In this article, we had two main aims, one theoretical and one empirical. Our theoretical 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 on the distributions of decision outcomes only, would continue to hold when both outcomes and RTs were taken into account. Our empirical 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 three pieces of convergent evidence 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 characterised 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 theoretical 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. Thirdly, a more flexible version of the circular diffusion model in the GVM which allowed for a gradual transition between continuous and threshold theories of retrieval yielded parameters that resulted in a GVM that was similar to the threshold account.

**Response Time Modelling**

In this paper we present a novel application of the circular diffusion model to memory data. This innovative approach

**Implications for Models of Memory**

In our study, jointly modelling RT along with response error provided evidence for a thresholded model of source memory. This work is paralleled in the change detection literature, Donkin and Nosofsky (2016) found RT evidence supporting a hybrid model of VWM consisting of a mixture of guessing and memory-driven states. Our findings fall within the growing body of work in memory research that suggests the architecture of memory involves a retrieval threshold. This threshold allows for a state where memory can fail discretely, but importantly, precision is variable for responses where retrieval was successful. In brief, memory appears to be some-or-none, not all-or-none (Onyper, Zhang, & Howard, 2010).

**Sequential Presentation of Stimuli**

In this study, we attempted to keep as methodologically close to the Harlow and Donaldson (2013) paradigm that introduced continuous report to the source memory literature. This motivated our decision to display item and source stimuli sequentially, rather than simply displaying the stimuli items in their source context, as is intuitive. The argument provided by Harlow and Donaldson (2013) was that presenting both aspects of the stimuli simultaneously would result in unitization that might allow for familiarity to contribute to what should otherwise be a pure recollection task in source memory retrieval. This operationalization of source memory, while theoretically convenient, may not be ecologically valid as real-world events, by definition, occur in their source context. It is possible that a future experiment where item and source are sequentially presented may result in different findings.

**Individual Differences**

**What did we find?**

Participants varied greatly in how they responded to the task. There appear to be three distinct groups of participants based on responses to the task. Some participants (1, 9, 13, 15) have very low memory precision (response error for 1 and 13 don’t significantly deviate from uniformity), and are not very diagnostic of model performance. On the other extreme, participant 10 is also qualitatively distinct from the rest of the participants, with a pattern of extremely precise responses with very long RTs (on the order of 3-5s- compare medians?). For this participant, all models struggle to account for the peak of the response error distribution, because they are trying to simultaneously capture the slow RTs (this participant is just very weird, outlier). Performance also varies between the remaining 14 participants, but they are at least engaging with the task at similar latencies and with non-zero precision. Harlow and Donaldson may have fit to individual data- check this!

**Relate data to intro**

Intro mentions fast error pattern. When we analysed group level data, we found a fast error pattern in the joint distribution of RTs and error (i.e. less accurate responses were associated with shorter RTs). The diffusion model is able to account for a fast error pattern by allowing trial-to-trial variability in the criterion (a) of the diffusion process, and so we incorporated this into all of the models. At an individual level however, there was no fast error pattern (refer to flat quantiles in quantile-quantile plot).

**Have other people looked at this? If so, what did they find?**

Averaging artefacts in previous source memory research- Slotnick and Dodson? Bit different though, that’s ROC curves.

**What does this mean for memory/decision-making?**

Important to model data from these tasks at an individual level.

**Imageability**

**What did we find?**

*Parameters*

Not much of a difference

**Relate data to intro**

**Have other people looked at this? If so, what did they find?**

**What does this mean for memory/decision-making?**

**Confidence**

**What did we find?**

Participants used confidence ratings mostly at the extremes (.79 of all ratings = 6, .83 = 5 & 6). Not a lot of data for items recognised with lower confidence or deemed new.

**Relate data to intro**

**Have other people looked at this? If so, what did they find?**

**What does this mean for memory/decision-making?**

Quick paragraph its possible work may be different with other source modalities, people have done voices, colour. “limitations” suggesting interesting way of doing this. An interesting future direction not a limitation.

**Subheading summarizing circular diffusion work**

**Colour naming**