When we recall a past experience, we often not only retrieve information about an item in memory, but also information about the conditions under which that memory was formed, or the *source* of that memory (Johnson et al., 1993). Episodic memory, which describes memory for events, has been studied experimentally using item recognition and source memory tasks, often in tandem. In a source memory task, subjects are shown stimuli (e.g., words, shapes, or objects) which are presented in some context (e.g., the voice of a speaker, location on a display). When later cued with the item, participants are then asked to report the source. Several models have been advanced to understand the processes governing both recognition and source judgements (Yonelinas, 1999; Slotnick & Dodson, 2005; Hautus et al., 2008).   
 A key question such models contend with is whether the retrieval of information from source memory is better characterized as a continuous or a discrete process. Models of memory retrieval as a continuous process, based upon Signal Detection Theory (SDT), assume that memory strength varies continuously, and so predict that performance in a source memory task declines gradually as memory strength decreases (Banks, 2000; Mickes et al., 2009). In contrast, threshold or discrete-state models assume that memory strength for an item must reach a certain threshold in order for that item to be retrieved, and so predict that source responses are either made with high precision when driven by memory or are guesses, made in the absence of information, when the memory is below the retrieval threshold (Batchelder & Riefer, 1990; Klauer & Kellen, 2010). Another alternative is the dual-process framework, in which different retrieval mechanisms are used in different kinds of memory tasks (Mander, 1980). Specifically, the two processes in the influential Yonelinas (1999) dual-process model are 1) familiarity, which yields a continuous measure of strength for an item in memory and 2) recollection, which yields richer contextual information about the study event through a search process which is thresholded. Successful recollection or familiarity can both contribute to recognition, because familiarity can distinguish between a studied and an unstudied item. On the other hand, familiarity does not distinguish between two items from different sources, which are both studied, and so the Yonelinas (1999) dual-process model predicts that source judgements should be thresholded as they can only be driven by recollection. This dual-process view of memory retrieval holds only if recollection, and therefore source memory performance, can be characterized as a thresholded process. Existing research which attempted to distinguish between continuous and thresholded models of source memory has been based on data from two-choice tasks, whereby confidence ratings and accuracy in two-choice tasks are used to construct Receiver Operating Characteristic (ROC) curves (Yonelinas, 1999; Slotnick & Dodson, 2005). Although the predicted shape of these curves were initially thought to distinguish between continuous and thresholded models, subsequent work found numerous conditions under which the models mimic each other (Yonelinas & Parks, 2007; Klauer & Kellen, 2010).

**Continuous-Outcome Tasks**

An alternative to using two-choice tasks is to use *continuous-outcome* tasks, in which responses are made on a continuous scale. The advantage of using such a task is that it allows direct measurement of response precision, as opposed to the proportion of responses in each of the discrete options in a traditional two-choice task. This richer, continuous measurement is more informative about the nature of mental representations, particularly in terms of the variability of decisions made about these representations (Smith et al., 2020). Continuous-outcome tasks were first used to study memory in the specific context of how visual working memory (VWM) representations change with the number of items stored in memory (Wilken & Ma, 2004). Just as the source memory literature has been concerned with the existence of a discrete subthreshold guessing state, the VWM literature has historically grappled with whether storage capacity is determined by a discrete number of “slots” to be filled, or a continuous resource that can be distributed across an increasing number of items that are represented with decreasing resolution in memory. Zhang and Luck (2008) modelled distributions of response outcomes in a color recall task under different set size conditions, and found the data was well described by a mixture model, specifically a mixture of a von Mises distribution[[1]](#footnote-1) and a uniform distribution.

Applying a similar approach to source memory modelling, Harlow and Donaldson (2013) used a continuous-outcome task in which source was operationalized as the locations of word stimuli along the circumference of a circle. At test, participants reproduced remembered locations on a response circle when cued with each word in the study list. The authors found that a mixture model consisting of a wrapped Cauchy and a uniform component was preferred over a pure wrapped Cauchy model, which was interpreted as evidence for a thresholded retrieval process which yields uniform guesses when memory strength is subthreshold (Harlow & Donaldson, 2013). In summary, both the Zhang and Luck (2008) and the Harlow and Donaldson (2013) analyses attribute observers’ response error to two sources in memory: 1) variability in memory precision and 2) the possibility that memory is absent and the response is a guess. A further insight from the continuous-outcome VWM literature that is relevant to our interest in source memory is that there is another source of error to be considered: the possibility that the observer responds with information about a different item than the one probed, referred to as a *non-target* response (Bays, 2016).

## Non-target Responding

The tendency for subjects to respond to non-target features or items has been observed in a wide variety of cognitive tasks, and the related types of errors that arise are referred to by various terms including *binding, transposition, intrusion,* and *swap errors*, each reflecting specific properties of the tasks used to study the phenomenon (Bays, 2016). The key distinction between a guess and an intrusion response is that a guess is generated in the total absence of information, while a non-target response is motivated by information for the wrong item. In this study, we consider to what extent the inclusion of a process for intrusion responses attenuates, or even eliminates, the need to invoke a uniform guessing process to account for error patterns in source memory and by extension, the theoretical justification for a thresholded model of source memory retrieval.

Most explanations of non-target responding attribute the phenomenon to confusion between items that are similar in one or more domain (Rerko et al., 2014; Bays, 2016; Oberauer & Lin, 2017; but see Pratte, 2018 for an alternative explanation). In the experimental paradigm we use to study source memory (lists of words presented in space), items may be similar in several ways, including the position of items in the study list i.e., the temporal similarity of the presentation of items, the spatial similarity in the location of the items, as well as features of the words used as stimuli such as their semantic and orthographic similarity. Here, we give a brief review of work which has investigated the effect of these different forms of similarity affect performance in a range of tasks.

### Temporal Contiguity

In serial recall tasks, where subjects must call lists of items in the sequence in which they are given, a classic finding is that incorrect responses tend to be items studied near the target in the study sequence (Lee & Estes, 1977). A similar effect of serial order has been observed in free recall tasks: when participants are asked to recall a list of items in any sequence, Kahana (1996) demonstrated that after recalling a given item, the next item to be recalled tends to be a neighboring item in the study sequence. Additionally, neighbors in the forwards direction were more likely to follow and item than backwards neighbors, referred to as forward asymmetry (Kahana, 1996). While presentation order has different effects on serial and free recall tasks (in the former, recalling neighbors is a type of error, while in the latter it can facilitate successful recall of lists), these examples illustrate the principle of *temporal contiguity*, that is that events that occur close in time become associated with each other (for a review concerning the temporal contiguity effect in free recall, see Healey et al., 2018).

In the most proximal   
Popov, So and Reder (2021) found that participants make an error due to failing to retrieve a word-location binding, they do not respond with a random non-target. Instead, locations for items presented in closer serial order proximity were more likely than locations for items from further away.

While estimated using free parameters, observe a pattern akin to the contiguity effect. We aim to systematically model this effect of serial position with a temporal similarity confusability account.

### Spatial Contiguity

.

Bays (2016) items closest in spatial proximity are swapped.

Spatiotemporal transposition gradient (Renko, oberauer). Visual working memory, simulataneous presentation of

Previous study of memory for spatial location has found that errors were composed entirely of swap errors, with little evidence of guess responses (Rajsic & Wilson, 2012, 2014; Pertzov et al., 2012). When intrusions between items in the source location task are accounted for in a similar way, the prior research would suggest that the contribution of uniform guesses should be similarly low, and that most errors should arise due to intrusions, or swap errors.

**Decision-Making in Continuous-Outcome Tasks**

* Why it be important
* Another source of variability in the task.

**Modelling Decision-Making using the Circular Diffusion Model**

While advantageous, a significant obstacle in using continuous-outcome tasks to model memory performance was that until recently, no formal models of response times (RT) and decision-making existed to account for decisions made on a continuous domain. Accounting for the decision-making process in generating a response to any task is crucial to understanding the underlying cognitive processes of interest in that task, such as memory retrieval. Starns etc.

In this study, we extend this line of reasoning to consider what experimental factors affect the probability of a non-target generating an intrusion response. At a broad level, as similarity between two items in memory increases, so does confusability and the probability of an intrusion. Previous studies have examined the contribution of similarity in terms of temporal context

With word stimuli, attributes of the words used, such as the semantic and orthographic similarity

Ultimately, our goal in introducing more sophisticated models of intrusion responding is to see to what extent the intrusion component mitigates, or even eliminates, estimations for the proportion of no-information uniform guessing.

## Description of Models

Our modelling approach was to start with a two-component mixture model equivalent to Zhang and Luck (2008), and then introduce successive elaborations on the intrusion component to make it sensitive to similarity first in terms of temporal, then spatial similarity of presentation, and finally semantic and orthographic features of the stimuli. The same stepwise process was also taken with the circular diffusion model, using the same calculations to weight intrusion probability by the various kinds of similarity, using the Zhou et al. (2021) thresholded circular diffusion model as a base in place of the Zhang and Luck (2008) model. The models are formally described in the sections to follow, and the key differences between models are summarized in Table X

|  |  |  |
| --- | --- | --- |
| Model | Guessing | Intrusions |
| 1 | Yes | None |
| 2 | No | Flat |
| 3 | Yes | Flat |
| 4 | Yes | Temporal |
| 5 | Yes | Spatiotemporal |
| 6 | Yes | Saturated |

### Model 1: Pure Guessing

As previously described, the Zhang and Luck (2008) model expresses the idea that responses are generated from a mixture of two process, one which is target-driven with Gaussian error, and another which is driven by guesses made at random:

|  |  |
| --- | --- |
|  | (1) |

where in the target-driven component, represents the target angle, is the reported angle, and represents a Von Mises distribution with a mean of 0 and a standard deviation of . The probability that a response is a guess is represented by .

### Model 2: Pure Intrusions

To test the strong prediction that all non-target responses can be accounted for with intrusions from non-target items without invoking any uniform guessing, Model 2 substitutes the guessing component in the mixture model with an intrusion component:

|  |  |
| --- | --- |
|  | (2) |

where the probability of an intrusion occurring is represented by , and the angle associated with the *i*th intruding item is represented by . Note that of the *m* non-target items, the probability of a particular non-target intruding is equal.

### Model 3: Intrusions + Guessing

Model 3 combines intrusion and guess responses in the three-component model of Bays et al. (2009):

|  |  |
| --- | --- |
|  | (3) |

### Model 4: Temporal Similarity Gradient

In contrast to Models 2 and 3 in which each intrusion is equally weighted (that is, the likelihood of each intruding item is divided by the number of intrusions), in Model 4 the weight of each intruding item is determined by its temporal similarity to the target represented by *t:*

|  |  |
| --- | --- |
|  | (4) |

|  |  |
| --- | --- |
|  | (5) |

which is an exponential decay function of *l,* the temporal distance between the target and the intruding item, or the *lag* of the intrusion. The choice of an exponential decay function   
To allow for asymmetry in terms of temporal similarity for backwards and forwards lags, scales the similarity slope in each direction such that when , intrusions presented after the target have greater temporal similarity, and hence are weighted more in calculating the overall intrusion likelihood, compared to intrusions preceding the target, commonly observed in studies of temporal contiguity (Kahana, 1996). The rate of exponential decay is estimated separately for the forwards and backwards similarity slopes.

### Model 5: Spatiotemporal Similarity Gradient

Using the same basic structure as the previous models, in Model 5 intrusion likelihood is a weighted product of temporal and spatial similarity:

|  |  |
| --- | --- |
|  | (6) |
|  | (7) |

where the overall weight given to each intruding angle, *w*, is determined by both the temporal similarity between the intruding item and the target, *t* as defined in (5), and the spatial similarity between the target and intruding angles *s*:

|  |  |
| --- | --- |
|  | (8) |

similarly to the temporal component, spatial similarity

The relative contribution of temporal and spatial similarity in determining overall spatiotemporal similarity is weighted by .

### Model 6: Spatiotemporal-Semantic-Orthographic Similarity Gradient

# Overview of Experiments

In experiment 1 we found qualitative improvements in fit with more sophisticated systematic intrusion processes, but insufficient data

# Experiment 1

## Method

### Participants

In Experiment 1, 10 participants were recruited online through the University of Melbourne undergraduate research experience program and 40 participants were recruited via *Prolific*, an online participant recruitment platform. Five participants from the undergraduate pool and seven participants from the Prolific pool did not complete all sessions of the online experiment, resulting in incomplete datasets which were excluded from the final analyses. Additionally, two participants recruited via Prolific were excluded due to at-chance performance in the memory retrieval task, measured by applying the Rayleigh test which indicated no evidence for a departure from uniformity, interpretable as completely random responding. After exclusion, there were five undergraduate participants and 31 Prolific participants, for a total sample of 36 participants. For their participation in each session, undergraduate students were granted credit towards course requirements, and Prolific participants were paid 6.50 GBP/hour. Participants were provided with plain language statements and consent forms and gave informed consent prior to the start of the first session of the experiment.

### Stimuli and Apparatus

Stimuli consisted of words generated from the SUBTLEXus database, filtered for words with a length of four letters, and with frequency ratings between one and five. Words were displayed in size 24 point “Courier New” white font positioned in the center of a uniform mean luminance field. The choice of a monospaced font and the restriction of words to strictly four letters were to ensure stimuli always occupied a consistent amount of space on the screen. Software written in Javascript using jsPsych (deLeeuw, 2015) controlled stimulus presentation and recorded responses.

### Procedure

Participants completed the experimental tasks over three sessions. Each of the three sessions consisted of 120 trials, which was broken up into 12 blocks of ten items each. Blocks consisted of a study phase, a math distractor phase, a recognition phase, and finally a source recall phase. There were additionally five practice trials at the beginning of each session, the data from which was not included for analysis. There were two conditions in this experiment, a simultaneous study condition and a sequential study condition, with all other phases being identical between the conditions. Participants were randomly allocated to either the simultaneous or the sequential presentation condition when beginning session one of the experiment, which would be the same for all subsequent sessions for that participant.

In the sequential study condition, participants were presented with a black marker positioned on a randomly generated angle on the outline of a circle at the start of each trial for 600 ms. The presentation of the marker was followed by the display of a word in the center 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 π/8 radians 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 “TOO DISTANT” was displayed for 1000 ms, then the location was then re-presented and the verification task was repeated.

In the simultaneous study condition, participants were presented with the marker and the word simultaneously for 1000 ms. Instead of being positioning the word in the center of the screen, in the simultaneous encoding condition, the word was positioned at the same angle as the marker, offset by a longer radius. The location of the word relative to the marker was determined by the sector the angle was in, with the word being offset to one of eight points on the bounds of the text box, corresponding to the middle of each of the four sides, and the four corners (i.e. in the North sector, the anchor was the bottom middle of the text box, while in the Northeast sector the anchor was the bottom left of the text box). As with the sequential condition, a verification task followed each presentation, which was repeated until participants reproduced the location to within π/8 radians of the presented angle.

After studying each of the items for that block, participants were then instructed to complete a distractor task, which involved 30 seconds of arithmetic problems. These problems were presented as three single digit integers, which summed to a fourth number which would either be the correct sum, or a number that was one higher or lower than the actual sum. Participants would indicate if the sum was correct by pressing the keys 0 (false) or 1 (true).

In the recognition phase, participants were shown a shuffled list of 10 previously studied items and 10 foils and asked to rate each item on a six-point Old/New confidence scale. Participants responded by pressing a number from 1 to 6 on their keyboard, with 1 representing “Sure New” and 6 representing “Sure Old”.

Finally, in the source memory retrieval task, participants were cued with the words for 1500 ms, and then indicated the recalled location by a moving the mouse from the starting point in the centre of the circle to a point on the circumference of the 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 1.

**Figure 1**

Schematic of display presented to the participant in one trial in each phase of the experiment.

A picture containing clock

Description automatically generated

## Results

To facilitate presentation of the analyses, we step through incremental elaborations of the response error model, and then repeat these steps with the circular diffusion model that models response error and response time data. First, we assess whether there is a difference between the sequential and simultaneous presentations of source and item, in terms of both recognition and source judgements. Second, we turn our attention to intrusions and compare a thresholded model without intrusions to a model with intrusions, and a model with intrusions and a threshold. Third, we introduce a more sophisticated intrusion component to the model that is sensitive to the similarity between items when determining individual pairwise intrusion probabilities. Finally, we repeat these steps with the circular diffusion model.

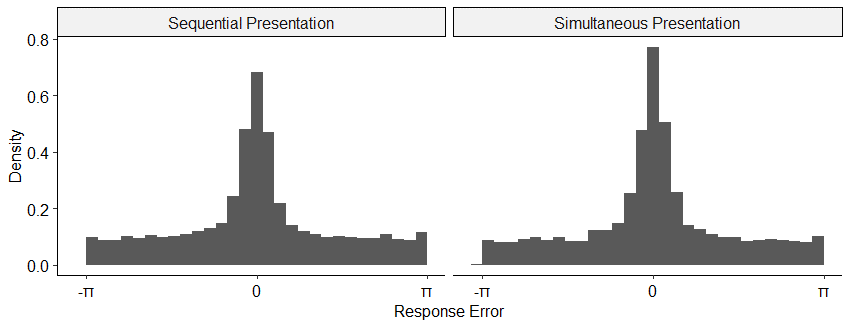
### Data Exclusion

In addition to the rejection of two participants previously described under methodology, individual responses from the remaining participants with a response time of faster than 300 ms or slower than 7000 ms were also excluded from subsequent analyses. This resulted in the omission of 1.72% of data.

### Simultaneous vs Sequential Presentation

With regard to performance in the source judgments, response error averaged within and compared between the simultaneous (M = .009, SD = 1.37) and sequential (M = .002, SD = 1.43) groups were not significantly different *t*(12460) = .28, *p* = .773. This can be confirmed visually by comparing the distributions of response error in the two conditions (Figure 2)

**Figure 2**   
Normalized Histograms of Source Error in Sequential and Simultaneous Presentation Conditions



Subsequent modelling analyses were conducted on an individual level, and significance tests on resultant parameter estimates between the presentation conditions were also not significant. For the sake of brevity, these analyses are provided as supplementary material and commentary on the modelling will not make further reference to the presentation manipulation.

### Model Comparisons

## Discussion

Successive qualitative improvement in fits when intrusion probabilities were determined by temporal and spatiotemporal gradients. However, this was not reflected in quantitative fit statistics, in which marginal improvements in model likelihood with the temporal and spatiotemporal models were outweighed by the additional parameters entailed by those models.

# Experiment 2

## Method

The method for Experiment 2 was identical to Experiment 1 with the following exceptions detailed below.

### Participants

In Experiment 2, participants were recruited solely via Prolific. Of the 10 participants recruited, four participants did not finish all sessions of the experiment, and one participant was excluded as the Rayleigh test indicated no deviance from uniform responding, leaving a final sample of five participants included for the analyses.

### Stimuli

Words were sampled anew from the entire list each time

### Procedure

Only simultaneous presentation. 10 sessions instead of three.

## Results

## Discussion

# General Discussion

## Mixture Models

Ambiguity about the relative contribution of multiple components (i.e. is the decrease in overall intrusion probability over the serial position of the target item associated with an increased probability of the memory component or the guessing component in the model?)

It does not seem reasonable to expect that the proportion of guesses remains the same across serial positions, but we do not have a formal alternative model of guessing. To take an extreme example, we can consider a potential interaction between recognition and intrusion probability where items that are not recognized do not intrude. In a list where no items are recognized, we would intuit that all responses should be guesses.

Need for a process model, like racing diffusion models, to address this ambiguity.

1. The von Mises distribution is a circular analogue of the Gaussian distribution. [↑](#footnote-ref-1)