# General Introduction

When drawing upon our memory, we may sometimes summon some information without being able to say from where or whom we learned it. This characteristic of human nature can result in errors such as retelling a joke to the person you heard it from in the first place. 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 presents 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 the 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 is not always neatly described by this taxonomy[[1]](#footnote-1), 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.

### Single and Dual-Process Theories of Remembering and Knowing

[I wanted to start by casting a wide net in my review, and the remember-know stuff seemed like an early point to start from. Reading back on this, I wonder if this is all a bit tangential, and whether I should pare this back or even cut it entirely and get straight to recognition/source? I do think I need to give more context than I did for my papers. Can’t quite bring myself to do it yet, but the intro might read better with some of this fat trimmed]

When we remember episodes, we can experience the retrieval of information in different ways. These different experiences can be observed experimentally in the remember-know (RK) paradigm introduced by Tulving (1985). In RK tasks, subjects must judge whether items were previously seen (i.e. recognised) and critically, report the basis for their judgements: either they “remembered” the prior occurrence of the item (R) or that they simply “knew” that they had seen it without any accompanying information about the prior occurrence (K). As a concrete example of these experiences, imagine seeing a man on the bus: you feel that you have met before, but you cannot quite remember who he is. Concentrating, you consider and eliminate some possibilities: he is not a co-worker, a relative, or a celebrity. Eventually, you remember seeing him behind a glass counter with cuts of meat-- he’s the butcher from the supermarket! This, Mandler’s (1980) classic “butcher on the bus” example, illustrates the difference between remembering and knowing, and the attached explanation makes some assumptions about how we arrive at these different states of awareness about the past, principally that there are two memory processes at work. First, there is an initial “context-free” feeling of familiarity that enabled recognition of the man without conscious control, followed by a conscious, effortful search process which enabled his identification. Explanations such as this one, which involve two qualitatively different memory components, are known as *dual-process* models of episodic memory retrieval. Generally, these processes are referred to as familiarity, which corresponds to the first process in the prior example, and recollection, which corresponds to the latter, although the precise distinctions between the two processes vary between specific models which fall under the dual-process framework (Atkinson & Juola, 1974; Gardiner, 2001; Jacoby et al., 1997; Rajaram, 1996; Raaijmakers & Shiffrin, 1992; Yonelinas, 1994). In the influential Yonelinas (1994) dual-process model, familiarity is characterised as a continuous measure of the likelihood the stimulus was previously encountered, while recollection functions in a discrete fashion and is subject to a threshold, such that it returns the full episode of the previous encounter on a proportion of attempts but fails absolutely otherwise. We return to this particular dual-process model later in relation to source memory. The dual-process framework naturally explains circumstances where the man is not identified as the butcher: familiarity has enabled recognition, but recollection has failed (Jacoby et al., 1997). Although intuition may tempt a direct mapping of familiarity and recollection, as processes, to knowing and remembering, as outcomes respectively, there are two complications that should be noted. The first reason is that RK judgements are not process-pure, meaning that recollection as well as familiarity can contribute to successful recognition, which is a general criticism of attempts to dissociate consciously controlled and unconscious influences on cognitive tasks (Wainwright & Reingold; 1996; Wixted et al., 2010). The second reason is that two different memory outcomes do not necessarily imply the existence of two retrieval mechanisms, and *single-process* models which take this view offer competing explanations of these distinctions.

In single-process accounts of the RK paradigm, “remember” and “know” judgements reflect different levels of confidence imposed on the product of a single retrieval mechanism, such that remembering requires a more stringent decision criteria (i.e. a higher level of confidence) to be met than knowing (Donaldson, 1996; Dunn, 2004; Hirshman, 1998; Inoue & Bellezza, 1998; Ratcliff et al., 1995). This approach draws on the terminology of the signal detection theory (SDT) so is referred to as the SDT interpretation (Dunn, 2004). SDT was 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 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. Like familiarity in the dual-process model, 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. The SDT interpretation of RK judgements is that the difference between responses is where this criterion is placed: stimuli which exceed an initial criterion elicit a “know” judgement, while stimuli that also exceed a higher criterion are further deemed to be “remembered” (Figure 2). [not sure how much detail is best, there is remember/know/guess judgements, and the actual dissociations seen in this paradigm. I intend to use RK to set up SDT vs dual process models, and I want to get into source ROC studies without too much fuss. Perhaps this is unnecessary background? But better too much than too little with early draft.]

Diagram

Description automatically generated

The difficulty with distinguishing between dual-process and SDT explanations of the RK paradigm lies with limitations of the task. To distinguish between recollective and nonrecollective memory, we cannot rely on introspection/subjective judgement as with the RK procedure (Wixted et al., 2010). One way of addressing this is to use separate tasks, like recognition and source memory tasks, which differ in the demands placed upon memory. [Need to expand if keeping in]

### Discrete and Continuous Models of Source and Item Memory

Instead of measuring presumed differences in recollective and nonrecollective memory with a single RK task, an alternative approach is to probe recollection with a separate task, which spurred interest in source memory tasks. The key difference between item recognition and source memory tasks is that in the former, subjects discriminate between studied and unstudied items, while in the latter they must discriminate between items from different sources, or between possible sources for each items. 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. [out of order? Kind of jarring to go back to the history of source tasks.]

The study of source memory originated with the source-monitoring paradigm, which tests subjects’ ability to remember the incidental features of items, such as the font of words (e.g. Hintzman et al., 1972; Light & Berger, 1976) of the gender of a voice (e.g. Craik & Kirsner, 1974; Light et al., 1973), or even between perceived and imagined features 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, reminiscent of dissociations in RK judgements (Hashtroudi et al., 1989; McIntyre & Craik, 1987 Mitchell et al., 1986). The challenge for the field was to develop models that jointly explain performance in recognition and source memory tasks and formalise the relationship between the two. I will review three main classes of models: discrete-state, bivariate SDT, and dual-process. [expand]

One approach to relating item and source memory is to model item and source responses using a multinomal processing-tree structure. discrete probability of a response falling in each category (Batchelder & Riefer, 1990; Klauer & Kellen, 2010). Pure multinomial trees grow very complex. Source guessing for unrecognised items.

In continuous models of source memory, which are based on Signal Detection Theory, memory strength is assumed to vary continuously, bivariate signal detection theory, predict that performance in a source memory task declines gradually as memory strength decreases [probably save this bit for the continuous report stuff] (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.

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. Unlike subjective distinction between remembering and knowing, where the two types of judgements are not process-pure, source memory tasks allow for a direct evaluation of dual-process theory.

Traditionally, evidence characterising the underlying memory distributions have been derived from the shape of Receiver Operating Characteristic (ROC) curves. In the next section, I give a brief review of evidence from ROCs, followed by the limitations of two-choice tasks which have lead the field to move towards continuous-outcome tasks like those used throughout this thesis.

### 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).

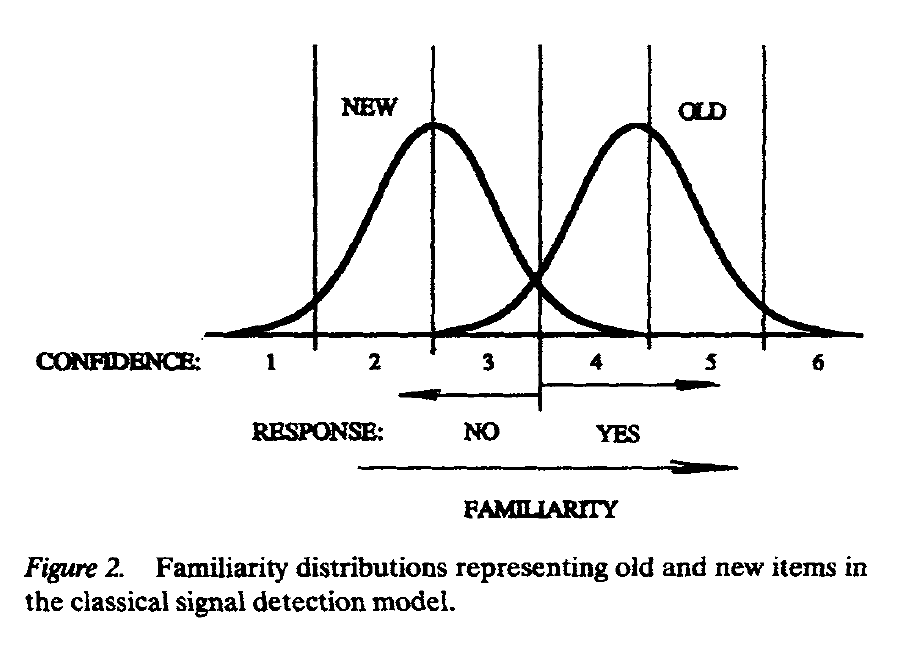


Figure 2 Yonelinas, 1994

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).

You can transform ROC 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, BUT 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).

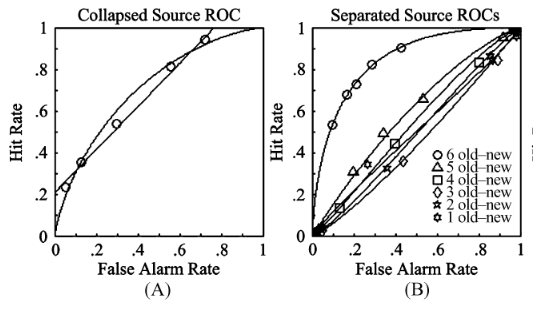


Figure 3. Source ROCs constructed from Yonelinas (1999, Experiment 2), comparing the relatively linear ROC when averaged across item familiarity confidence ratings (A), and when separated (B). Note that the ROC of the highest separated band (6) appears curvilinear, while the average is linear. Adapted 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.

The summary of these findings is that despite the reliance on ROCs in divergent models of memory retrieval, ROC data 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

Continuous-outcome tasks, in which responses are made on a continuous scale, have been widely used in the visual working memory literature (Wilken & Ma, 2004). [Strictly speaking, not developed there. Blake et al. (1997) and Prinzmetal al. (1998) predated them.]

Continuous-outcome tasks offer a range of advantages over traditional two-choice tasks, but one drawback to using them is that until recently, there were no models of RT and decision-making, as existed for two-choice tasks. In the next section, I outline why we need to model decision-making and what RT buys us, and the development of decision-making in continuous-outcome tasks such as the Smith (2016) circular diffusion model.

## 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

### 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

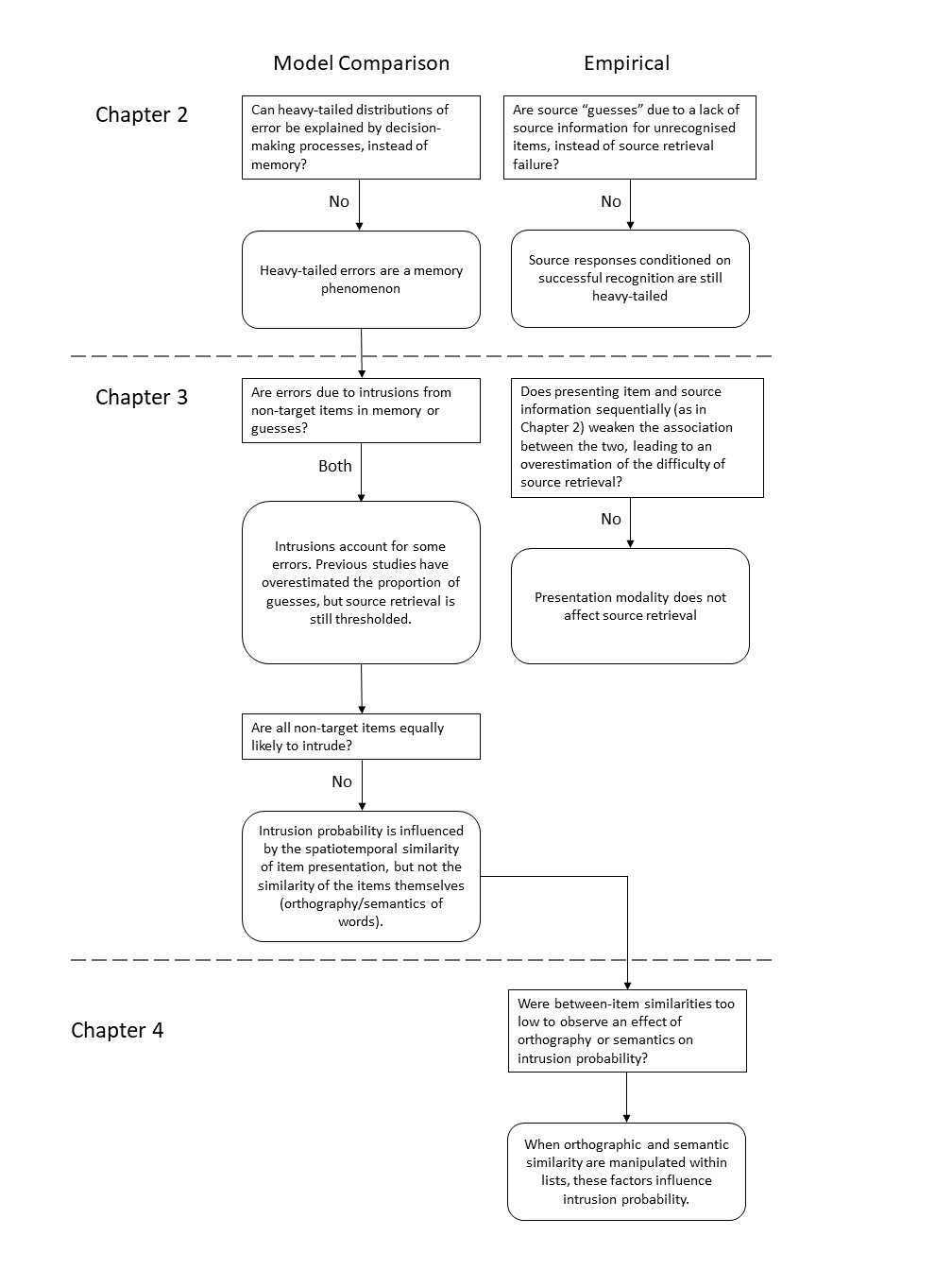
### Preview of Chapter 3

### Preview of Chapter 4

## Summary

**Figure 1**

*The Structure of the Thesis*



1. Think of example where the boundary between types of memory is not clear. [↑](#footnote-ref-1)