# General Introduction

In order to effectively use remembered information to inform our beliefs and behaviours, we constantly need to make judgements about origin or *source* of that information. Judgements of this kind are studied in the laboratory using source memory tasks, in which subjects, when cued with a given item, report the context in which that item was encountered. In the source memory literature, a key question is whether retrieval from source memory is better characterized as a thresholded (i.e. discrete) process, in which retrieval either succeeds or fails absolutely, or a continuous process, in which retrieval always returns some information, with varying degrees of quality or precision. In the following review, I aim to connect this specific question to the broader episodic memory literature to illustrate the significance of accurately characterising the nature of source retrieval.

The body of this thesis comprises a series of challenges to the thresholded view of source memory. In Chapter 2, I ask firstly whether heavy-tailed distributions of source errors, which have been previously interpreted as guesses according to thresholded models of source memory, could instead be a result of properties of the decision-making process rather than those of memory. Secondly, given these errors are a memory phenomenon, I ask whether source guesses are instead source responses for items that are not recognised. In Chapter 3, I firstly distinguish between errors due to guesses, made in the absence of any information, from intrusion responses, which are driven by information from an incorrect item, and I secondly investigate whether the heavy-tailed error pattern is robust to changes in how source and item information is presented. Chapter 4 continues the investigation into intrusions, specifically pursuing whether item similarity affects intrusion probability. Through these empirical chapters, I ultimately find that the threshold account of source memory retrieval holds under scrutiny from each of these angles. In the final chapter, I discuss the implications of my findings and the conclusions we can subsequently draw about episodic memory.

## Episodic Memory

Memory is an essential part of the human experience. As simple organisms, memory allows us to maintain a record of past experiences and use it to build expectations of and prepare reactions to our environment. As complex human beings, memory allows us to integrate our experiences into a sense of self, and influences everything that entails, such as our beliefs, personalities, and abilities. Given the fundamental importance of memory, it is unsurprising that the study of its properties and processes is one with a long tradition. Modern cognitive science commonly distinguishes between procedural memory, that is, forms of memory that we use without conscious manipulation, such as remembering how to ride a bicycle, and declarative memory which we consciously retrieve, such as how many wheels a bicycle has, or the time you were first taught to ride a bicycle. A further distinction within declarative memory can be drawn between semantic memory, which is knowledge for factual information like the first example, and episodic memory, which is memory for specific events like the second example (Hintzman, 1990; Shimamura, 1989; Squire, 1987; Tulving, 1972). Although the complexity of memory in practice does not always fit neatly into this taxonomy, distinctions between different kinds of memory are useful in orienting ourselves in the voluminous literature. This thesis concerns the nature of episodic memory, specifically, what source memory judgements reveal about episodic memory retrieval.

### Source Memory

The study of source memory was popularised by Johnson and her colleagues, and has roots in the literature on false memory, particularly with regard to how people mistake memories of imagined events for memories of actual events (Johnson, 1988; Johnson & Raye, 1981). A typical source-monitoring experiment sees subjects study items that are associated with one of two sources, A and B, and then are tested on a mixture of A items, B items, and new items that were not studied, N (Batchelder & Riefer, 1990). This paradigm allows for measurement of the ability to discriminate between studied items and unstudied distractors, and separately, the ability to discriminate between the sources of studied items. Source-monitoring experiments have been used with a variety of different source features, such as the font of words (e.g. Hintzman et al., 1972; Light & Berger, 1976), the gender of a voice (e.g. Craik & Kirsner, 1974; Light et al., 1973), or even discriminating between perceived features and imagined ones that subjects are merely instructed to visualise (known as reality-monitoring; Johnson & Raye, 1983; Johnson et al., 1982). Typically, healthy subjects are able to discriminate between correct and incorrect source features with high accuracy (Eich & Metcalfe, 1989; Batchelder & Riefer, 1990), but when placed under time constraint, source performance was negatively impacted to a greater extent than recognition (Johnson et al., 1994). Source discrimination differs in populations, such as subjects of advanced age or diagnosed with age-related disease (Hashtroudi et al., 1989; McIntyre & Craik, 1987), diagnosed with schizophrenia (Harvey, 1985), amnesia (Hirst, 1982; Mayes et al., 1985; Shimamura & Squire, 19991), and frontal lobe lesions (Janowsky et al., 1989). Critically, these deficits were also found to result in a greater degree of impairment to source discrimination than item recognition, leading researchers to suggest a dissociation between the capacity to perform the two types of tasks (Hashtroudi et al., 1989; McIntyre & Craik, 1987 Mitchell et al., 1986). A key focus for general models of episodic memory is to jointly explain performance across tasks and formalise the relationship between the two (Banks, 2000). This comprehensive approach is particularly important if we assume that in doing either task, subjects build a mental representation, aspects of which are used to support both item and source judgements (Hautus et al., 2008). In this section, I introduce three classes of models: discrete-state, continuous, and dual-process models, followed by a review of the data from two-choice tasks these models were tested against, and finally argue that a more diagnostic paradigm in the continuous-outcome task is required to differentiate between these accounts.

### Threshold Models

The earliest approach to relating item recognition and source judgements with a mathematical model, taken by Batchelder and Riefer (1990), was to place the two in a processing-tree structure, with binary outcomes at each stage. This structure proposes that when retrieving information about the item in memory, a decision is first made about whether it was previously seen, and then if the item is deemed to have been studied, a subsequent decision is made about its source. The processing tree terminates in one of several possible outcomes and is described as multinomial because it models the probability of each of these outcomes occurring. Multinomial models are based on threshold theories of signal detection, which assume that the decision space can be divided into a number of discrete areas which describe different cognitive states subjects are in when they make a response (Krantz, 1969; Luce, 1963a; Luce, 1963b). A key goal of multinomial models of source memory is to separately measure source sensitivity and response bias, two measures that previously difficult to isolate empirically (Batchelder & Riefer, 1990; Batchelder et al., 1994). Although the intent of Batchelder and Riefer (1990) in introducing multinomial models of the source monitoring paradigm was to provide these separate measurements, the discrete branching paths of the processing tree can also be interpreted as a model of the process underlying recognition and source responding, namely that responses are generated as a result of a sequence of discrete states.

In the Batchelder and Riefer (1990) multinomial model, the first stage that determines if items, from either source, are detected (i.e., recognised as previously studied). In the item recognition judgement, the decision space is partitioned by a single threshold into two states with probability *D* and 1-*D*. This framework known as the *high-threshold* view of signal detection, which means that on lure trials where there is no memory signal (when the item is actually new), “old” responses are only generated as a result of a guess with probability *b* and lures are otherwise correctly identify the item as new. In other words, a nonsignal never exceeds the detection threshold, which lends the high-threshold theory its name. Because there is only a single threshold partitioning the space, the recognition component of the Batchelder and Riefer (1990) model can be referred to as a one high-threshold (1HT) model, which distinguishes it from the source judgement which is assumed to be a two-high threshold (2HT) process. The two thresholds in the source judgement divide the decision space into three areas, so that if one threshold is crossed the source is identified as A, if the other threshold is also crossed the source is identified as B, while if neither threshold is crossed then no information is available and source A is guessed with probability *a*, and source B is guessed with probability 1-*a*.

Diagram

Description automatically generated

Historically, 1HT models have been criticised for making predictions about the shape of the receiver-operating characteristic (ROC; discussed in detail later) that are contrary to sensory and recognition memory data, which Kinchla (1994) argued renders the Batchelder and Riefer (1990) model inadequate as a model of source-monitoring (Green & Swets, 1966; Luce, 1963; Murdock, 1974). Notably, the adoption of a 1HT model of recognition is a specific feature of the Batchelder and Riefer (1990) model, and not of all multinomial models. Instead, multinomial models with a single low threshold (LT; Batchelder et al., 1994; Macmillan & Creelman 1991) and 2HT recognition processes (Bayen et al., 1996) have also been proposed, although the source judgement is consistently modelled as a 2HT process and so these multinomial models made the same predictions about the shape of source ROCs.

### Continuous Signal-Detection Models

Continuous models of source memory are based on Signal Detection Theory. Perhaps the single most influential advancement in the study of episodic memory was the application of Signal Detection Theory (SDT) to understanding recognition memory by Egan (1958). SDT was first developed in the context of the Second World War to describe how radar operators made decisions about the presence or absence of targets in radar signals, based on the level of the underlying strength of the signal (Marcum, 1947). As a general framework for understanding how people make decisions about noisy stimuli, SDT since proved useful for distinguishing between the effects of sensitivity and response bias in sensory perception (Creelman, 1965; Green & Swets, 1966). In the application of the SDT framework to recognition memory by Egan (1958), the underlying signal is the degree of memory strength elicited by a stimulus. This signal is assumed to be continuously distributed, and we can refer to the degree of memory strength as familiarity, such that studied items are on average higher along the continuum of familiarity than unstudied items. As with radar operators, subjects in recognition tasks are thought to decide whether the degree of familiarity elicited by the stimulus is the presence of a target embedded in noise (studied), or simply an instance of noise (unstudied; Osth & Dennis, 2015). This decision is made by comparing the signal to a criterion, specifically a level of familiarity above which items are judged to be recognised.

Item recognition has been fully incorporated in signal detection theory (SDT) models (Egan, 1958; Gillund & Shiffrin, 1984; Glanzer, Adams, Iverson, & Kim, 1993; Green & Swets, 1974; Murdock, 1965; Murdock, 1974; Ratcliff, McKoon, & Tindall, 1994; Ratcliff, Sheu, & Gronlund, 1992; Shiffrin & Steyvers, 1997) either by SDT in isolation (e.g., Egan, 1958; Murdock,

1965) or by SDT incorporated in other models (Gillund & Shiffrin, 1984; Glanzer et al., 1993; Murdock, 1982; Shiffrin & Steyvers, 1997).

Continuous models of source memory extend SDT such that memory strength is assumed to vary continuously on two dimensions, bivariate SDT.

(Banks, 2000; Glanzer, Hilford, & Kim, 2004; Mickes et al., 2009).

In its application to the study of memory, SDT proposes that recognition judgements are based on its familiarity- which by analogy is a signal which varies in strength.

### Dual-Process Models

In a dual-process view (Yonelinas, 1999), one can respond by directly retrieve an item from memory through recollection, or by simply making a judgement about whether the item is memory or not without retrieving it, based on a feeling of familiarity. In this way, both recollection and familiarity can contribute to successful recognition. On the other hand, in a source memory task, familiarity cannot distinguish between two studied items from different sources, as both items are present in memory and should therefore be equally familiar. Thus, source judgements are thought to reflect a pure recollection process. When performing a recognition task, one can respond by directly retrieve an item from memory through recollection, or by simply making a judgement about whether the item is memory or not without retrieving it, based on a feeling of familiarity. In this way, both recollection and familiarity can contribute to successful recognition. On the other hand, in a source memory task, familiarity cannot distinguish between two studied items from different sources, as both items are present in memory and should therefore be equally familiar. Because source judgements rely only on recollection, the Yonelinas (1999) dual-process model predicts that source judgements should rely purely on a high threshold recollection process. Thus, dual-process and discrete-state models make identical predictions about source memory but are distinguishable on item recognition.

### Evidence from Receiver Operating Characteristic Curves

ROC curves are constructed by plotting hit rates against false alarms at multiple criterion points (Fawcett, 2006). These criterion points are typically obtained by recording subjects’ self-ratings of confidence in the accuracy of their response (Wixted, 2007). ROCs were first used with Signal Detection Theory (SDT) to study how performance varies with response bias, or as a decision criterion is varied (Green & Swets, 1966; Norman & Wickelgren, 1969). With specific reference to models of memory, researchers have used the shape of this curve to infer properties of underlying distributions of memory strengths (Macmillan & Creelman, 1991; Slotnick & Dodson, 2005). In particular, thresholded and continuous models of memory were thought to predict different signature ROC shapes, which lead to a great deal of interest in comparing ROCs across different memory tasks. We will first consider early applications to recognition memory, which will explain the emerging interest in source judgements which followed in the literature.

In recognition memory tasks, ROC curves tend to be curvilinear (Figure 2B). Curvilinear ROCs are well explained by SDT and this was one of the strongest pieces of evidence for early SDT explanations of memory (Yonelinas, 1994). Consider a standard two-choice recognition task where subjects judge whether and item is old or new. SDT says that old and new items are each associated with a normal distribution of memory strengths. In Figure 2B, the first criterion point represents the strictest possible response criterion (say, rated 1 out of 6), so that only the items remembered with the highest confidence are considered hits (in the case of items that were actually old) and false alarms (in the case of new items). The area under the “old” curve to the right of the [red] line (i.e. the proportion of old items with strength exceeding the response criterion associated with the highest confidence rating) are hits, while the area under the “new” curve above the same line are false alarms. This hit rate is plotted against this false alarm as the [red dot]. Each subsequent point in the ROC incrementally relaxes the response criterion, so for the second point items rated with either the highest or next highest level of confidence (1 or 2 out of 6) are considered, and so on for each point on the confidence scale (Yonelinas, 1994).

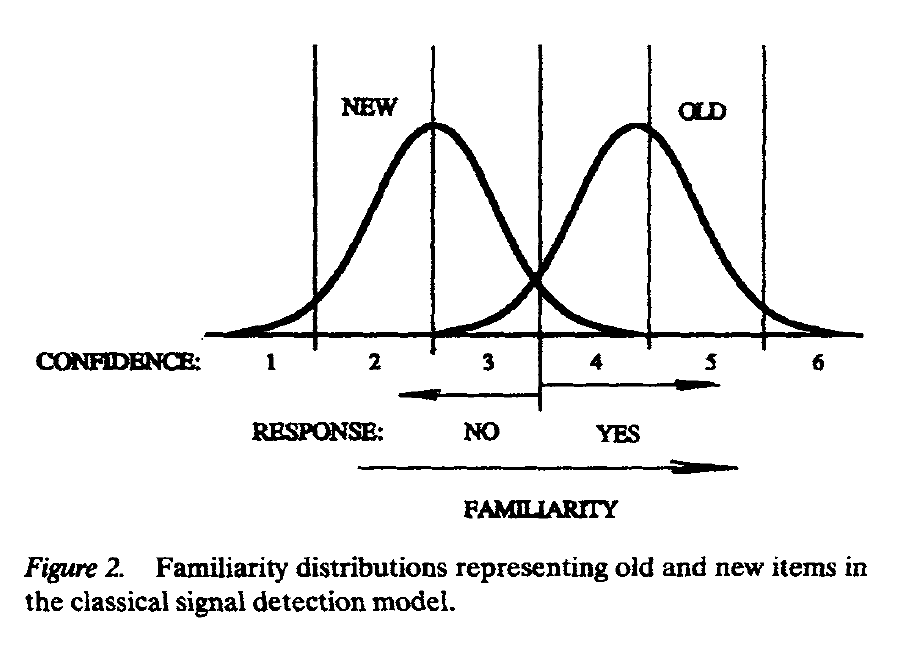


Figure 2 Yonelinas, 1994

In addition to comparing the probabilities of hit rates and false alarms, these probabilities can also be transformed into z-scores, which when plotted against each other, are referred to as a z-transformed ROC (z-ROC).ROC curves can be transformed to z-space to evaluate the degree of symmetry between the old and new distributions. If both are identical distributions, varying only in their mean, then the z-ROC should have a gradient of 1.0, a straight diagonal line. However, recognition zROC slope sometimes less than 1. Yonelinas (1994) explains this as an increase in variance of the old distribution, or more specifically as a skew to the right. However, unequal variance signal detection theory model is also able to explain this, and has been popular in other paradigms (find some examples). So it is difficult to distinguish between the dual-process signal detection model and the unequal variance signal detection model on the basis of recognition ROCs

Hirstman & Hostetter (2000): change in presentation time can affect z-ROC slope, contrary to Yonelinas (1994), which was a dual-process evidence.

A key piece of evidence used by Yonelinas (1994; 1997; 1999) to support a dual-process model was the observation that ROCs look different in item recognition compared to associative recognition and source judgements, implying the distributions in the response space are different for these tasks. 1997, Straight ROC for associative memory. However, Kelley and Wixted

and then in 1999 linear ROC for source memory (which we already established is a special case of assoc. memory). This is new, stronger evidence for the Yonelinas dual-process model

Now, consider a source memory task where subjects must attribute previously studied items to one of two sources. Yonelinas (1999) found that source ROCs are linear

Yonelinas 1994- familiarity is the same thing as unequal variance SDT

Yonelinas 1997- Straight ROC for associative memory

‘99- linear ROC for source

Last two are the big ones

But

The dual process framework, which has been influential on how we understand aging (blah), and blah and blah, rests on the premise that recollection is a thresholded process. Scrutiny is placed on source memory tasks which, as a pure test of recollection, should exhibit strictly thresholded retrieval.

Diagram, engineering drawing

Description automatically generated

Slotnick & Dodson for source:

However, there are limitations to ROCs that undermine their supposed diagnostic value in distinguishing the two models. The dual process model made the prediction that item recognition ROCs should be curvilinear and source memory ROCs should be linear, which was confirmed by Yonelinas (1999). However, Slotnick and Dodson (2005) reanalysed the same data, which included item recognition confidence ratings, collected before source memory confidence, allowing for conditionalization of source memory performance on item recognition. This reanalysis demonstrated that if source ROCs were plotted separately for different levels of target confidence, the highest confidence source ROCs were in fact curvilinear, contrary to the predictions of the dual process model. The authors argued that only the items with high familiarity confidence ratings contained diagnostic source information, and that the linearity of source ROCs was an artefact of collapsing across all items even if no source information was encoded, and was thus not evidence of a recollection threshold (Figure 2; Slotnick and Dodson, 2005).

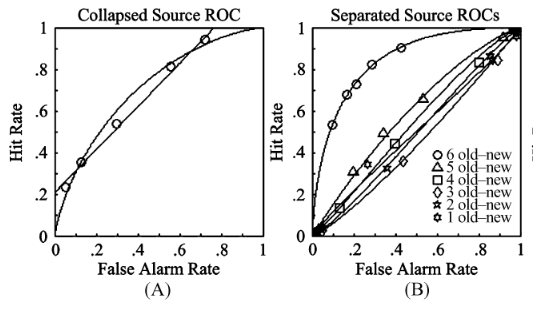


Figure 3. Source ROCs constructed from Yonelinas (1999, Experiment 2), comparing the relatively linear ROC when averaged across item familiarity confidence ratings (A), and when separated (B). Note that the ROC of the highest separated band (6) appears curvilinear, while the average is linear. Adapted from “Support for a continuous (single-process) model of recognition memory and source memory,” by S. D. Slotnick and C. S. Dodson, 2005, Memory & Cognition, 33(1), 151-170.

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. On such condition was when an item and a source were treated holistically as one item, known as *unitised familiarity*. Other conditions sufficient for more curvilinear source ROCs included increasing study event complexity, and better overall performance on the task (Yonelinas & Parks, 2007). With several possible explanations for the curvilinear shape of source ROCs, no single interpretation was diagnostic of the underlying process producing the observed pattern of responses.

Ultimately, two-choice tasks and ROC analysis are too ambiguous to sufficiently distinguish between continuous and threshold accounts of source memory and that a more diagnostic paradigm was needed to determine if source memory performance is really associated with a threshold. Harlow and Donaldson (2013).

### Continuous-Outcome Tasks

In continuous-outcome tasks, subjects make responses on a continuous, often circular, domain instead of selecting between discrete choices. These tasks have their methodological origins in the study of sensory thresholds in classical psychophysics, specifically in the method of adjustment, in which subjects adjust the magnitude of some property of some sensory stimulus (e.g. the loudness or pitch of a tone) along a continuum until they judge it to be equivalent to a standard (Smith, 2016; Woodworth & Schlosberg, 1954). Another way to measure sensitivity is to present a set of stimuli (e.g. multiple tones of different pitches) and have subjects make a binary comparison to the standard (“this tone is higher than the standard tone”), a technique related to 2AFC tasks which grew to ubiquity in cognitive psychology (Smith, 2016; Woodworth & Schlosberg, 1954). Compared to binary judgements, the response continuum of the method of adjustment has notable advantages and disadvantages.

Continuous-outcome tasks allow direct observation of the variability of responses and in turn, allows comparison of models

Source responses made on a continuous domain allow for an objective, direct measurement of the precision of the retrieved information, as opposed to relying on different ratings of confidence in a two-alternative forced choice as a proxy that is compromised by the variability in what each level of confidence means across participants, and even across trials for one participant. The problem of subjectivity is just like the reason why we use different tasks instead of the single remember-know. The same sorts of issues recur and lead to cycles of development. Better measurements from new experiments, extension of models to explain these measurements.

Continuous-outcome tasks were first applied to memory research in the study of the capacity limits of visual working memory (VWM; Wilken & Ma, 2004). [Strictly speaking, not developed there. Blake et al. (1997) and Prinzmetal al. (1998)

Continuous tasks were applied to source memory by Harlow and Donaldson (2013). Popov et al. also.

Another disadvantage to continuous outcome tasks is that while many models of decision-making and response times were developed for two-choice tasks, such models for continuous outcomes were only recently developed (Ratcliff, Smith, Kvam)

## Decision-Making

Response not a direct readout of memory. Properties of the decision-making process can affect responding. Ratcliff & Starns RTCON

### Modelling Decision-Making in Memory Tasks

Diffusion models have emerged as increasingly influential accounts of both RT and accuracy data in decision tasks, and naturally explains well-documented phenomena like speed-accuracy trade-off effects (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 in visual working memory are made using a diffusion process (Smith & Ratcliff, 2009). The diffusion decision model conceptualises decision making as occurring by a process of noisy evidence accumulation. In a two-choice task, evidence is accumulated between a pair of decision boundaries that represent the decision criteria for each choice. (Ratcliff, 1978). In order for a decision to be made, the evidence accumulator must reach one of these boundaries, with whichever criteria being met first determining the response (Ratcliff et al., 2016). Although alternative models with two competing accumulators exist, the standard diffusion model assumes a single evidence accumulator, which at each moment in time may move towards either of the two boundaries due to noise in the evidence accumulation process but tends to move toward the correct boundary more than the error boundary depending on the quality of evidence entering the decision process (Figure 5; Ratcliff et al., 2016).

### The Circular Diffusion Model

Slow errors

### Preview of Chapter 2

## Intrusions from Non-target Items

### Models of Non-target Responding

Bays swap errors

Interference Model Oberauer & Lin

Temporal Contiguity Healey Kahana

Rerko

### Similarity Effects in Memory

### Preview of Chapter 3

### Preview of Chapter 4

## Summary

**Figure 1**

*The Structure of the Thesis*

