target location. This suggests that the addition of drift variability does not improve the fit of the threshold model enough to justify the additional complexity introduced into the model.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Table X  *Mean Parameter Values Across Participants for Circular Diffusion Models* | | | | | | | |
| Model Name | Model Parameters | | | | | | |
|  | *μ* | *η* | *a1* | *a2* | *π* | *Ter* | *st* |
| Continuous | 0.92 | 0.52 | 1.46 | - | - | -0.02 | 0.14 |
| Threshold | 3.64 | - | 3.00 | 1.32 | 0.42 | -0.08 | 0.09 |
| Hybrid | 3.66 | 0.18 | 3.20 | 1.30 | 0.41 | -0.07 | 0.06 |

**Non-normal Distributions of Phase Angles**

Schurgin, Wixted, and Brady (2018) argued that the peaked high-tailed distribution of decision outcomes found in visual working memory may be due to a nonlinear scaling of the psychological space, such that the distances between items far from the true value become increasingly compressed. Smith, Saber, Corbett, and Lilburn (in press) used the circular diffusion model to model continuous outcome decisions about the hues of noisy color patches and compared two models of the drift rates in the decision process. One was a two-component model similar to Zhang and Luck's (2008) memory-plus-guessing model, and the other was a continuous model with non-normal phase angles, with similar properties to Schurgin et al.'s nonlinear scaling model. They found that the two models gave almost identical pictures of the evidence entering the decision process: On the majority of trials, the phase angle of the drift rate, which represents the encoded stimulus identity, was clustered around the true value, but on the remaining trials it was distributed uniformly around the circle. We implemented a similar model for our task, with non-normal distributions of phase angles of the drift rate. We found the non-normal model performed similarly to our threshold model and the estimated parameters of the model supported the idea of a threshold or threshold-like retrieval process. We omit the details of this model.

# Discussion

In this article, we had two main aims. Our first aim was to attempt to characterize performance on a continuous report source memory task using a mathematical model of the decision process, the circular diffusion model, to ascertain whether it could predict the distributions of decision outcomes and RT from such a task. In applying the model to this kind of task, we sought to ascertain whether the conclusions of Harlow and Donaldson (2013), which were based only on the distributions of response error, would continue to hold when both error and RTs were taken into account. Our second aim was to ascertain whether Harlow and Donaldson’s (2013) conclusion that source memory is thresholded would continue to hold for memory when conditioned on item recognised with high confidence.

We found evidence suggesting that source memory retrieval is indeed best characterized as a thresholded process. Firstly, we found that even when source responses were conditioned on successful recognition, the marginal distribution of response error was well characterized with the Zhang and Luck (2008) mixture model, consisting of a von Mises and a uniform distribution . This corroborates the Harlow and Donaldson (2013) finding which used a wrapped Cauchy to similarly account for a greater number of very high accuracy and very low accuracy responses, with fewer responses with moderate accuracy than in a wrapped normal distribution. Secondly, in fulfilling our aim of predicting joint distributions of source response error and RT using the circular diffusion model, we found that the threshold and hybrid models, which both assumed a mixture of guessing and memory-based responses, fit the data better than the continuous model which did not.

**Implications for Models of Memory**

In our study, jointly modelling RT along with response error provided evidence for a thresholded model of source memory. Our findings fall within the growing body of work in memory research that suggests the architecture of memory involves a memory strength threshold. Our findings support the Yonelinas (1994) dual-process model and discrete state models of source memory, while posing a challenge to continuous models of source memory (Banks, 2000). It is unclear from our findings whether the “some-or-none” hybrid model (Onyper, Zhang, & Howard, 2010) offers a consistent advantage over the discrete “all-or-none” model represented by the threshold model.

A similar long-standing debate in the working memory literature is whether memory capacity is better described as a number of discrete slots or by a continuous shared resource (Bays, Catalao, & Husain, 2009; Luck & Vogel, 1997; Pashler, 1988; Wilken & Ma, 2004; Zhang & Luck, 2008). Adam, Vogel and Awh (2017) found evidence for a discrete item limit, using a whole report paradigm in which participants performed increasingly worse over three to four items in a six item list, and subsequent responses were best characterized as guesses. In the change detection literature, Donkin and Nosofsky (2016) found RT evidence supporting a hybrid model of visual working memory consisting of a mixture of guessing and memory-driven states. These models fall in a general family of models akin to the Zhang and Luck (2008) “slots + resources” model where a continuous resource is divided among slots, much like performance in our task can be described as variable but thresholded.

**Application of the Circular Diffusion Model**

In this paper we present a novel application of the circular diffusion model to memory data. (Van den berg, 2014) plateaus in memory, such as the uniform distribution in a response error distribution, are statistically meaningless. Looking at RT distribution alleviates this by providing richer data for more constrained models. Because the model has to account for both jointly, including RT distributions in the analysis becomes a more stringent test of the model.

This study presents the first application of the circular diffusion model to continuous report memory data. Unlike other applications of continuous-outcome models in the literature, the circular diffusion model allows the precision of the empirical distribution of decision outcomes to be decomposed into psychologically meaningful components via Equation 1. This equation shows that empirical precision depends jointly on the quality of the information in the stimulus and the amount of evidence needed for a response. The theoretically relevant quantity is the quality of the evidence entering the decision process, and this can only be obtained using a model-based analysis, like the circular diffusion model provides.

**Conclusion**

This study represents the first attempt to model RT data alongside response error in a source memory task with continuous report outcomes. We used the circular diffusion model, which provides for an elaborated account of decision-making, to fit joint RT and error data. This allowed for more constrained analysis than previous studies which accounted only for response error. We found evidence supporting the existence of a threshold in source retrieval, corroborating the overall conclusion of Harlow and Donaldson (2013).

**Appendices**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Table X | | | | | |
| *Parameter Values for Best Fits of the Simple Mixture Model to Recognized Individual Data.* | | | | | |
|  | | Low Imageability | | High Imageability | |
| Participant | | Precision | *π* | Precision | *π* |
| High Accuracy |  |  |  |  |  |
|  | 2 | 24.51 | 0.66 | 17.36 | 0.73 |
|  | 3 | 11.76 | 0.45 | 10.47 | 0.54 |
|  | 4 | 32.44 | 0.48 | 44.47 | 0.54 |
|  | 5 | 18.51 | 0.47 | 16.80 | 0.65 |
|  | 6 | 12.90 | 0.27 | 8.86 | 0.23 |
|  | 7 | 12.17 | 0.65 | 9.11 | 0.69 |
|  | 8 | 37.43 | 0.87 | 42.93 | 0.89 |
|  | 9 | 0.57 | 0.29 | 0.10 | 1.00 |
|  | 10 | 49.40 | 0.87 | 49.08 | 0.87 |
|  | 11 | 7.90 | 0.35 | 13.10 | 0.39 |
|  | 12 | 36.72 | 0.73 | 46.67 | 0.69 |
|  | 15 | 5.39 | 0.13 | 85.49 | 0.04 |
|  | 16 | 15.53 | 0.55 | 11.01 | 0.64 |
|  | 17 | 0.27 | 0.64 | 5.81 | 0.10 |
|  | 18 | 9.77 | 0.69 | 27.61 | 0.59 |
|  | 19 | 14.27 | 0.19 | 5.45 | 0.32 |
|  | 20 | 20.61 | 0.15 | 10.15 | 0.25 |
|  |  |  |  |  |  |
| Low Accuracy | 1 | 157.24 | 0.05 | 250.00 | 0.02 |
|  | 13 | 45.64 | 0.03 | 250.00 | 0.02 |

References

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

Banks, W. P. (2000). Recognition and source memory as multivariate decision processes. *Psychological Science*, *11*(4), 267-273.

Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological review*, *97*(4), 548.

Bell, R., Mieth, L., & Buchner, A. (2017). Emotional memory: No source memory without old–new recognition. *Emotion*, *17*(1), 120.

Bowers, J. S., & Schacter, D. L. (1990). Implicit memory and test awareness. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(3), 404.

Dube, C., Starns, J. J., Rotello, C. M., & Ratcliff, R. (2012). Beyond ROC curvature: Strength effects and response time data support continuous-evidence models of recognition memory. *Journal of Memory and Language*, *67*(3), 389-406.

Harlow, I. M., & Donaldson, D. I. (2013). Source accuracy data reveal the thresholded nature of human episodic memory. *Psychonomic Bulletin & Review*, *20*(2), 318-325.

Hautus, M. J., Macmillan, N. A., & Rotello, C. B. (2008). Toward a complete decision model of item and source recognition. *Psychonomic Bulletin & Review*, *15*(5), 889-905.

Janowsky, J. S., Shimamura, A. P., & Squire, L. R. (1989). Source memory impairment in patients with frontal lobe lesions. *Neuropsychologia*, *27*(8), 1043-1056.

Klauer, K. C., & Kellen, D. (2010). Toward a complete decision model of item and source recognition: A discrete-state approach. *Psychonomic Bulletin & Review*, *17*(4), 465-478.

Malejka, S., & Bröder, A. (2016). No source memory for unrecognized items when implicit feedback is avoided. *Memory & cognition*, *44*(1), 63-72.

Mickes, L., Wais, P. E., & Wixted, J. T. (2009). Recollection is a continuous process: Implications for dual-process theories of recognition memory. *Psychological science*, *20*(4), 509-515.

Onyper, S. V., Zhang, Y. X., & Howard, M. W. (2010). Some-or-none recollection: Evidence from item and source memory. *Journal of Experimental Psychology: General*, *139*(2), 341.

Osth, A. F., Bora, B., Dennis, S., & Heathcote, A. (2017). Diffusion vs. linear ballistic accumulation: Different models, different conclusions about the slope of the zROC in recognition memory. *Journal of Memory and Language*, *96*, 36-61.

Osth, A. F., Fox, J., McKague, M., Heathcote, A., & Dennis, S. (2018). The list strength effect in source memory: Data and a global matching model. *Journal of Memory and Language*, *103*, 91-113.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological review*, *85*(2), 59.

Ratcliff, R., & Starns, J. J. (2013). Modeling confidence judgments, response times, and multiple choices in decision making: Recognition memory and motion discrimination. *Psychological review*, *120*(3), 697.

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in cognitive sciences*, *20*(4), 260-281.

Rouder, J. N., Morey, R. D., Cowan, N., Zwilling, C. E., Morey, C. C., & Pratte, M. S. (2008). An assessment of fixed-capacity models of visual working memory. *Proceedings of the National Academy of Sciences*, *105*(16), 5975-5979.

Schurgin, M. W., Wixted, J. T., & Brady, T. F. (2018). Psychological scaling reveals a single parameter framework for visual working memory. *bioRxiv*, 325472.

Slotnick, S. D., & Dodson, C. S. (2005). Support for a continuous (single-process) model of recognition memory and source memory. *Memory & cognition*, *33*(1), 151-170.

Smith, P. L. (2016). Diffusion theory of decision making in continuous report. *Psychological Review*, *123*(4), 425.

Smith, P. L., & Ratcliff, R. (2009). An integrated theory of attention and decision making in visual signal detection. *Psychological review*, *116*(2), 283.

Smith, P. L., Saber, S., Corbett, E. A. & Lilburn, S. D. (2019). Modelling the speed and accuracy of continuous outcome colour decisions: Metric and categorical effects. Paper presented at the The Australian Mathematical Psychology Conference, Melbourne, Australia, February.. Abstract retrieved from: http://mathpsy.ch/abstracts/smith-26/

Starns, J. J., Hicks, J. L., Brown, N. L., & Martin, B. A. (2008). Source memory for unrecognized items: Predictions from multivariate signal detection theory. *Memory & Cognition*, *36*(1), 1-8.

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(6), 1341.

Yonelinas, A. P. (1999). The contribution of recollection and familiarity to recognition and source-memory judgments: A formal dual-process model and an analysis of receiver operating characterstics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(6), 1415.

Yonelinas, A. P., & Parks, C. M. (2007). Receiver operating characteristics (ROCs) in recognition memory: a review. *Psychological bulletin*, *133*(5), 800.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233.