

A DYNAMIC APPROACH TO RECOGNITION MEMORY

Gregory E. Cox

Submitted to the faculty of the University Graduate School

in partial fulfillment of the requirements

for the degree

Doctor of Philosophy

in the Department of Psychological and Brain Sciences and Cognitive Science Program,

Indiana University

December 2015

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the
requirements for the degree of Doctor of Philosophy.

Doctoral Committee

Richard M. Shiffrin, PhD

Michael N. Jones, PhD

Jerome Busemeyer, PhD

Robert M. Nosofsky, PhD

3 December 2015

Acknowledgements

First, I extend my thanks to my committee members for their time, consideration, and guidance. I would especially like to thank my advisors, Rich Shiffrin and Mike Jones, for first convincing me to come to Indiana, making me feel welcome, and guiding me into the world I now inhabit. I must also thank Isaiah Harbison, Michael Dougherty, and Sharona Atkins for introducing me to the very idea of mathematical psychology and for teaching me and trusting me.

Enumerating the reasons why the following wonderful people mean so much to me and my scientific and personal development would be a dissertation unto itself. Beyond that, trying to dissect the ways in which they have positively influenced my life would seem unfair at best. Therefore, I list them in alphabetical order, trusting that each of them knows how important they have been and continue to be to me: Eran Agmon, Eamon Anderson, Angela Aubuchon, Sheila Blanchard, Erin Biebuyck, Maxim Bushmakin, Lisa Cantrell, Rui Cao, Emma Cohen, Mariah Coley, Patricia Crouch, Chris Donkin, Melody Dye, Brandi Emerick, Seth Frey, Rob Goldstone, Alisa Grubb, Michael Hansen, Chris Harshaw, Drew Hendrickson, Jared Hotaling, Joe Houpt, Matthew Hurley, Brendan Johns, Tessa Johnson, George Kachergis, Brent Kievit-Kylar, Artemy Kolchinsky, John Kruschke, Nick Lewis, Dan Little, Cotie Long, Jessica Montag, Zach Moore, Zander Nevitt, Brian Page, Francisco Parada, Olga Rass, Gabe Recchia, Alejandra Rossi, Tim Rubin, Geoff Stewart, Umay

Suanda, Jim Townsend, Jennifer Trueblood, Richard Veale, Charlotte Walker, Evan Weisman, Jon Willits, Dan Yurovsky.

Finally, I thank my parents, Christine and David Cox, for their continued love and support.

Gregory E. Cox

A DYNAMIC APPROACH TO RECOGNITION MEMORY

We argue that taking a dynamic approach to the understanding of memory will lead to advances that are not possible via other routes. To that end, we present a model of recognition memory that specifies how memory retrieval and recognition decisions jointly evolve over time and show that it is able to jointly predict accuracy, response time, and speed-accuracy trade-off functions. The model affords insights into the effects of study time, list length, and instructions. The model leads to a novel qualitative and quantitative test of the source of word frequency effects in recognition, showing that the relatively high distinctiveness of the features of low frequency words provide the best account. We also show how the dynamic model can be extended to account for paradigms like associative recognition and list discrimination, leading to another novel test of the presence of recall-like processes. Associative recognition, list discrimination, recognition of similar foils, and source exclusion are all better explained by the formation of a compound cue rather than recall, although source memory is found to be better modeled by a recall process.

Richard M. Shiffrin, PhD

Michael N. Jones, PhD

Jerome Busemeyer, PhD

Robert M. Nosofsky, PhD

Contents

1	Introduction	1
1.1	Models of memory	2
1.1.1	Separate and composite representations	2
1.1.2	Item and context noise	5
1.2	Theories of response dynamics	7
1.3	Overview	11
2	A Dynamic Model for Recognition Memory	14
2.1	Structure of memory traces	17
2.2	Feature sampling	22
2.3	Familiarity	23
2.3.1	Comparison between probe and trace	23
2.3.2	Activation and familiarity	25
2.3.3	Familiarity over time	26
2.4	Predicting response-signal data	28
2.5	Predicting free-response data	31
2.5.1	Collapsing boundary considerations	32
3	Validations and Insights	36

3.1	Fitting procedures	36
3.1.1	Quantile maximum likelihood	37
3.1.2	Parameter search	38
3.2	List length and study time: Nobel & Shiffrin, 2001	38
3.2.1	The data	40
3.2.2	Model fit	41
3.2.3	Discussion	46
3.3	Word Frequency	51
3.3.1	Signal to respond	57
3.3.2	Rae et al. 2014	59
3.3.3	Starns et al. 2012	67
3.3.4	Discussion	72
4	Secondary Processes in Recognition	75
4.1	Highly similar foils	78
4.1.1	Other kinds of similarity	84
4.2	Associative recognition	89
4.2.1	Modeling associative recognition	95
4.2.2	Model fits	100
4.2.3	Preexperimental associations	107
4.2.4	Discussion	110
4.3	List discrimination	111
4.4	Source memory	120
4.4.1	A model for recall dynamics	121

4.4.2	Source memory vs. source exclusion	128
4.5	Discussion	133
5	Discussion	136
5.1	Other aspects of memory	137
5.1.1	Short-term recognition and lag	138
5.1.2	Encoding	139
5.1.3	Extended events	140
5.1.4	Categorization	141
5.2	Prospects for neuroscience	142
5.3	Model extensions	143
5.3.1	Modeling confidence judgments	143
5.3.2	Alternative decision mechanisms	146
5.3.3	Alternative memory representations	147
5.4	Conclusion	151
A	Continuous approximation to the model	152
A.1	Likelihood as a log-normal	152
A.2	Familiarity as an average of log-normals	154
A.2.1	Log-normal sum approximation	155
A.2.2	Computing the average	156
A.2.3	Familiarity over time	157
A.3	Decision probabilities and latency	158
	References	161

Curriculum vitae

Chapter 1

Introduction

Recognition is a critical function performed by the memory system, and must therefore form part of the bedrock upon which any theory of memory can be constructed. When engaging in a recognition task, a participant must decide whether a given test item was or was not present in a particular study context. Recognition judgments thus indicate the ability of a participant to discriminate between experienced and unexperienced events, where an “event” is defined by the conjunction of an item and context. By characterizing the situations in which studied items can be correctly detected and unstudied items can be correctly rejected, we gain understanding of how memory is organized. We also learn about how the test situation is used to probe memory to collect evidence for recognition decisions.

Evidence from various recognition paradigms has supported the development of a variety of theories of memory (e.g., Atkinson & Juola, 1974; Metcalfe & Shimamura, 1994; Murdock, 1982; Gillund & Shiffrin, 1984; Hintzman, 1988; Shiffrin & Steyvers, 1997; McClelland & Chappell, 1998; Dennis & Humphreys, 2001). These theories are concerned primarily with recognition accuracy and confidence ratings and can be viewed as extensions of a signal detection approach to recognition (Green & Swets, 1966). According to this approach, experiencing an item in a particular context leads to a value of “familiarity” or memory strength. This value is a random sample from some distribution of familiarity values, and

accurate recognition consists in being able to tell whether any given sample came from the distribution associated with “old” items versus the distribution associated with “new” items. Some memory theories, such as dual process theory (e.g., Reder et al., 2000; Yonelinas, 1994), augment this familiarity-based process with a recall process such that, if recall is successful, recognition performance is perfect. The memory theories cited above specify how the distributions of “old” and “new” memory strength vary with experimental conditions and, in some cases, specify the likelihood that recall contributes to the recognition judgment. Many of these memory theories have been reviewed from a variety of perspectives (e.g., Clark & Gronlund, 1996; Diana, Reder, Arndt, & Park, 2006; Malmberg, 2008; Kahana, 2012), but it behooves us to review some of the principles behind many of these models.

1.1 Models of memory

Although many memory theories pertain to tasks other than recognition (e.g., recall, judgments of learning, judgments of recency, judgments of frequency, etc.), we will focus on aspects of memory models that enable them to perform a recognition task; many of the mechanisms and representational structures assumed by these models are shared across tasks in any case. As mentioned above, these theories, when applied to recognition, aim to predict the probability that a particular item will be judged as having been seen in a particular context, thus turning recognition into a form of signal detection in noise.

1.1.1 Separate and composite representations

One key distinction among most theories of memory is the representational structure of memory itself. Separate-trace or exemplar theories (e.g., Gillund & Shiffrin, 1984; Hintz-

man, 1988; Shiffrin & Steyvers, 1997; McClelland & Chappell, 1998) postulate that a distinct memory element is created each time an event is experienced. These memory elements—we will adopt the term “memory trace”—contain information about both the content of the event (e.g., the identity of a word on a study list) and the context in which the event occurred (where “context” can be defined at various levels, e.g., the time and location of the event, encoding task, and/or other nearby items or events; see Klein, Shiffrin, & Criss, 2007). To determine whether a given test item was or was not seen in a particular study context, a participant forms a “memory probe” consisting of some combination of the test item and the study context. This probe is compared to each of the memory traces. This comparison can take many forms, including a dot product (Hintzman, 1988) or likelihood ratio (Shiffrin & Steyvers, 1997; McClelland & Chappell, 1998); in any case, the comparison results in a single value reflecting the degree to which each memory trace is similar to the memory probe. This comparison is often assumed to take place in parallel across all memory traces (Gillund & Shiffrin, 1984; Hintzman, 1988; Shiffrin & Steyvers, 1997; McClelland & Chappell, 1998), but may take place over the course of a serial search among the various memory traces (Sternberg, 1966) or via a combination of parallel matching and serial search (Atkinson & Juola, 1974). Finally, these various match values are treated as evidence for whether the test item was or was not experienced in the target study context: In parallel matching models, one takes the mean (Hintzman, 1988; Shiffrin & Steyvers, 1997), maximum (McClelland & Chappell, 1998), or sum (Nosofsky, Little, Donkin, & Fific, 2011) of these match values and judge the item “old” if this is above a criterion value, and “new” otherwise. In serial search models, one continues making comparisons until either a match is found that is sufficiently high at which point an “old” response is given or until all comparisons have been made without finding a sufficiently high match,

at which point one responds “new” (Sternberg, 1966). Finally, it is possible to combine both approaches, where an initial parallel match results in an “old” response with a sufficiently high value, a “new” response with a sufficiently low value, or a serial search after an intermediate value (Atkinson & Juola, 1974).

Single-store or composite-trace models of memory (e.g., Murdock, 1982; Metcalfe Eich, 1982) assume that each experienced event is added to a single memory representation that is analogous to a “prototype” or “average” over a set of events. These models are sometimes termed “holographic” (Plate, 2003), in the sense that a single storage medium (usually represented as a vector) is used to store many separate elements. And, just as in light holography, although it is impossible to perfectly retrieve a single event from a composite trace, it is possible to do so approximately via a convolution operation (essentially “viewing” the memory from a different “angle”; see Murdock, 1982; Plate, 2003, for details). Modeling recognition is straightforward with a composite trace, since all representations are assumed to be of the same dimensionality. The test item/context is, again, represented as a memory probe, and the inner product between the memory probe and the single composite memory trace results in a scalar match value. If this value is above a criterion, the test item is judged “old”, otherwise it is considered “new”.

Composite trace models of memory are, generally, more limited than separate-trace models because the greater flexibility afforded by storage of separate traces allows separate-trace models to mimic the behavior of composite models (Nosofsky, 1988; Kwantes, 2005). Composite trace models also have trouble predicting the list strength effect, in which strengthening of some items at study via increased study time, repetition, or deeper encoding does not harm memory for the non-strengthened items; a composite trace model would tend to incorrectly predict that strengthening would “wash out” the influence of

weaker items (Ratcliff, Clark, & Shiffrin, 1990; Shiffrin, Ratcliff, & Clark, 1990). Conversely, given a sufficiently large representation, composite trace models can approximate the behavior of separate trace models; essentially, with a large enough composite vector, representations of individual memories become sparse enough to easily tell apart (Plate, 2003). There is also no reason to believe that composite- and separate-trace representations are mutually exclusive: Simultaneous storage of a sparse, distinct event trace along with addition to a growing composite trace underlies a variety of theories of the distinction between episodic and semantic memory (Nelson & Shiffrin, 2013), and is supported by differences in representation and function found between medial temporal lobes and neocortex (McClelland, McNaughton, & O'Reilly, 1995; Norman & O'Reilly, 2003; Kumaran & McClelland, 2012).

1.1.2 Item and context noise

A second critical distinction among models of memory is the chief source of noise in the recognition process. Even a completely novel item will generally produce some degree of familiarity, and rarely will two presentations of the same item lead to exactly the same performance. Most of the separate-trace models reviewed above assume that the chief source of noise is other items experienced within the target study context. That is, context cues alone are generally sufficient to “narrow down” the set of memory traces that are relevant for the decision, such that the memory probe is compared only with traces stored in the target context (Gillund & Shiffrin, 1984; Hintzman, 1988; Shiffrin & Steyvers, 1997; McClelland & Chappell, 1998) and noise results from the fact that different item can still partially match (a novel item might partially match one or more study items, while a studied item might also match some of the other studied item traces). The same is true for the

TODAM (Murdock, 1982) and CHARM (Metcalf & Eich, 1982) models, which employ a composite-trace representation. In the case of composite traces, the question is what events are added into the composite trace; in these models, only those events from the target study context are directly added (although background noise added to the composite trace might be considered to represent events from other irrelevant contexts). By assuming item noise resulting from comparison between the test item and the traces formed during the target study context, such models offer a natural account for both length effects (more studied items in the same contexts leads to more chances for a partial match by chance) and similarity effects (a foil that is similar to a studied item is more likely to be called “old” than a random foil). A pure item noise model, however, does not offer an account for how experience from pre-experimental contexts can influence recognition performance.

At the other extreme are models that assume that items can be perfectly identified, but that confusion between different contexts is the main source of noise in recognition. For instance, BCDMEM (Dennis & Humphreys, 2001) assumes that every item is associated with a noisy distributed context representation, and study of an item results in the addition of some study context features to the item’s context representation (in this way, BCDMEM can be viewed as a type of composite-trace theory). At test, every item is perfectly identified such that only the context features associated with the item can serve as a source of noise. Only if the associated context is sufficiently similar to the target study context is the item called “old”. Context noise offers a direct account of word frequency effects in recognition—items that occur more often in life are also more likely to have been associated with contexts that match the study context by chance. They also explain the null list strength effect, since strengthening an item has no effect on the item-context associations formed between other weaker items. However, context noise models must assume

other mechanisms in order to account for length effects (Dennis & Humphreys, 2001) or similarity effects (Dennis & Chapman, 2010).

Of course, these two sources of noise are not mutually exclusive (Criss & Shiffrin, 2004a). A variety of context-noise models assume that the context is, itself, partially determined by the studied items, thus allowing these models to mimic item-noise effects (Howard & Kahana, 2002; Dennis & Chapman, 2010). It is also straightforward to include pre-experimental memory traces in a separate-trace model such that context cues alone are no longer sufficient to isolate the memory traces relevant for the recognition task, an approach we will take in the model we present here. Finally, prior experience can influence how items are encoded (Nelson & Shiffrin, 2013), thus allowing for a certain type of long-term context noise. We defer fuller discussion of these issues until after we present our model.

1.2 Theories of response dynamics

Just as many theories of memory can be considered extensions of signal detection theory, a variety of models of response time represent dynamic versions of signal detection. As in signal detection, a decision-maker is tasked with making a binary decision (e.g., “old” vs. “new” in recognition) on the basis of incoming evidence (e.g., familiarity). However, instead of taking a single sample from a noisy distribution of evidence, models of response dynamics assume that many samples are taken and that these are accumulated over time until the decision-maker is sufficiently certain about which distribution the evidence is coming from. The evidence distributions are, as in signal detection theory, typically characterized as Gaussian, with a particular mean and variance. In statistical decision theory, this sequential sampling process is known as the sequential probability ratio test (SPRT; Wald, 1945,

1947), and was later explored as a description of human decision making behavior (Stone, 1960; Edwards, 1965). More generally, the SPRT is a kind of random walk, where each sample provides evidence for or against each of two options; this evidence is formalized as a likelihood ratio and, assuming evidence samples are independent, these ratios are multiplied (equivalently, the log-likelihood ratios are summed) until the accumulated likelihood for one option is sufficiently greater than the likelihood for the other.

Random walk models have also been used as models of human decision making (e.g., Link, 1975; Link & Heath, 1975), but none more so than the diffusion model proposed by Ratcliff (1978). A continuous-time version of a random walk, the key insight in Ratcliff's approach was that the mean of the evidence distribution from which samples are drawn—called the “drift rate”—could change from trial to trial. In particular, the distributions from which the drift rate in the diffusion process are drawn on each trial are, just as in most applications of signal detection theory, Gaussian distributions with particular means and variances. By allowing for trial-to-trial variability in drift rate, Ratcliff's diffusion model alleviates various problems faced by other random walk models, in particular, the often-incorrect prediction that responses of a particular type (e.g., an “old” or “new” response) have the same reaction time distributions regardless of whether the response was correct or not (Swensson & Green, 1977). Especially when the decision-maker's initial bias and motor execution/stimulus encoding time are also allowed to vary between trials, a diffusion model of this kind provides an excellent fit to two-choice response times in a variety of paradigms (Ratcliff & Rouder, 1998; Ratcliff & Smith, 2004). In addition, the accumulation of evidence samples to one of two fixed decision criteria is “optimal” in the sense of achieving a desired accuracy level in minimal time, suggesting that, to the extent that humans have evolved or learned to make decisions optimally, a diffusion process should

characterize their decision-making behavior (Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006).

Recently, it has been found that a very different type of model also fits data well and leads to similar conclusions about the underlying decision evidence: a ballistic accumulator model (S. Brown & Heathcote, 2005, 2008; Donkin, Brown, Heathcote, & Wagenmakers, 2011). Unlike random walk models, accumulator models do not necessarily assume that evidence in favor of one response is evidence against the other. Instead, a decision maker accumulates evidence in favor of each response and makes a decision as soon as one of the response accumulators reaches a criterial value. Although accumulator models have a long history in psychology (LaBerge, 1962; Vickers, 1970; Pike, 1973; Townsend & Ashby, 1983; Smith & Vickers, 1988; Smith & Van Zandt, 2000)—including in the study of recognition memory (Van Zandt, 2000)—they typically assume that the evidence accumulated within a trial is noisy. A ballistic accumulator model assumes no noise within a trial—hence the term “ballistic”. The model is nonetheless able to predict choice probabilities and response times quite well by allowing the drift rate and initial bias to vary randomly (according to particular distributions) between trials, although ballistic accumulator models are, by nature, unable to address situations in which the decision evidence changes over time (e.g., Tsetsos, Usher, & McClelland, 2011; Teodorescu & Usher, 2013); such situations provide evidence that competition between choice options—an integral feature of random walks—is necessary to explain observed decision behavior. While the assumption of non-interacting accumulators may not suffice to describe human decision making mechanisms, a model of evidence accumulation that assumes that noise only arises from the stimulus itself—i.e., not from the random perturbations assumed in a diffusion model—does a good job of explaining perceptual decision making in both humans and rats (Brunton,

Botvinick, & Brody, 2013). Thus, the use of random perturbations in the diffusion model and of trial-to-trial stochastic drift rates in the linear ballistic model may simply be statistical descriptions of stimulus-driven noise, rather than inherent properties of decision making mechanisms.

While it may be difficult behaviorally to distinguish between a random walk model like Ratcliff's diffusion model and a ballistic accumulator model (especially if one allows the form of the evidence distributions to vary; M. Jones & Dzhafarov, 2014), there is some neural evidence for the kind of noisy accumulation to a criterion posited by random walk models. In perceptual decision making in primates, a variety of cortical areas appear to be involved in integrating information over time, akin to the integration of information in a random walk or diffusion. These include “build-up” neurons in the superior colliculus (Munoz, Dorris, Paré, & Everling, 2000) that control saccades, and neurons in the frontal eye fields (Gold & Shadlen, 2000) and lateral intraparietal area (LIP; Roitman & Shadlen, 2002) that appear to integrate information from area MT to determine the dominant direction of random dot motion. Although neurons in LIP exhibit stimulus-driven changes in firing rate up to 800 ms after stimulus onset, the ability of a new stimulus to perturb the activity of these neurons diminishes over time (Huk & Shadlen, 2005). This is consistent with a neuron that integrates information in a linear fashion up to a boundary, at which point integration ceases (Kiani, Hanks, & Shadlen, 2008). Ditterich (2006), however, found that linear integration, even with trial-to-trial variability in drift rates, did not account for the behavioral data of monkeys in a motion-discrimination task: Observed RT distributions were not as skewed as those predicted by the diffusion model. Instead of linear integration to a boundary, LIP neurons appeared to implement a time-varying gain function on the input, such that later inputs have more influence than earlier inputs, consistent with either a

“leakage” of early information (e.g., Busemeyer & Townsend, 1993; Usher & McClelland, 2001) or an “urgency” signal that ramps up the gain on evidence over time (Cisek, Puskas, & El-Murr, 2009).

1.3 Overview

Despite the many parallel advances in the understanding of memory and of response dynamics, theorists have only rarely attempted to link the two, despite a wealth of evidence in this domain. Detailed analyses of response time distributions (e.g., Sternberg, 1969; Ratcliff & Murdock, 1976) lead to the development of the diffusion model (Ratcliff, 1978), originally a model of recognition, which has since developed into a general model of decision making. Yet while this theory and the subsequent decision model of Hockley and Murdock (1987) were able to predict many patterns of observed data, these models remain agnostic about two key components of the recognition process: 1) The source of the memory evidence on which decisions are based; and 2) how this evidence itself evolves over time. To address the first point, it would be possible, as suggested by Hockley and Murdock (1987), to use one of the static memory models discussed above to generate evidence distributions that are then fed into a second decision stage. One prominent model of both categorization and recognition, the Exemplar-Based Random Walk (Nosofsky & Palmeri, 1997; Nosofsky et al., 2011), results in a model of this form, where the Generalized Context Model (Nosofsky, 1986) defines a distribution of evidence values in terms of the probability of retrieving particular exemplars in response to a probe; this sequence of retrieval events is then used to drive either a random walk (e.g., Nosofsky et al., 2011; Nosofsky, Cox, Cao, & Shiffrin, 2014) or a set of linear ballistic accumulators (Donkin & Nosofsky, 2012b; Nosofsky, Cao, Cox, & Shiffrin, 2014), closely matching both accuracy

and response time distributions.

As to the second point, researchers have traditionally investigated the evolution of memory evidence over time using speed-accuracy trade-off procedures. These employ signals that interrupt processing at various times to assess how decision evidence evolves. Evidence from these paradigms has supported the development of models in which information about the test stimulus grows over time (Brockdorff & Lamberts, 2000; Cohen & Nosofsky, 2003), although such models have proven difficult to extend to free response paradigms (Lamberts, Brockdorff, & Heit, 2003), impeding their wider acceptance.

The purpose of this research is to develop a dynamic approach to recognition memory that explicitly models the structure of memory as well as how retrieval evolves over time. This approach codifies a variety of principles within a quantitative model that closely fits both speed-accuracy trade-off data as well as accuracy and response time distributions. By taking a dynamic approach, we then investigate various phenomena in memory that are difficult to adjudicate when considering accuracy alone, affording insight into both decision processes and retrieval. Model fits to free-response data from several experiments (Nobel & Shiffrin, 2001; Starns, Ratcliff, & McKoon, 2012; Rae, Heathcote, Donkin, Averell, & Brown, 2014) illustrate how the joint evolution of retrieval and decision are affected by both study manipulations (study time, list length) and instructions (bias and speed vs. accuracy emphasis). They also provide evidence that word frequency effects in recognition are mediated by semantic knowledge rather than interference from specific prior episodes, an inference that would be impossible by considering accuracy alone. We then show how the dynamic model can be extended to account for paradigms like associative recognition and list discrimination in which secondary recall-like processes have been implicated. Echoing Brockdorff and Lamberts (2000), we show that in many cases, such paradigms are bet-

ter accounted for by a model in which different information is used to prompt retrieval at different times, with only source memory providing evidence for the operation of recall. This inference is, again, only possible by taking a dynamic approach: When the relevant target item is made stronger (e.g., by repetition or increased study time), a recall model would predict that this trace would be retrieved *faster* and would thus enable a faster rise to asymptotic performance. Only source judgments (Hintzman & Caulton, 1997) show this pattern, suggesting that recall is not necessary to explain associative recognition, list discrimination, or the rejection of highly similar foils and that such situations are better described by a model in which different information enters the memory probe at different times.

Chapter 2

A Dynamic Model for Recognition Memory

The dynamic approach to recognition memory that we adopt (introduced in Cox & Shiffrin, 2012) has its roots in a variety of modern approaches to the dynamics of decision making (Ratcliff, 1978; S. Brown & Heathcote, 2008) and categorization (Nosofsky & Palmeri, 1997; Lamberts, 2000), and reflects a recent trend toward developing models of the time-course of recognition memory (Brockdorff & Lamberts, 2000; Nosofsky et al., 2011). A central assumption that we share with all of these models is that information accumulates over time. While often taken for granted, this assumption is not shared by, say, the recognition model of (Hockley & Murdock, 1987), in which memory evidence is sampled repeatedly but is not accumulated over time, entailing the assumption of separate fast guessing processes or differences in encoding to explain speed-accuracy trade-offs (Gronlund & Ratcliff, 1991; Hockley & Murdock, 1992). What we do here is lay out a framework that describes the information used for recognition, and how it is accumulated over time to make a decision. This framework is formalized within a mathematical model that can then be used to quantitatively test claims of theoretical interest, which we proceed to do in subsequent sections. We adopted this particular formalization because we wished to represent these theoretical claims in the clearest way possible in order to show how the proposed mechanisms lead to particular predictions. Alternative formalizations are certainly possi-

ble without changing the core of our dynamic approach. Some alternatives might lead to better quantitative fits, but may not aid understanding of the principles and mechanisms needed to produce those fits. Nonetheless, several variations on our current formal model are suggested throughout this section and in the discussion as targets of future work.

The process leading to a recognition decision is based not on a constant rate of evidence accumulation, but a time-varying interaction of retrieval of features from knowledge, growth of an episodic memory probe, retrieval that varies as the probe varies, and evidence accumulation that varies with what is retrieved. When a test stimulus is presented, its physical form plus context are used to access knowledge (for a word, this would be a lexical trace in semantic memory). Features—extracted both from perception and knowledge—gradually over time and join the features of the current context to form an accumulating probe in short-term memory. At each epoch of time, the probe is compared in parallel to traces in episodic memory, producing similarity-based activations of each trace characterized as likelihood ratios. The average likelihood ratio at time t is termed the familiarity at time t . Familiarity changes over time as the probe changes; successive changes in familiarity are accumulated as time passes, with a response occurring when the accumulated change reaches an “old” or “new” boundary.

Note that the probe of memory begins with context features only, before any features are extracted from the test item. Although context might not change radically from one test trial to another following a single list, it certainly would change enormously between retrieval situations in everyday life. We therefore assume the recognition system has been constructed to adjust for such changes, and uses the starting familiarity value—based on context—as a reference point. That is why we assume that the decision is based on accumulating changes in familiarity from an initial value.

Most of the traces in memory will tend to match the probe very poorly, of course, so only those traces whose likelihood exceeds a threshold value are considered “activated” at any one time and participate in the average that produces familiarity at that time. The traces that do participate are of three main types: 1) A list trace that had been produced by the target (if a target is tested)—this trace tends to match well in both content and context; 2) list traces that do not match the test item—these traces tend to match in context but not content; 3) traces of the test item from events prior to the current list, in the experiment or in life—these traces tend to match in content but not context.

The test stimulus is represented as a “probe”, which consists of a set of features held in short-term memory. The probe is compared, in parallel, to all the traces stored in memory; each probe/trace comparison results in a likelihood value that represents the probability that the features stored in the probe and the features stored in the trace encode the same event; likelihood is, then, a measure of probe/trace similarity. Most of the traces in memory will tend to match the probe very poorly, of course, so only those traces whose likelihood exceeds a threshold value are considered “activated” at any one time. Familiarity is the average likelihood ratio over these activated traces, reflecting a sense of global match between current experience (the test stimulus) and the contents of memory (the set of activated traces).

Familiarity changes over time as the probe changes. Before any stimulus is presented, only context features are available and, thus, the probe only contains context features. Once a test stimulus appears, content features begin to accrue in the probe. If a new feature increases familiarity (i.e., it tends to increase the match between the probe and the set of active traces), this is evidence that the test stimulus is “old” (it was present in recent experience on the study list). Conversely, if a new feature decreases familiarity, this is evidence

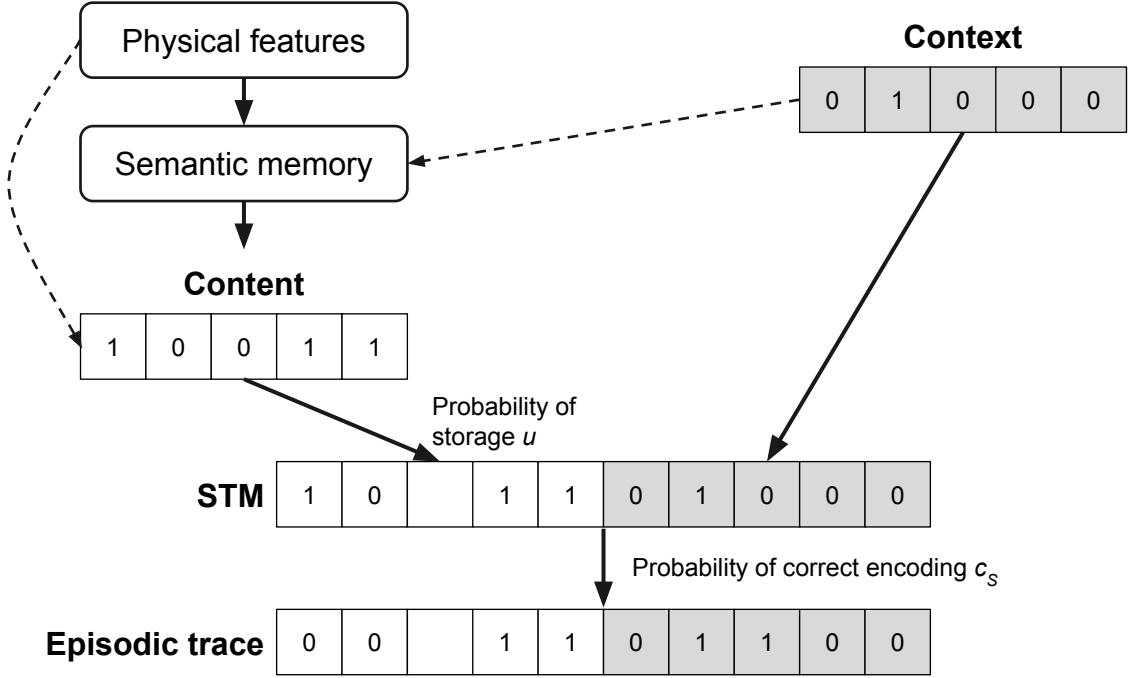


Figure 2.1: Schematic depiction of the formation of an episodic memory trace. See text for detail.

that the test stimulus is “new”. Thus, we presume that an old/new recognition decision is made on the basis of *accumulated changes in familiarity* from a starting point. When these accumulated changes reach one of two time-varying thresholds—an upper one for “old” and a lower one for “new”—the corresponding response is made. The time taken to reach one of those thresholds is related to the response time. We now describe a quantitative model that embodies these ideas. For reference, the parameters of our model are summarized in Table 2.1 while other model notation is summarized in Table 2.2.

2.1 Structure of memory traces

Studying a list of items leads to the formation of a set of traces in memory. One trace is formed for each studied item. Each trace consists of a set of features, some of which represent the content of the studied item, and others of which represent the context in

Table 2.1: A summary of the parameters of the model of recognition dynamics.

<i>Parameter</i>	<i>Description</i>
u	Probability of storing a content feature in a memory trace at study.
c_S	Probability of encoding a feature correctly in a memory trace at study.
c_T	Probability of encoding a feature correctly in the memory probe at test. Usually assumed to be 1.
g	Probability of a feature in the probe and a memory trace having matching values by chance. Usually assumed to be $\frac{1}{2}$.
N_C	Number of content features. Arbitrarily set to 30 unless otherwise noted.
N_X	Number of context features. Arbitrarily set to 30 unless otherwise noted.
N_C^*	Participant’s estimate of the number of content features needing to be sampled. Usually assumed equal to N_C unless otherwise noted.
A_0	Initial separation between “old” and “new” response boundaries in free response.
b	Response bias; $b > \frac{1}{2}$ reflects a bias to say “old” while $b < \frac{1}{2}$ reflects a bias to say “new”.
θ	Response criterion for response signal tasks.
T_R	Residual time attributable to stimulus detection and response execution.
t_0	Time at which feature sampling begins (usually a component of T_R).
ρ	Amount of real-time (seconds or milliseconds) for each time-step in the model.

Table 2.2: Notational reference.

<i>Notation</i>	<i>Description</i>
$\lambda_i(t)$	Likelihood ratio for trace i at time t , denoting the relative probability that the probe and trace i encode the same event versus the probability that they encode different events, given by Equation 2.6.
$\phi(t)$	Familiarity at time t , defined as the logarithm of the average of all likelihood ratios $\lambda_i(t)$ that are greater than one, given by Equation 2.8.
$x(t)$	Accumulated change in familiarity from starting time t_0 , given by Equation 2.9. This is the evidence used for recognition decisions.
$v(t)$	True probability that all content features have been sampled into the probe by time t , given by Equation 2.1.
$v^*(t)$	Participant’s estimated probability that all desired content features have been sampled into the probe by time t , given by Equation 2.10. May differ from $v(t)$ if participants are willing to commit to a decision with greater or fewer sampled features.
$B_O(t), B_N(t)$	In free-response, the “old” and “new” decision boundaries, given by Equation 2.12. When $x(t)$ crosses one of these boundaries, the corresponding decision is made.

which the item was studied. We presume that the number of such features that can be encoded in a brief time (most study times are no longer than a few seconds) is limited, with a maximum of N_X context features and N_C content features, although these limits may differ depending on condition. Here, we use “context” to mean the general situation in which an event occurs, rather than the other items that are presented nearby in time (cf., Howard & Kahana, 2002; Klein et al., 2007). Thus, context features pertain to the time, location, and internal state of the participant at the time of study. These features are presumed to be ambient in the environment and are thus able to be encoded rapidly (indeed, additional context does not seem to be stored after about 1 second; Malmberg & Shiffrin, 2005). Content features, on the other hand, pertain to the semantic and perceptual aspects of the studied item, and the degree to which they are stored in the trace depends strongly on the amount of time available for study. For a word, such features would include the word’s meaning as well as its phonological and orthographic characteristics. For an image, content features might be primarily perceptual, but also include semantic features to the extent that the image contains known objects or scenarios. Critically, then, some content features might come directly from the percept of the studied item (e.g., the font in which a word is printed), but many will come from the participant’s own semantic memory (e.g., the meaning of a word). Thus, how a study item is encoded—which features are used and what their values are—depends on the participant’s prior knowledge stored in semantic memory. Access to semantic memory depends not only on the physical features of the stimulus, such as the letters of a word, but can also be affected by context, as has been found in studies of lexical decision and naming (Wagenmakers et al., 2004). Finally, although we do not model it here, just as semantic memory is used to determine the features encoded in an episodic trace, the episode augments the knowledge stored in semantic memory (Nelson

& Shiffrin, 2013). This process of forming a memory trace that stores both content and context features is depicted schematically in Figure 2.1.

In a given study condition, an available content feature is stored in the memory trace with probability u . Various factors can influence u , but it is especially susceptible to variations in study time or number of repetitions of the item (for simplicity, we model repetition as strengthening an existing trace rather than forming a new trace; Shiffrin et al., 1990)¹. Factors that influence access to semantic memory—and therefore the availability of knowledge-based features like word meaning—may also affect u , such as study modality, priming, or item type (e.g., pseudowords or random dot patterns are unlikely to have any semantic features). For simplicity—and because we are generally agnostic about exactly what these features represent—we assume that all features are binary-valued. If a feature is stored in the trace, its value (e.g., “0” or “1”) is stored correctly in the trace with probability c_S . Otherwise, the stored value is chosen at random according to the base-rate probabilities for that feature, where g is the probability of storing “1” (typically $g = \frac{1}{2}$). The base-rate information for a given feature is presumed to reside in semantic memory, given that it represents information accrued over a long period of time.

¹It might be plausibly argued that, when an item is repeated, different features are used to encode it a second time. If these new features joined those already encoded in prior presentations, this would effectively increase the size of the memory trace beyond the limits of short-term memory. However, because those same limits are imposed on the memory probe during recognition, such a “lengthening” of the trace has the same effect as increasing u . At recognition, only N_C content and N_X context features can be used in the first place, so the match between the probe and a given trace depends only on those features, which may be just a subset of those actually encoded in the trace. Thus, it is still reasonable, in this situation, to model repetition via the u parameter.

2.2 Feature sampling

When an item is presented at test, a similar encoding process is used to construct a memory probe in short term memory. We assume that the same capacity limits that apply at study also apply at test, with a maximum of N_X context and N_C content features able to be held in the probe. Before any test stimulus is presented, the probe contains only the N_X features of the test context, which we typically assume to be identical to the target study context (although we address list discrimination later, in which this is not necessarily the case). Once a test item appears, the physical features of the test stimulus are used in conjunction with context to access semantic memory. At this point, which we call time t_0 , content features are available to enter the probe, but they do not do so all at once. Instead, they are sampled over time from the pool of available features. At any given time after t_0 , one of the N_C content features is chosen uniformly at random. If that feature does not yet have a value in the probe, the test item's value on that feature enters the probe. With probability c_T , the correct value enters the probe; otherwise, a value is chosen at random according to the base-rate for that feature, as above. In most applications, we assume $c_T = 1$. If a feature already has a stored value, we assume that value is retained in the probe. Thus, the probability of sampling a new value into the probe decreases over time, with the probe eventually becoming saturated with features. The probability that the probe is complete at time t is

$$v(t) = 1 - \left(1 - \frac{1}{N_C}\right)^t. \quad (2.1)$$

2.3 Familiarity

2.3.1 Comparison between probe and trace

At each time-step t , the probe consists of N_X context features along with whatever content features have been sampled by that time. At each time-step, the probe is compared in parallel to all of the traces in memory. The number of traces is quite large, so for practical reasons we only explicitly model comparisons with the N traces formed during list study (one per study item, unless there are spaced repetitions) as well as K traces of the test item from prior life history. If the test item is a target, one list trace will tend to match in both content and context. Otherwise, list traces will tend to match in context, but only randomly in content. History traces, conversely, tend to match in content but not in context (here we simplify and assume that context features are randomly sampled for each history trace). The many other traces in memory tend not to match well in either content or context, so we assume they are not activated and do not model them explicitly.

The comparison between the probe and a given trace i results in a likelihood ratio $\lambda_i(t)$, reflecting the relative probability that the probe and trace encode the same event (the same item in the same context) or a different event. The likelihood ratio is computed by aligning and comparing the features in the probe and trace. There are five possible outcomes of any particular feature comparison:

∅: No value is stored for this feature in either the probe or trace.

P: A value is stored in the probe, but not in the trace.

T: A value is stored in the trace, but not in the probe.

M: A value is stored in both the probe and the trace, and the values match.

N: A value is stored in both the probe and the trace, but the values do not match.

An outcome of \emptyset does not provide any evidence about whether the probe and trace encode the same or different events. Although it will not be relevant to the studies reviewed here, we note that outcomes P and T can provide evidence that the probe and trace encode different events, as described by Cox and Shiffrin (2012). This occurs when trace encode items that are not of the same type, for example, if the study list contains both words and pictures. A word is likely to be encoded with different features (e.g., orthographic, phonological) than a picture. If tested with a picture, the probe will tend to be encoded with features that overlap with picture traces, but not with word traces, leading to a larger number of M and N outcomes with picture traces and P and T outcome with word traces. Thus, outcomes P and T can be used to correctly reject a test item that comes from a different class, in accord with findings that interference only manifests as a function of the number of studied items of the same type as the test item (Criss & Shiffrin, 2004b).

Assuming the probe and trace both have a value encoded for a particular feature, the probability of a feature match (M) between the probe and trace, given that they encode the same item, is

$$p_{M|S} = c_S c_T + c_S(1 - c_T)g + (1 - c_S)c_T g + (1 - c_S)(1 - c_T)[g^2 + (1 - g)^2], \quad (2.2)$$

which simplifies to $p_{M|S} = c_S + (1 - c_S)g$ when $c_T = 1$. The probability of a match if the probe and trace encode different items is

$$p_{M|D} = g, \quad (2.3)$$

where g is the probability of a match by chance. The probability of a mismatch (N) given that probe and trace encode the same item is

$$p_{N|S} = c_S(1 - c_T)(1 - g) + (1 - c_S)c_T(1 - g) + 2(1 - c_S)(1 - c_T)g(1 - g), \quad (2.4)$$

which simplifies to $p_{N|S} = (1 - c_S)(1 - g)$ when $c_T = 1$. Finally, the probability of a mismatch if the probe and trace encode different items is the probability of a mismatch by chance

$$p_{N|D} = 1 - g. \quad (2.5)$$

Because we assume independence between features, the likelihood ratio for the entire trace is the product of the likelihood ratios of each feature. Letting $N_M(t)$ and $N_N(t)$ denote the number of matching and mismatching features between the probe and trace i at time t (and ignoring for now the number of P and T features), we have

$$\lambda_i(t) = \left(\frac{p_{M|S}}{p_{M|D}} \right)^{N_M(t)} \left(\frac{p_{N|S}}{p_{N|D}} \right)^{N_N(t)}, \quad (2.6)$$

which simplifies to

$$\lambda_i(t) = \left[1 + c_S \left(\frac{1 - g}{g} \right) \right]^{N_M(t)} [1 - c_S]^{N_N(t)} \quad (2.7)$$

when $c_T = 1$.

2.3.2 Activation and familiarity

Familiarity, denoted $\phi(t)$, is the logarithm of the average likelihood ratio among those traces that are active at time t^2 . It can thus be considered a “log-odds” that there exists a trace in memory that encodes the same event as the probe. The notion that a trace must be “active” to be considered arises from the fact that episodic memory undoubtedly contains a great many traces which are not relevant to the current recognition situation. By imposing an activation threshold, we ensure that memory is not overwhelmed with irrelevant traces. We

²Although aggregation schemes other than the average are possible, like a sum or maximum, it is unlikely that adopting these other formulations would substantially alter the model’s predictions (Shiffrin & Steyvers, 1998).

propose a simple rule, that a trace is considered “active” if its likelihood ratio is greater than one; this means that the trace matches sufficiently well in either content or context features in order to be a plausible match to the test item. Familiarity can then be denoted

$$\phi(t) = \log \langle \lambda_i(t) : \lambda_i(t) > 1 \rangle. \quad (2.8)$$

2.3.3 Familiarity over time

Initially, the probe only contains context features. Thus, at time t_0 , familiarity reflects the match between the test context and the contents of memory. Traces of recent events—such as study items, which tend to share test context features—will thus tend to be active at this time. Familiarity changes after time t_0 as content features are sampled and join the probe. On any one trial, familiarity will tend to be noisy, governed by the order and timing with which features are sampled, as well as the various properties of the traces stored during study. Each new feature sampled provides more statistical evidence about whether there is a trace in memory that encodes the same event as the probe. Some historical traces—which match in content but not in context—may become active as their matching content features overcome their initially mismatching context features. On the other hand, non-matching list traces—which match in context but not in content—will have likelihood ratios that tend to decrease over time as the increasing number of mismatching content features lowers their likelihood ratios. If, however, there is a matching target trace, its likelihood ratio will tend to increase beyond the initial good match to context; this single trace that matches well on both content and context will grow to dominate the average likelihood.

On average, familiarity increases for targets and decreases for foils, as shown in Figure 2.2. This change in familiarity over time is thus statistically diagnostic of whether an item had been studied or not, and we presume this is the basis for a recognition decision.

Familiarity distributions

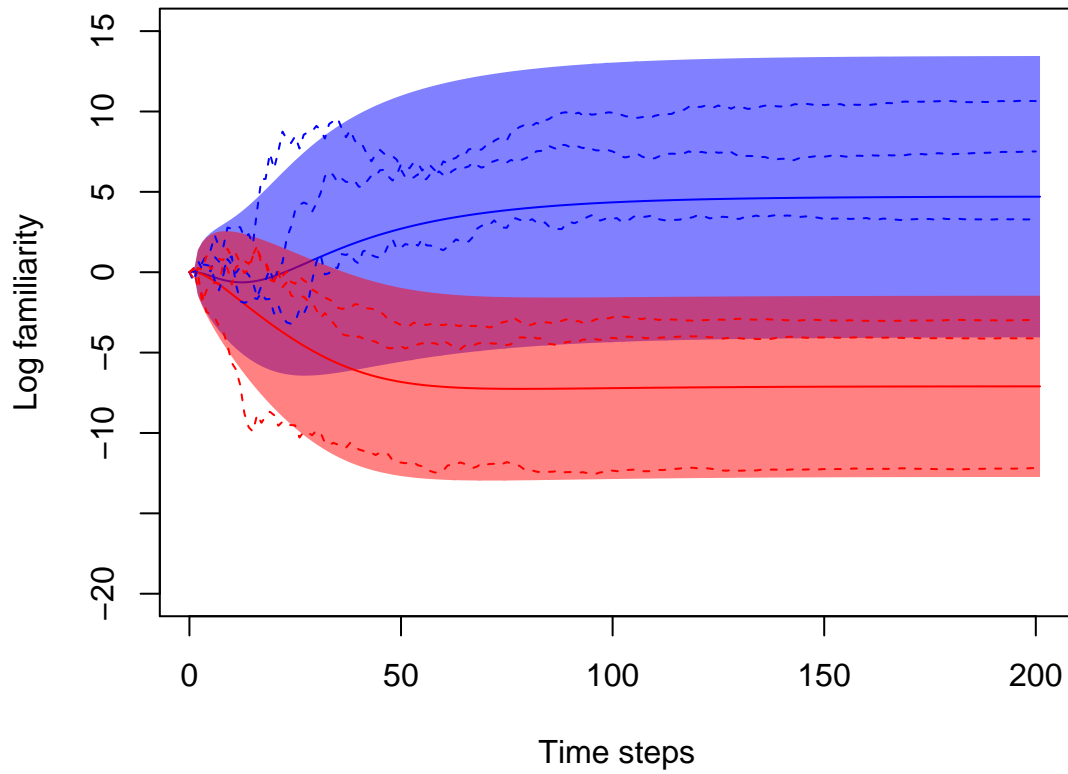


Figure 2.2: Examples of the evolution of familiarity over time. Solid lines depict mean familiarity over time, shaded regions are the central 95% of familiarity values at each time, and dashed lines show example familiarity trajectories from single trials.

Specifically, recognition decisions are based on the accumulated change in familiarity from the starting time t_0 , denoted $x(t)$:

$$x(t) = \sum_{\tau=t_0}^{t-1} \phi(\tau+1) - \phi(\tau) = \phi(t) - \phi(t_0). \quad (2.9)$$

By making recognition dependent on the *change* in familiarity over time, rather than on an absolute level of familiarity, one does not need to assume differing criteria for experimental conditions or stimulus classes which may differ in their absolute familiarity (Cox & Shiffrin, 2012). One can also view $x(t)$ as conditionalizing familiarity $\phi(t)$ on an initial value determined by context, $\phi(t_0)$. Because context features are persistent in the environment, rather than changing rapidly like content features, they provide a natural baseline level of familiarity against which to judge any new information. Finally, using the accumulated change in familiarity from a starting point emphasizes the important role played by the features that initially occupy the probe. Although these will usually be content features, other features might be present initially as a function of experimental manipulations like priming; the effect of this will be explored later.

2.4 Predicting response-signal data

The response signal method (Reed, 1973, 1976) involves presenting a test stimulus for a fixed amount of time before a signal—either visual or auditory—is given at which point the participant must respond immediately with their current best guess about whether the item is old or new. By varying the lag between stimulus onset and the response signal, it is possible to estimate a speed-accuracy trade-off (SAT) curve which measures how the probability of responding “old” to a test item changes as additional time is allowed to process the stimulus. Although it is possible to make more complex assumptions about

the cognitive processes that underlie SAT curves, for instance that they reflect a mixture of ongoing and already-finished processes (Meyer, Irwin, Osman, & Kounios, 1988; Ratcliff, 1988), we adopt a much simpler assumption. As just described, we assume that features of the test item begin to be sampled to join the context features already present in the probe of memory, starting at time t_0 after stimulus onset. Each model time step is assumed to take a constant amount of time p .³ Participants accumulate changes in familiarity as just described until a response signal occurs. At that point, a participant is predicted to make an “old” response if the accumulated change in familiarity is above a threshold θ ; otherwise, the participant makes a “new” response. If a response signal occurs before t_0 , a participant guesses “old” with probability

$$\frac{1}{1 + \exp(-\theta)},$$

in accord with the assumption that $x(t)$ lies on a log-odds scale.

Although it may be possible for people to change their response criteria between different response signal lags, this is rare (Hintzman, Caulton, & Curran, 1994) and only seems to occur when experimental conditions require such adaptation, such as when the proportion of targets varies across lags (Heit, Brockdorff, & Lamberts, 2003). More important for our purposes, allowing for θ to change while the accumulated change in familiarity is also changing would seem to make the model unidentifiable. Thus, we assume θ is constant across time. Typical approaches to response signal data assume that the response time also contributes to stimulus processing time, but we assume instead that only the time between stimulus onset and response signal allows for sampling features from the stimulus and that the response time reflects only the time for response selection and execution. Response times are typically quite brief (200 ms) and, although they do tend to be longer at earlier

³This assumption is probably overly simplistic, but suffices to explain the data we consider here.

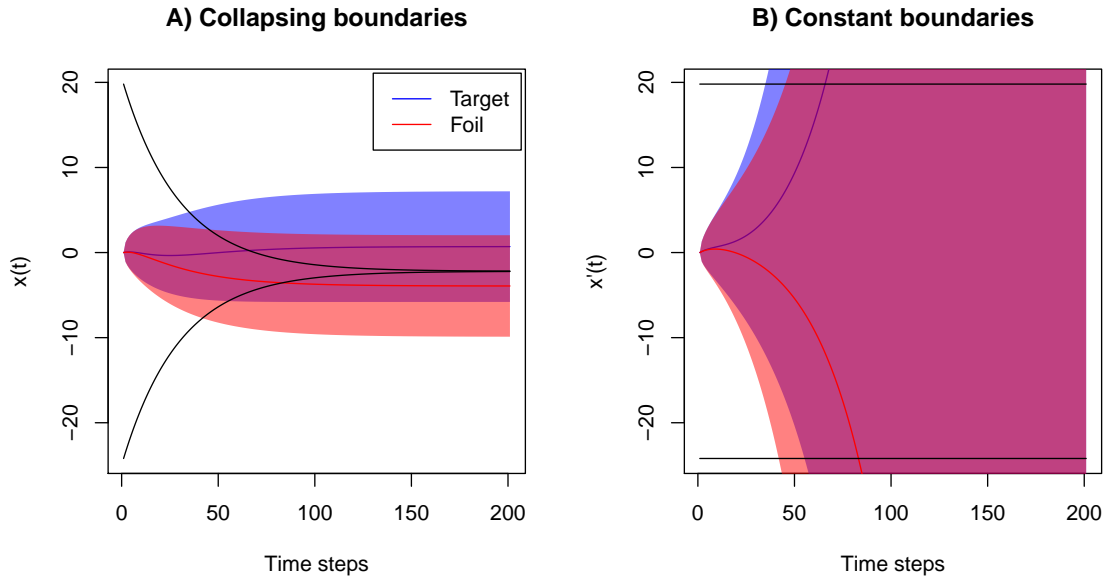


Figure 2.3: The model’s predicted distributions of familiarity over time for target and foil items, in relation to an upper (“old”) and lower (“new”) response boundary. Solid colored lines depict mean familiarity while colored regions show 95% confidence regions about the mean. A) In the standard formulation of the model, boundaries collapse over time to ensure a response is made. B) An equivalent formulation “stretches” familiarity differentially over time such that response boundaries are constant.

lags, they are generally not affected by the identity of the stimulus, suggesting that little additional processing of the stimulus itself could occur between the response signal and the response itself. Although this is surely an oversimplification, we were able to obtain good fits despite it. Moreover, because response times tend to vary between signal lags only at the very earliest lags, they do not affect the qualitative form of either the observed or predicted SAT curves.

2.5 Predicting free-response data

When participants are allowed to make their response at any time, we assume that participants continue to accumulate changes in familiarity until either an upper or lower boundary is reached. Reaching the upper boundary leads to an “old” response while reaching the lower one leads to a “new” response. However, because familiarity eventually reaches a stable asymptote as the probe becomes saturated with features, these boundaries cannot be constant over time. If they were, it would lead to situations in which a decision was never made if asymptotic familiarity lied between the two bounds. Thus, we assume that the decision boundaries over time as a function of the expected completeness of the probe, namely,

$$v^*(t) = 1 - \left(1 - \frac{1}{N_C^*}\right)^t, \quad (2.10)$$

which differs from the equation for the *actual* probe completeness, $v(t)$ in Equation 2.1, only in that N_C^* might differ from the true value of N_C in situations like speeded responding, when participants might be willing to commit to a response with fewer features than are actually available (typically we assume $N_C^* = N_C$). The decision boundaries begin a certain distance apart, A_0 , at starting time t_0 . The distance between the “old” and “new” boundaries at time t , denoted $A(t)$, diminishes over time:

$$A(t) = (1 - v^*(t))A_0.$$

The parameter b , $0 < b < 1$, reflects the degree of asymmetry between the two bounds, such that the “old” and “new” boundaries at time t are, respectively,

$$B_O(t) = (1 - b)A(t) \quad (2.11)$$

$$B_N(t) = -bA(t). \quad (2.12)$$

The initial “old” boundary is, then, $(1 - b)A_0$, the initial “new” boundary is $-bA_0$, and the boundaries collapse symmetrically toward a common asymptote, $(\frac{1}{2} - b)A_0$. If $b = \frac{1}{2}$, the bounds are symmetrical, if $b < \frac{1}{2}$ boundaries are “new”-biased, and if $b > \frac{1}{2}$ boundaries are “old”-biased. The form of the boundary collapse is depicted in Figure 2.3A.

The number of time-steps needed to reach a boundary, t , determines the response time. Just as in the response-signal paradigm, predicted response time \widehat{RT} is assumed to be a linear function of model time-steps

$$\widehat{RT} = T_R + \rho t \quad (2.13)$$

where T_R represents the residual time needed to detect the test stimulus and execute a motor response and ρ is the amount of real time per model time-step. Note that T_R includes the amount of time before stimulus sampling begins (t_0), although this exact time cannot be estimated in free response tasks. Although it would certainly be more realistic to allow T_R (and t_0) and ρ to be random variables that vary from trial to trial, it was not necessary to obtain the good quantitative fits we report here, therefore we chose not to introduce this additional complexity.

2.5.1 Collapsing boundary considerations

Collapsing decision bounds means that, as the probe accumulates more features, a participant is willing to commit to a decision on the basis of weaker and weaker evidence. Collapsing boundaries have been implicated in a variety of decision-making models (Balakrishnan & Macdonald, 2011; Busemeyer & Rapoport, 1988; Ditterich, 2006; Frazier & Yu, 2008; Ratcliff & Frank, 2012; Zhang, Lee, Vandekerckhove, Maris, & Wagenmakers, 2014), including the recognition decision model of Hockley and Murdock (1987). Al-

though the Hockley & Murdock model also assumed an exponential form for collapsing boundaries, in our case, this assumption is justified as a function of our model’s feature sampling mechanism, as outlined above. We also note the analogy between collapsing bounds and a time-dependent “urgency signal” (Cisek et al., 2009). Urgency is reflected in a time-dependent rise in firing rates of all lateral intraparietal neurons—whose firing rates are typically associated with accumulated evidence (Roitman & Shadlen, 2002)—regardless of which response the neuron is associated with. Studies finding an urgency signal typically employ simple perceptual decisions, so it is not clear if exactly the same mechanism should be implicated in memory retrieval, but it is certainly reasonable to expect that participants will not search memory indefinitely, and collapsing decision boundaries allow for a straightforward implementation of this notion.

It is, however, possible to consider our model as having fixed boundaries, but with an accelerating growth of familiarity over time, as shown in Figure 2.3B. This equivalence results from a simple warping of familiarity over time in order to “stretch” the values of familiarity in order to keep the boundaries constant over time. Specifically, at each time step, the current accumulated change in familiarity, $x(t)$, is transformed according to

$$x'(t) = \frac{x(t) - \left(\frac{1}{2} - b\right) A_0}{1 - v^*(t)} + \left(\frac{1}{2} - b\right) A_0$$

where $v^*(t)$ is the expected probe completeness at time t , given in Equation 2.10, A_0 and b are as defined above, and $\left(\frac{1}{2} - b\right) A_0$ is the midpoint between the “old” and “new” boundaries. Similar translations between different process representations have been used to deal with non-homogeneous processes in psychology and in the mathematical literature on diffusion processes (e.g., Ricciardi, 1976; Heath, 1992; Smith, 2000). This kind of acceleration toward a final response might result from the operation of self-excitation and/or

competition between accumulators corresponding to “old” and “new” responses. In this way, a response that had accumulated a large amount of evidence might be able to accelerate toward a boundary and possibly even suppress other responses that had only accumulated weak evidence; this could occur even in the absence of new incoming evidence. Such excitatory and inhibitory processes are embodied in various neural models of decision making, such as the leaky competing accumulator (Usher & McClelland, 2001), and there is good evidence that response competition, beyond input competition, is necessary to explain patterns of data in perceptual decision making (Teodorescu & Usher, 2013).

Although it would be possible to reformulate the model in the ways just described without changing its predictions, we believe it is easier to interpret under the view that familiarity asymptotes whilst decision boundaries collapse over time. Our model of familiarity over time without any kind of warping closely matches the response curves obtained using the speed-accuracy trade-off procedure, as will be shown throughout this article. It is unclear to us, then, why response-signal experiments would not entail warping familiarity over time but free-response instructions would; it seems more sensible to us to assume that it is the decision bounds that change. If one still wanted to adopt constant boundaries with warping evidence, it would still be possible to model SAT functions by assuming that responses in the SAT paradigm are a mixture of guesses and already completed decisions (i.e., those that have already reached one of the fixed decision bounds; Ratcliff, 1988).

Finally, as an aside, we consider the notion of “optimality” as it relates to decision boundaries. In two-choice situations with stationary evidence distributions, it is optimal to accumulate information until one of two fixed boundaries is reached (Wald, 1945, 1947; Stone, 1960; Edwards, 1965; Bogacz et al., 2006). However, in situations where there is a stochastic response deadline (Frazier & Yu, 2008) or one is limited by total time

and not number of trials in a task, it is optimal to allow response boundaries to collapse, reflecting the diminishing returns on collecting additional evidence (Thura, Beauregard-Racine, Fradet, & Cisek, 2012; Drugowitsch, Moreno-Bote, Churchland, Shadlen, & Pouget, 2012). The memory evidence posited by our model can be seen as imposing a kind of stochastic deadline, since no new information accrues once the probe is saturated with features. It is possible to compute the optimal decision boundaries under the assumption that evidence has a zero mean on average but reaches an asymptotic value according to Equation 2.10 via dynamic programming (Bellman, 1952). The details of this computation would take us too far afield; suffice it to say that although there is no closed-form expression for these optimal boundaries, the numerical approximation obtained by dynamic programming is approximated well by an exponential function, albeit one that tends to approach asymptote more slowly than $v^*(t)$ in the fits reported here (Equation 2.10). This suggests that participants are either not allowing decision boundaries to collapse optimally, or are optimizing something other than expected accuracy. In any event, our model fits observed data considerably better with the form of boundary collapse outlined above than with one derived from optimal control theory.

Chapter 3

Validations and Insights

Our goals in this section are twofold: First, to demonstrate that the dynamic model for recognition just described provides a good account of both free response and response signal paradigms. Second, by fitting different versions of our model and comparing their quantitative fits, we show how our dynamic model can shed light on memory phenomena that are obscure when considering only recognition accuracy. To accomplish these goals, we fit our model to three sets of free response data and two sets of response signal data (additional response signal fits are presented in the next section). These datasets manipulate a variety of factors and come from different participants in different labs at different times. Thus, the ability of our model to fit all of these datasets is diagnostic of its ability to explain recognition dynamics generally.

3.1 Fitting procedures

In order to efficiently fit our model to data, we developed a continuous approximation to the model that enables much faster computation of the trajectories of familiarity over time and the resulting predictions of joint RT and accuracy. The approximation, which allows us to avoid tedious simulation, is described in detail in the Appendix. Suffice it to say, however, that the approximation produces behavior that is nearly indistinguishable from simulations

of the full model, given the same parameters.

3.1.1 Quantile maximum likelihood

To jointly fit accuracy and free response time distributions, we employ the quantile maximum likelihood procedure (QML; Heathcote, Brown, & Mewhort, 2002). In this procedure, the .1, .3, .5, .7, and .9 RT quantiles are first computed for each participant in each condition and then averaged to generate estimates of the group RT distribution (also referred to as “vincentizing”; Ratcliff, 1979). This gives, for each condition, a set of bins representing the number of responses of each type (hit, miss, false alarm, correct rejection) that fall within particular RT quantile ranges (0–0.1, 0.1–0.3, 0.3–0.5, 0.5–0.7, 0.7–0.9, 0.9–1). The model makes joint predictions not just of the probability of making a particular response, but also the probability that the response is made within each of the ranges, making it straightforward to calculate the multinomial likelihood assigned by the model to the vectors of observed counts (Ratcliff & Tuerlinx, 2002). Admittedly, Vincentizing on its own is prone to inconsistency (Rouder & Speckman, 2004), but combined with QML and a sufficient number of observations, the parameters obtained are still quite stable without needing to re-describe RT distributions in terms of a parametric family (Rouder, Lu, Speckman, Sun, & Jiang, 2005, although we will resort to parametric models of RT distributions when the number of observations is small or only summary statistics are available).

An important advantage to finding parameters by maximum likelihood is that the Akaike and Bayesian Information Criteria (AIC and BIC, respectively) can be used to compare different models and different parameterizations within the same model class. AIC and BIC represent “badness of fit”, such that the preferred model minimizes these values, which depend on both the negative log-likelihood achieved by the best-fitting parameters (denoted

NLL) and on the number of model parameters k . Additionally, BIC depends on the number of observations N , imposing a greater penalty on the number of free parameters as the number of observations increases:

$$AIC = 2 \times NLL + 2k$$

$$BIC = 2 \times NLL + k \log N.$$

3.1.2 Parameter search

For all the fits we report (including those of other models), we found parameters that maximized the quantile likelihood by first using a stochastic Differential Evolution search algorithm (Price, Storn, & Lampinen, 2006; Mullen, Ardia, Gil, Windover, & Cline, 2011) to find a good initial set of parameters, then used a simplex routine (Nelder & Mead, 1965) to find the maximum. Although the possibility of arriving at a local maximum can never be completely ruled out, this possibility is minimized by the initial stochastic search, which does a good job of exploring the full parameter space.

3.2 List length and study time: Nobel & Shiffrin, 2001

Nobel and Shiffrin (2001) present a series of studies of the dynamics of recognition memory using a variety of response procedures, including free response old/new recognition, cued recall, signal-to-respond, and associative recognition. Here, we specifically address single-item recognition under both free response and signal-to-respond conditions. These were the conditions that motivated the ARC model (Assessment of Retrieval Completion; Diller, Nobel, & Shiffrin, 2001), particularly the critical—and somewhat surprising—finding that although accuracy in free response differed as a function of both study time and list length,

response time distributions did not. Thus, Diller et al. (2001) postulated that instead of terminating their decisions on the basis of accumulated evidence, participants instead made a decision once a sufficient number of probe features had been sampled, regardless of the degree of match between the probe and memory. The degree of match determined the actual response made once feature sampling was terminated, but was not used as the basis for termination; thus, response time was decoupled from accuracy, in accord with the data. To put it another way, participants appeared to make no use of the *quality* of evidence when terminating their decisions, only the *quantity* of evidence.

Our proposed decision mechanism in free response is based on *both* the quality and quantity of evidence, since we assume decision boundaries collapse over time as a function of expected probe completeness. As a result, it is not clear *a priori* whether or how our model can best account for these data; perhaps decision boundaries will need to vary between conditions, but perhaps the use of collapsing boundaries is sufficient to account for the similarity in response time distributions across conditions. The model proposed by Diller et al. (2001) is only entailed if one assumes that decision criteria are constant across conditions. Given that study time and list length were manipulated between lists, rather than within list, it seems reasonable that participants could adjust their criteria in response to the experimental condition. For example, Ratcliff and Murdock (1976) and Ratcliff (1978) found that although response times increase with study time when study time is varied between lists—a result that seems counterintuitive—response times decrease with study time when study time is varied within lists, as one would expect. This result makes sense if participants maintain constant decision criteria within a list, but can vary them between lists. More recently, it has been found that even when study conditions are held constant, participants are able to adjust aspects of processing, like decision criteria,

in response to instructions or expectations at test (Starns, White, & Ratcliff, 2010; Kılıç & Öztekin, 2014).

3.2.1 The data

The data to which we fit our model comes from Nobel and Shiffrin (2001), Experiment 1. In this experiment, participants studied lists of word pairs which contained either 10 or 40 pairs. Each pair on each list were studied for either 670 ms or 2000 ms. Thus, list length and study time were varied between lists. After study, participants were tested with either cued recall test or single-item recognition; we focus here on only the recognition data (Diller et al., 2001, provide a model for cued recall). Half of the experimental sessions called for a free response while the other half used a signal-to-respond procedure of the kind already discussed. The reader is directed to the original papers for further detail.

Nobel and Shiffrin (2001) report several summary statistics of the observed RT distributions in each free response condition, but full quantile information was not available. Therefore, in order to fit accuracy and RT jointly using QML, we used the observed mean, standard deviation, and median RT's to estimate the parameters of an ex-Gaussian distribution. This distribution is the convolution of an exponential and Gaussian distributions, with density function

$$f(t) = \frac{1}{\tau\sqrt{2\pi}} \exp\left(\frac{\sigma^2}{2\tau^2} - \frac{t-\mu}{\tau}\right) \int_{-\infty}^{\frac{t-\mu}{\sigma} - \frac{\sigma}{\tau}} \exp\left(-\frac{y^2}{2}\right) dy$$

where μ and σ are the mean and standard deviation of the Gaussian component and τ is the rate of the exponential component. The ex-Gaussian distribution was chosen not for theoretical reasons, but simply because it has the same number of degrees of freedom as the data (three, the mean, s.d., and median) and has been found to accurately describe RT

distributions in memory tasks (Ratcliff & Murdock, 1976). The best-fitting ex-Gaussian parameters were found by minimizing the squared error between the observed and predicted mean, standard deviation, and median RT's; the resulting parameter values and estimated RT quantiles are given in Table 3.1. We computed the .1, .3, .5, .7, and .9 RT quantiles from the best-fitting ex-Gaussian distribution for each response type in each condition. This, combined with the accuracy data and knowledge of the number of observations in each condition, allowed us to fit our model using quantile maximum likelihood. Fitting to a derived distribution rather than directly to the observed data is not ideal, but has been found to preserve the important features of the observed data (Ratcliff, 1979).

3.2.2 Model fit

We assume that each word of each pair is stored as a separate trace in memory, such that 20 traces are stored in a 10-pair list and 80 traces are stored in a 40-pair list; for simplicity, we assume that only list traces are involved and that history traces play little to no role (this assumption receives qualitative and quantitative support in the next section). The probability of feature storage u varies as a function of study time (u_{670} and u_{2000}). We fit single values of c_S , T_R , and ρ across all conditions, and assumed that g and c_T were fixed at their “default” values of $\frac{1}{2}$ and 1, respectively. Because study conditions were varied between lists, allowing participants to adapt their behavior between conditions, we began by fitting a model that allowed parameters governing decision boundaries (A_0 , b , and N_C^*) to vary between conditions and used AIC and BIC measures to investigate the contribution of each parameter to performance difference across conditions.

As pointed out by Diller et al. (2001), it is highly impractical to find best-fitting parameters for the ARC model, so in order to have a reasonable baseline against which to compare

Table 3.1: Data to be fit from Nobel and Shiffrin (2001); p is the probability of a given response and N is the number of responses of a given type, summed across all sessions and participants. RT quantiles were computed from the best-fitting Ex-Gaussian distribution for each response in each condition; see main text for details. FA = False Alarm, CR = Correct Rejection.

Study time (ms)	Number of pairs	Response	p	N	RT Quantiles					Ex-Gaussian parameters			
					0.1	0.3	0.5	0.7	0.9	μ	σ	τ	
670	10	Hit	0.69	8280	502	623	726	861	1144	527.68	110.41	257.32	
670	10	Miss	0.31	3720	585	766	909	1079	1412	664.76	183.74	300.24	
670	10	FA	0.21	2520	616	829	983	1147	1413	798.57	247.96	205.43	
670	10	CR	0.79	9480	534	676	789	927	1202	588.93	140.57	249.07	
2000	10	Hit	0.83	9960	501	612	706	829	1088	524.68	100.79	235.32	
2000	10	Miss	0.17	2040	597	787	940	1128	1504	667.36	187.86	340.63	
2000	10	FA	0.14	1680	592	820	988	1169	1472	770.28	260.78	246.72	
2000	10	CR	0.86	10320	535	665	774	915	1209	566.63	119.83	267.38	
670	40	Hit	0.65	7800	516	630	729	861	1139	535.56	101.47	253.44	
670	40	Miss	0.35	4200	622	759	872	1014	1305	663.94	131.51	264.06	
670	40	FA	0.31	3720	594	780	924	1090	1403	692.16	196.29	278.84	
670	40	CR	0.69	8280	550	684	793	930	1208	592.91	128.85	253.09	
2000	40	Hit	0.78	9360	506	625	730	871	1171	523.13	103.93	272.88	
2000	40	Miss	0.22	2640	577	794	957	1140	1467	714.23	236.23	284.77	
2000	40	FA	0.25	3000	560	793	969	1166	1521	705.86	253.84	309.14	
2000	40	CR	0.75	9000	540	683	801	951	1261	580.59	134.71	281.41	

our model’s performance, we also fit a diffusion model to these data (Ratcliff, 1978). In the full version of the model, we allowed it a great deal of parametric freedom, with the goal of achieving the best quantitative fit. The mean and variance of the drift rate distributions were allowed to vary between targets and foils in each condition, as were boundary separation and response bias (boundary separation and response bias have a similar interpretation in our dynamic model, although the decision bounds are constant over time in the diffusion model). In most applications of the diffusion model, including ours, residual time is assumed to vary randomly from trial to trial according to a uniform distribution characterized by a mean and range (Ratcliff & Tuerlinx, 2002). We found that start-point variability was unnecessary to fit these data and so was not included (this is not surprising given that errors in this experiment are consistently slower than correct responses; Ratcliff & Rouder, 1998). The diffusion model was fit using the same QML procedure we used to fit our model, using the algorithms described by Tuerlinx (2004) to compute the joint accuracy-RT distribution functions and quantiles of the diffusion model. While other models of joint accuracy and RT (e.g., S. Brown & Heathcote, 2008) would likely have sufficed for comparison purposes, we selected a diffusion model for its demonstrated ability to fit a wide variety of data and for its resemblance to our own model, which is also a type of random walk, making it easier to align parameters with a similar function in each model.

The best-fitting parameters of our dynamic recognition model are given in Table 3.2, with the best-fitting model’s predictions in Figure 3.1. As expected, the probability of feature storage u increases with study time. As measured by both AIC and BIC, unlike the ARC model, our model needs to allow response boundaries to change between conditions in order to obtain the near-equivalence of response time distributions between the different conditions. In particular, allowing all aspects of the response boundaries to change as a

Table 3.2: Best-fitting parameters of the dynamic recognition model to the free-response data of Nobel and Shiffrin (2001), as well as comparable parameters from the best-fitting diffusion model. Conditions are labeled according to the number of studied pairs and study time. NLL = Negative Log-Likelihood, AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion.

<i>Parameter</i>	<i>Condition</i>	<i>Full model</i>	<i>A₀ restricted</i>	<i>b restricted</i>	<i>N_C[*] restricted</i>
u_{670}		0.394	0.391	0.401	0.398
u_{2000}		0.500	0.490	0.496	0.499
c_S		0.904	0.908	0.903	0.903
A_0	10-670	46.926	46.539	46.653	47.271
	10-2000	55.676		56.748	57.366
	40-670	43.638		38.149	37.444
	40-2000	39.454		43.532	44.140
b	10-670	0.536	0.536	0.537	0.536
	10-2000	0.538	0.545		0.537
	40-670	0.529	0.526		0.534
	40-2000	0.544	0.537		0.540
N_C^*	10-670	31.982	32.277	32.240	31.987
	10-2000	32.613	35.795	32.108	
	40-670	29.477	28.819	31.417	
	40-2000	34.163	31.512	32.370	
T_R		20.649	20.960	20.952	21.004
ρ		10.479	10.476	10.476	10.475
NLL		1411	1596	1575	1536
AIC		2857	3221	3177	3100
BIC		3018	3353	3310	3232

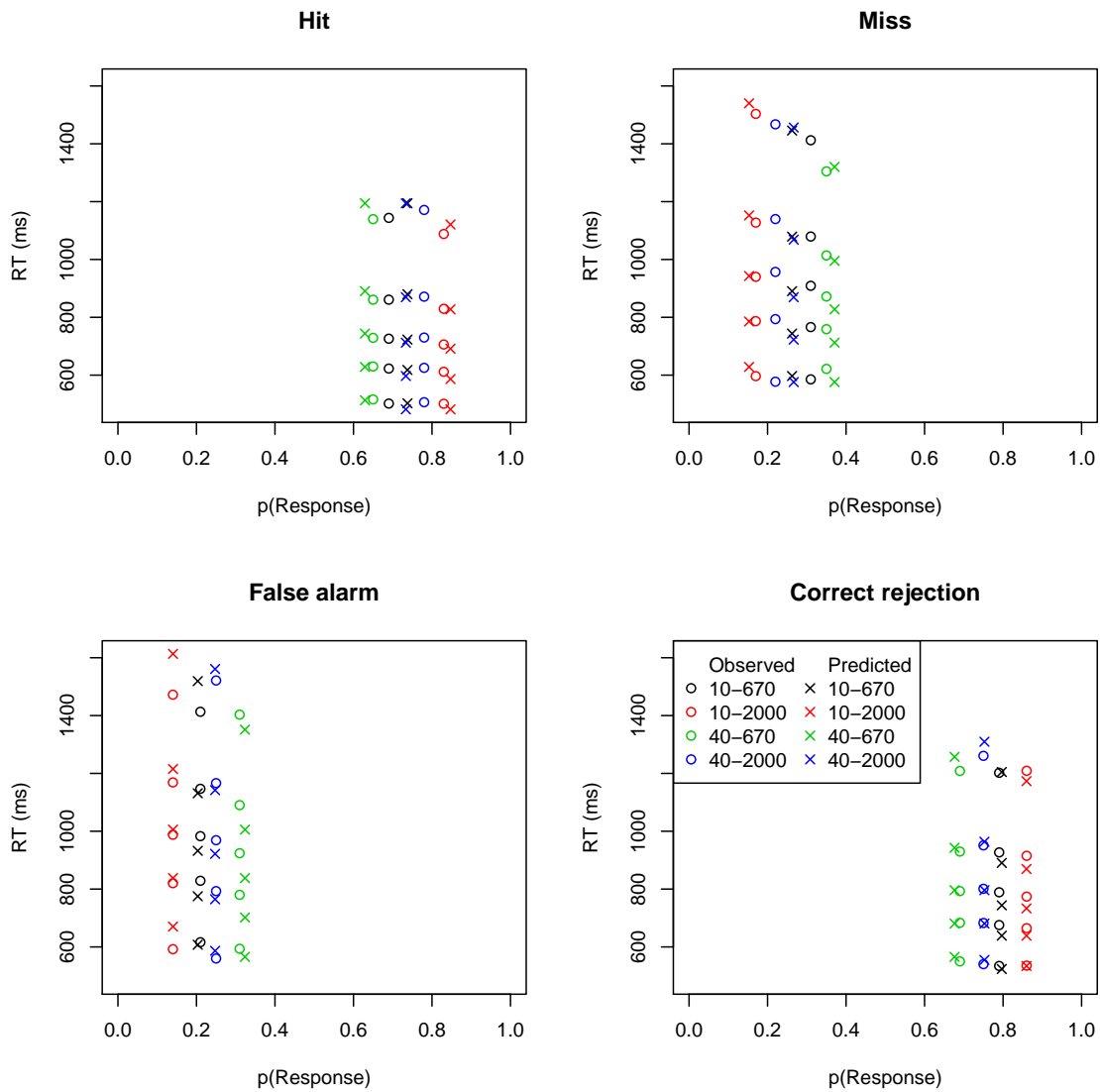


Figure 3.1: Observed response proportions and RT quantiles (0.1, 0.3, 0.5, 0.7, and 0.9) for each response type from Nobel and Shiffrin (2001), Exp. 1, along with predictions from the best-fitting dynamic recognition model. Conditions labeled according to the number of studied pairs and study time per pair.

function of condition, including initial separation, bias, and rate of collapse, produces the best fit, although fixing N_C^* produces a smaller harm than fixing A_0 or b .

The best-fitting values of various diffusion models are given in Table 3.3. Despite the larger number of free parameters, even the fully parameterized diffusion model is not able to fit these data as well as our best-fitting dynamic model. Nonetheless, the fit is still reasonably good, and just as in our model, it is clear that boundary separation needs to vary between conditions to achieve a good fit. On the other hand, the diffusion model does not require bias to vary by condition (the increase in NLL is quite minuscule) and according to BIC—which imposes a stronger penalty on the number of free parameters than AIC—the diffusion model is preferred that fixes bias and the standard deviation of the drift rate distribution across conditions. We note, however, that the diffusion model’s fit is significantly harmed if the trial-to-trial variability in residual time is removed. This result is not simply a function of using derived distributions instead of raw data, as it will be replicated in subsequent fits; thus, it suggests that the RT distributions in recognition memory are not as skewed as those predicted by the diffusion model (Ratcliff & Tuerlinex, 2002; Ratcliff & Smith, 2004). Note that our model does not assume any variability in residual time from trial-to-trial.

3.2.3 Discussion

The finding that participants adjust their response boundaries across conditions is consistent with the idea that memory evidence reaches an asymptote over time, as in our model, or is highly variable from trial to trial, as in the diffusion model. Although admittedly a post hoc explanation, if it is the case that the quality of evidence does not improve over time, such that additional accumulation would not provide additional information, there is no

Table 3.3: Best-fitting parameters and fit diagnostics of diffusion models to the free response data from Nobel and Shiffrin (2001), Exp. 1. NLL = Negative Log-Likelihood, AIC = Akaike Information Criterion, BIC = Bayesian Information Criterion.

<i>Parameter</i>	<i>Condition</i>	<i>Restricted</i>					
		<i>Full</i>	<i>Drift s.d.</i>	<i>Bound. sep.</i>	<i>Bias</i>	<i>Drift s.d. and bias</i>	<i>Resid. time var.</i>
Target mean drift	10-670	0.146	0.139	0.157	0.150	0.141	0.113
	10-2000	0.317	0.277	0.308	0.310	0.279	0.200
	40-670	0.109	0.112	0.130	0.111	0.114	0.080
	40-2000	0.209	0.215	0.210	0.214	0.214	0.145
Foil mean drift	10-670	-0.193	-0.188	-0.207	-0.192	-0.188	-0.139
	10-2000	-0.239	-0.238	-0.237	-0.248	-0.241	-0.270
	40-670	-0.119	-0.123	-0.140	-0.119	-0.124	-0.122
	40-2000	-0.168	-0.157	-0.170	-0.166	-0.160	-0.121
Target drift s.d.	10-670	0.279	0.259	0.305	0.285	0.262	0.214
	10-2000	0.296		0.290	0.297		0.238
	40-670	0.250		0.301	0.254		0.200
	40-2000	0.250		0.254	0.254		0.173
Foil drift s.d.	10-670	0.190	0.182	0.209	0.191	0.184	0.120
	10-2000	0.189		0.185	0.195		0.221
	40-670	0.173		0.225	0.178		0.169
	40-2000	0.197		0.202	0.199		0.145
Boundary separation	10-670	0.154	0.153	0.156	0.155	0.153	0.178
	10-2000	0.163	0.159		0.164	0.159	0.210
	40-670	0.145	0.148		0.146	0.148	0.180
	40-2000	0.158	0.158		0.159	0.158	0.175
Bias	10-670	0.531	0.532	0.534	0.524	0.527	0.524
	10-2000	0.511	0.526	0.513			0.570
	40-670	0.527	0.527	0.526			0.561
	40-2000	0.531	0.522	0.533			0.518
Mean residual time		0.513	0.511	0.519	0.513	0.512	0.394
Resid. time range		0.425	0.421	0.433	0.425	0.423	
NLL		1528	1549	1580	1531	1551	5721
AIC		3108	3138	3206	3107	3136	11493
BIC		3355	3327	3424	3325	3297	11729

point in spending additional time accumulating evidence for a decision after a certain point, because either no new information is available (the probe is saturated with features) or any new information is not informative (if it is too variable). The lowest boundary separations occur for long lists (40 pairs), where interference from the many studied traces reduces the separation between target and foil familiarity trajectories. The fastest rates of boundary collapse occur for short study times, when fewer features are stored and the chance that a new feature will lead to a change in familiarity is low. Both of these results suggest that participants adapt their decision criteria to the quality of the memory evidence, which had previously been shown when study time is varied between lists (Ratcliff & Murdock, 1976; Ratcliff, 1978). The design of the current experiment, in which participants experienced the various conditions many times over many sessions, would certainly afford ample time for such adaptation to occur. Why it would occur in such a way as to produce roughly equivalent RT distributions between conditions is not clear, and the chief virtue of the ARC model (Diller et al., 2001) is that it provides an *a priori* reason for such a finding. Our model does not entail an equivalence in RT distributions between conditions, but like the diffusion model, is capable of fitting a much broader range of findings than the ARC model (Nosofsky & Stanton, 2006). We can only say that the good fit of our model to the Nobel and Shiffrin (2001) data means that it is possible to explain their results without recourse to a model built especially to fit their data.

Given that both our model and the diffusion model assume that evidence accumulation for recognition is a random walk (albeit with very different properties for each model), it is informative to compare parameters between the two models to see if they have the same interpretations. Because evidence is nonstationary in our model, we only examine its asymptotic form (i.e., as time $t \rightarrow \infty$ or, equivalently, when all probe features have been

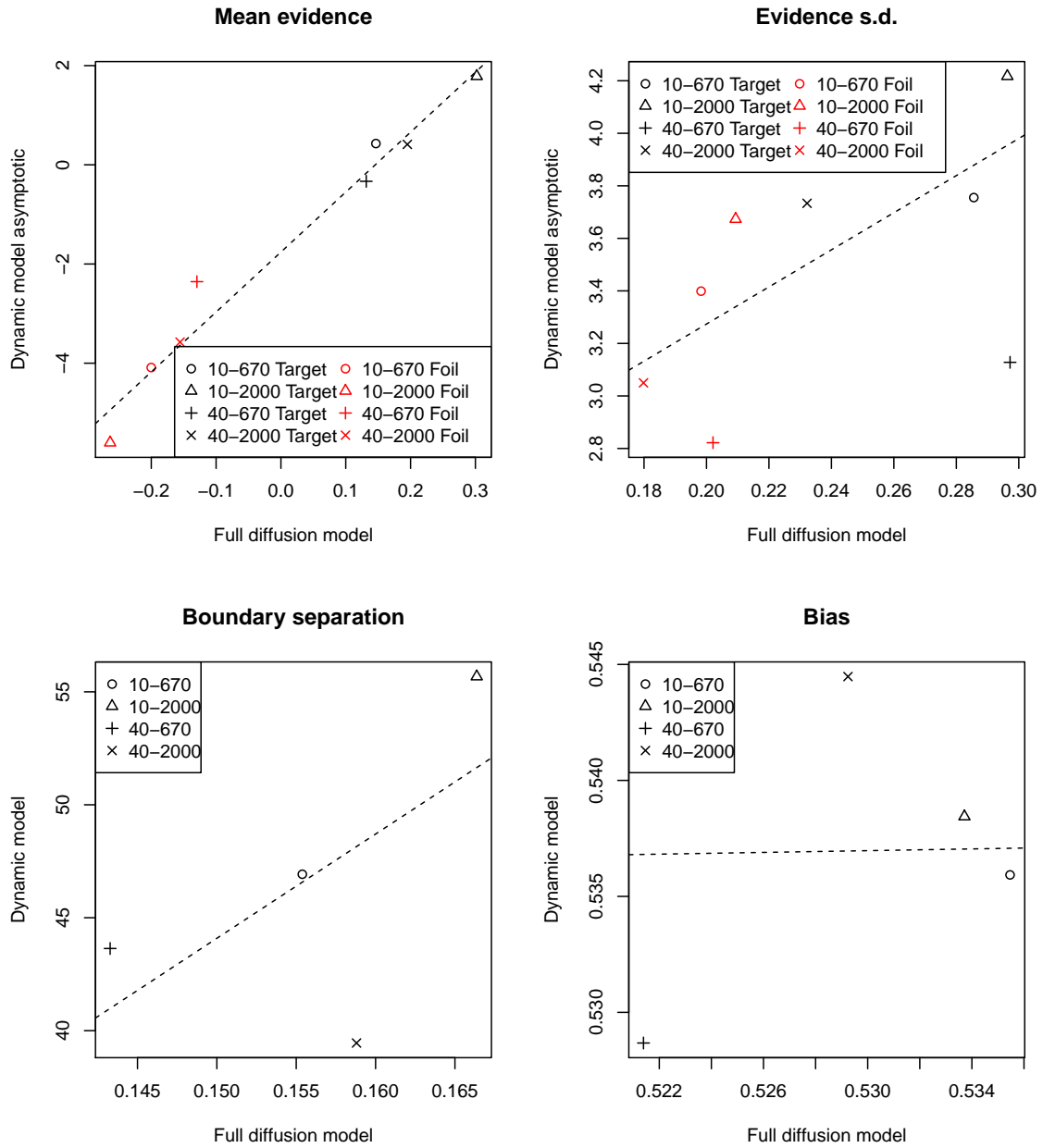


Figure 3.2: Comparison of asymptotic familiarity distributions produced by our model and parameter estimates for the diffusion model fitted to the free response data of Nobel and Shiffrin (2001), Exp. 1. Dashed lines depict the least-squares regression line.

sampled). In Figure 3.2, we plot the mean and standard deviations of the asymptotic distributions of $x(t)$ against the mean and standard deviation of the drift rate distributions in the full diffusion model. Both the means and standard deviations are correlated—particularly the means—indicating at the least that “memory evidence” would seem to have a similar interpretation within both models, although our model provides a set of principles from which to derive these values rather than letting them be free parameters. In contrast, decision parameters like boundary separation and response bias (also shown in Figure 3.2) are only weakly related between the two models, but given the vastly different character of the decision bounds in the two models, this is not surprising.

A final comparison between the two models is in their treatment of residual time (T_R): In our model fits, this is estimated to be quite low (around 20 ms) while in the diffusion model it is quite high (roughly half a second). In addition, our model assumes residual time is a constant, whereas the diffusion model requires substantial trial-to-trial variability in T_R to fit the data (see Table 3.3). While our assumption of constant residual time is undoubtedly over-simplistic, the facts that it is shorter in our model and does not need to vary to achieve a similar quantitative fit suggest that the processes of feature sampling and probe/trace comparison posited by our model account for a large amount of the variability in response time, more so than the linear evidence accumulation process posited by the diffusion model. Indeed, “residual time” in applications of the diffusion model is typically assumed to include time for stimulus encoding (Ratcliff, 1978), a process made explicit in our model. Finally, with regard to the very brief residual time estimates for our model, we note that although residual time is also presumed to include components related to stimulus detection and response execution, such processes need not occur serially (see, e.g., Sternberg, 1969; Townsend & Ashby, 1983). They may well occur in cascaded fashion

(e.g., McClelland, 1979; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001), with various stages of stimulus detection and response execution “leaking” into the retrieval process.

It is clear that our dynamic model for recognition provides a good quantitative account of these data, on par with the diffusion model. Our model has fewer free parameters and offers a mechanistic account of the recognition process in terms of degree of feature storage (u), encoding noise (c_S), and interference from other list items. Thus, in addition to fitting data, our model offers additional insight into the recognition processes that generate the data.

3.3 Word Frequency

In this section, we continue to demonstrate that our proposed dynamic model for recognition does a good job of fitting both accuracy and response time distributions and that, in doing so, it provides insights that would not be available with either a static memory model or a generic decision model. Although various issues will be discussed, our primary focus is on how word frequency manifests in recognition. Recognition performance is better for low frequency (LF) words than for high frequency (HF) words, with this advantage typically taking the form of a “mirror effect”: not only are hit rates higher for LF words, but false alarm rates are lower (Glanzer & Adams, 1985). Such a finding is counter-intuitive from a simple familiarity-based perspective on recognition: If LF items simply possessed less memory strength or lower familiarity, this would predict a lower false alarm rate, but not a lower hit rate. For this reason, the word frequency mirror effect has led to a great deal of theoretical and empirical investigation, although it remains difficult to decide between the various theories proposed. We show that, by taking a dynamic approach, it is possible to rule out a certain class of theories that would be otherwise indistinguishable.

Before we can make this argument, we must lay out the various explanations provided for the word frequency mirror effect. One type of explanation is based on the idea that participants are aware of how often an item occurs and in what situations, and are able to use this information to make a “meta-memory” judgment during a recognition test. For instance, participants do not commit false alarms to their own names, despite their obvious familiarity, suggesting that, in some cases, participants know if they *would have* encoded their name had it been studied (J. Brown, Lewis, & Monk, 1977). The notion that recognition judgments in general depend on meta-memory was codified in Attention Likelihood Theory (ALT; Glanzer & Adams, 1990), which posits that studying a word leads to certain features of the word becoming “marked” (in the language of our theory, we would say that these features are “stored”). Then, when an item is presented at test, a participant compares how many “marked” features the word has to how many it would have normally by chance; if the number of marked features exceeds this amount, a participant calls the word “old”, operating under the assumption that studying the item is what increased the number of marked features. A word frequency mirror effect obtains because participants know that they would have attended to and marked the features of a LF word more so than a HF word. Thus, when a studied LF word is tested, it has many marked features as would be expected if it has been studied; if a novel LF word is tested, it has many fewer marked features than it would have had if it had been studied.

An acknowledged drawback of ALT is that it is not clear how participants form these meta-memory expectations from one trial to the next. A participant must first decide what class the item belongs to (HF or LF) such that the correct expectations can be brought to bear. A different account that nonetheless implicates attention as the source of the LF advantage is the early-phase elevated attention hypothesis (Malmberg & Nelson, 2003).

This hypothesis holds that LF words present a kind of “desirable difficulty”: because LF words require more attentional resources to initially identify (Becker, 1976), this greater investment of effort enables more features of the word to be detected and encoded. However, although participants do in fact look longer at LF words at study—and looking time is predictive of subsequent memory performance—there is no interaction between looking time and word frequency (Pazzaglia, Staub, & Rotello, 2014). That is, if LF words are advantaged only to the extent that they attract or require more attention, as the early-phase elevated attention hypothesis proposes, then this advantage should be reduced with reduced looking time. That this does not occur suggests that, although LF items do indeed attract more attention, this alone is not the source of their mnemonic advantage.

A second group of explanations for the word frequency effect holds that memory differences by word frequency depend not on attention or meta-memory judgments, but on how the words are encoded in memory. This group may be divided into dual- and single-process accounts, although we will argue later that this division may obscure a deeper underlying similarity between these two types of account. Dual coding theory (Reder et al., 2000) is based on the idea that there are two routes to making a recognition judgment: one based purely on a continuous value of memory strength and another based on explicit recollection of a memory trace. Memory strength is presumed to increase with frequency, such that a high frequency item will have high memory strength and a low frequency item will have low memory strength. However, this is counterbalanced by a “fan” effect (e.g., Anderson, 1983) whereby the large number of memory traces stored for high frequency items compete with one another making any one of them difficult to retrieve. The LF advantage results, then, from the joint operation of these two factors on the two processes presumed to underlie recognition: LF foils have lower memory strength, leading to a lower false alarm

rate; LF targets, however, have fewer memory traces that compete to be retrieved and so are more likely to result in a hit via the recollection of the target trace.

Single-process theories provide a different explanation for word frequency effects that nonetheless relies on encoding differences. In particular, such theories posit that the representations of low frequency words in memory are more distinctive (Schmidt, 1991; Nosofsky & Zaki, 2003). According to the REM model (Shiffrin & Steyvers, 1997, 1998), this distinctiveness manifests in the features used to encode the words themselves. Specifically, low frequency words are comprised of both semantic (Adelman, Brown, & Quesada, 2006) and orthographic (Malmberg, Steyvers, Stevens, & Shiffrin, 2002) features that are more rare. In our model, then, studying a LF word would result in a trace with features that have a lower g parameter than HF words, and a LF test word would produce a memory probe with a lower g value for its features. If g is lower for probe features, there will be fewer traces in memory that happen to share the probe's value on that feature by chance, thus lowering the false alarm rate. Contrariwise, if there is a target trace in memory, a match when g is low provides stronger evidence that the probe and trace encode the same event, thus increasing the likelihood ratio for the target trace (termed "self-match" by Nosofsky & Zaki, 2003) and thereby increasing the hit rate.

Low frequency items not only occur less often, they also tend to occur in fewer different contexts (Adelman et al., 2006). Thus, the word frequency effect could result from context noise (Dennis & Humphreys, 2001), since LF words are less likely than HF words to have occurred in a context that is confusable with the study context. How this context noise operates depends on whether a separate- or composite-trace view is adopted. BCDMEM takes a composite-trace approach, such that a word is associated with a particular set of context features reflecting that word's pattern of use; when the word is studied, it becomes

associated with some additional features of the study context, and the degree of match between the features of the study context and a word's associated context features determines whether or not it will be judged "old". Because common words tend to be associated with more context features, it is more difficult to tell whether a feature of the study context is associated with that word because of studying it, or simply by chance, leading to a lower hit rate and higher false alarm rate for HF words. An analogous separate-trace implementation of context noise is to assume that memory contains more traces of HF words from prior life history, represented by parameter K in our model. Although history traces match in content features, they mismatch in context. Just as adding more list traces—which match in context, but not content—leads to a mirror effect, adding more history traces will do the same by increasing the odds that one of those traces was formed in a context similar to the experimental study context (higher false alarms to HF words) and introducing many poor matches that swamp the influence of a well-matching target trace (lower hit rate to HF words; Nelson & Shiffrin, 2013). Regardless of implementation, the contexts in which a word appears in prior life history affect episodic recognition: When normative word frequency is held constant but contextual diversity is varied, words that occur in more diverse contexts have a lower hit rate and higher false alarm rate than words that only occur in a few contexts (Steyvers & Malmberg, 2003).

Each of these various theories accounts for the effects of word frequency on recognition accuracy, and there is little basis for deciding between them on the basis of accuracy alone. Many of these theories do make predictions about the *dynamics* of the word frequency effect, at least at a qualitative level. If the word frequency effect depends on first determining the frequency class to which the word belongs such that an appropriate criterion can be set (e.g., Gillund & Shiffrin, 1984; Glanzer & Adams, 1990; Hintzman, 1994), then the

canonical mirror effect should only obtain when enough time is allowed for criterion setting. Similarly, if the LF hit rate advantage depends on recalling a LF target trace (Reder et al., 2000), this should take additional time to manifest. If LF words attract more attentional resources by virtue of more difficult initial identification (Malmberg & Nelson, 2003), there should be an overall delay in processing LF words during that initial period, but after that the mirror effect should appear. If word frequency effects depend on interference from specific prior episodes, this should only be apparent later in processing, when enough content features are sampled to overcome the initial mismatching context features. On the other hand, if frequency effects depend on feature distinctiveness (Shiffrin & Steyvers, 1997) or retrieval of associated context features (Dennis & Humphreys, 2001), then the frequency effect will persist throughout the trial since those features—whether content or associated context—are available from the very beginning.

These dynamic predictions can be tested using both a response signal paradigm, and free response under both speed and accuracy emphasis. Hintzman et al. (1994) conducted a response signal study of recognition of both HF and LF words, and we will examine this first, since it most clearly tests the dynamic predictions above. We also examine two free response studies, one by Rae et al. (2014) and another by Starns et al. (2012), in which both word frequency and response emphasis, speed versus accuracy, was varied. Starns et al. (2012) also varied study time and response bias, providing further opportunities to investigate the properties of our dynamic model. Across all these experiments, we find that the word frequency mirror effect persists across response signals at various times and as instructional focus shifts from speed to accuracy emphasis, arguing against any account of word frequency that predicts a difference in the effect as a function of time.

3.3.1 Signal to respond

Hintzman et al. (1994) studied the source of the word frequency effect using both divided attention and response signal procedures, but we focus on the response signal procedure used in their Experiment 6. On each study-test block, participants studied a list of 20 words, 5 of which were low frequency, 5 were high frequency, and 10 were of moderate frequency (these were used to test a different mirror effect—between concrete and abstract words—but will not be discussed here). At test, all of these words were tested along with 20 foils. To fit these data, we minimized the sum of squared error between predicted and observed probability of saying “old” at each signal lag, and assessed model fit via adjusted R^2 ,

$$R_{\text{Adj}}^2 = 1 - \frac{\frac{\sum_{i=1}^N (\hat{d}_i - d_i)^2}{N-k}}{\frac{\sum_{i=1}^N (\bar{d} - d_i)^2}{N-1}},$$

where \hat{d}_i is the predicted probability of responding “old” at point i , \bar{d} is the mean probability of saying “old” across lags and item types, N is the number of observed points, and k is the number of free parameters. Thus, R_{Adj}^2 implements a trade-off between model fit and complexity.

To model frequency effects arising from interference from prior life history, we allowed the number of history traces to differ between LF and HF words (parameters K_L and K_H , respectively). These traces match the test word in content features, but have random context features, assuming that most life experience bares little similarity to the experimental situation. We compared this account of word frequency effects with a feature distinctiveness account, which was embodied by a single parameter g_L , reflecting the probability that a content feature of a LF word would match a probe feature by chance; we simply let this value for HF words be $g_H = 1 - g_L$. Historical traces will not be activated until sufficient

Table 3.4: Best-fitting parameters of models representing different account of the word frequency effect in the Hintzman et al. (1994, Exp. 6), along with fit diagnostic R^2_{Adj} . Values fixed for each model are marked with an asterisk.

Parameter	Word frequency source		
	Feature distinctiveness	Prior history traces	Both
u	0.371	0.373	0.371
c_S	0.982	0.986	0.982
K_L	0*	668	0
K_H	0*	2886	0
g_L	0.484	0.5*	0.484
θ	0.037	0.062	0.037
t_0	121.809	94.305	121.809
ρ	6.686	7.718	6.686
R^2_{Adj}	0.958	0.914	0.950

content features have been sampled to allow them to be active, overcoming their initial mismatching context features, leading to the prediction that differences between frequency classes will not arise until later response signals. Feature distinctiveness, on the other hand, predicts frequency differences at all lags, although the size of the effect will grow as more content features are sampled. We fit three versions of our dynamic model, one under the assumption that only feature distinctiveness leads to frequency effects, another that assumes only interference from history traces leads to frequency effect, and a model that allows both mechanisms to operate. Fitted parameter values are given in Table 3.4.

The preferred model is that in which frequency effects are due only to feature distinctiveness; when both that and history traces are allowed to vary, the data are best fit by assuming $K_L = K_H = 0$. The reason is clear by examining the data, which are shown along with predictions of the feature distinctiveness model in Figure 3.3: A word frequency mirror effect is, indeed, present across signal lags, while a difference in history traces predicts that such a difference would not appear until later. Beyond the fact that it is unwise to draw strong conclusions from a single study, we note that it is also possible that the unusual demands of the response signal procedure entail a different kind of processing from that normally encountered in free response settings. We thus turn to two large-scale studies of free response recognition.

3.3.2 Rae et al. 2014

Rae et al. (2014) report results from a recognition memory experiment with 47 participants and a total of 18,001 trials; participants studied lists containing 56 words, of which 50 appeared as targets on a subsequent test along with 50 foils. For each study/test block, participants were instructed to emphasize either accuracy or speed in their responses. Al-

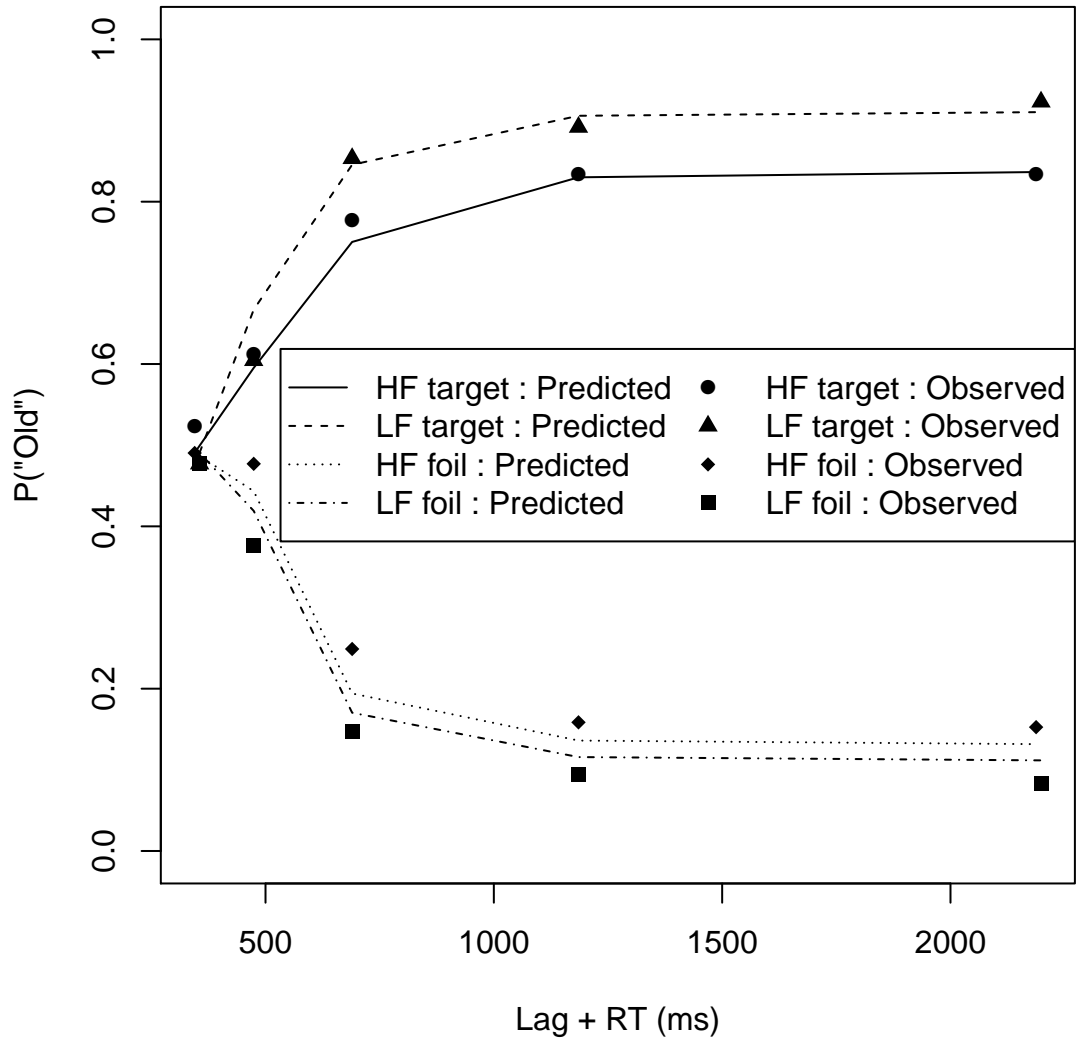


Figure 3.3: Observed probability of responding “old” as a function of signal lag and response time in Hintzman et al. (1994, Exp. 6), with predictions from the dynamic model that assumes feature distinctiveness underlies differences between frequency classes.

though Rae et al. (2014) did not analyze and report effects of normative word frequency, we divided their stimuli into LF and HF words based on a median split in word frequency (using the MRC database; Coltheart, 1981). For further details of the experimental procedure, the reader is directed to their paper. In this case, we were able to find parameter estimates via QML, enabling more robust quantitative model comparisons using both AIC and BIC.

Frequency

Once again, we fit three versions of the model each embodying different assumptions about the source of frequency effects: One that assumes only differences in the number of history traces (model 1), one that assumes only differences in feature distinctiveness (model 2), and another that allows both mechanisms (model 3). As evidenced by both AIC and BIC (Table 3.5, rows 1–3), and by the optimization routine hitting the minimum K values of 0 for both HF and LF items, the model (#2) is preferred where history traces play no role ($K_L = K_H = 0$) and frequency effects are accounted for solely in terms of different base-rates between HF and LF items ($g_L < g_H$). Once again, this quantitative result is anticipated by the qualitative finding that frequency effects persist under both speed instructions (where responses are made quickly) and accuracy instructions (where responses are made slowly), as shown in Figure 3.5. If the source of frequency effects was time-dependent, as would be predicted by the historical interference account, then frequency effects should be mitigated under speed instructions, but this is not the case.

Table 3.5: Best-fitting parameter values and fit criteria for the data from Rae et al. (2014). Values fixed for each model are marked with an asterisk. Models 1-3 only vary frequency-related parameters while models 4-8 vary parameters related to speed/accuracy trade-off; the overall preferred model is indicated in boldface.

Parameter	Condition	Model							
		1	2	3	4	5	6	7	8
u		0.393	0.394	0.394	0.407	0.407	0.412	0.396	0.408
c_S		0.932	0.931	0.931	0.938	0.935	0.938	0.934	0.939
K_L		0	0*	0	0*	0*	0*	0*	0*
K_H		0	0*	0	0*	0*	0*	0*	0*
g_L		0.5*	0.479	0.479	0.478	0.474	0.474	0.473	0.473
A_0	Accuracy	46.6	48.3	48.3	45.4	45.1	42.5	45.5	44.5
	Speed	22.6	21.9	21.9	22.5	26.3	28.3	28.3	30.2
b	Accuracy	0.541	0.54	0.54	0.541	0.543	0.545	0.542	0.543
	Speed	0.573	0.572	0.572	0.579	0.56	0.559	0.557	0.555
T_R	Accuracy	315	320	320	323	323	336	323	335
	Speed	327	331	331	319	304	292	298	285
ρ	Accuracy	6.27	6.1	6.1	6.38	6.38	6.37	6.38	6.39
	Speed	4.17	4.17	4.17	4.17	5.62	5.62	5.62	5.83
N_C^*	Accuracy	30*	30*	30*	30*	30*	30*	29.2	29.1
	Speed	30*	30*	30*	30*	22.6	21.9	22.1	21.3
c_T	Speed	1*	1*	1*	0.961	1*	0.971	1*	0.975
NLL		830	779	779	749	704	678	693	669
AIC		1685	1579	1583	1522	1432	1381	1413	1366
BIC		1778	1665	1685	1616	1526	1483	1514	1475

Speed-accuracy trade-offs

Rae et al. (2014) assessed the possibility that instructions to focus on speed, rather than accuracy, affect not just response boundaries, but decision evidence as well. Starns et al. (2012) argue that a focus on response speed may impair participants' ability to form effective retrieval cues. In our model, this concept can be embodied in two ways. Speed instructions may cause sampling of test item features to be less accurate, represented by a value of $c_T < 1$. Different instructions may also lead to a more rapid collapse of response boundaries, represented as a lower value of N_C^* in Equation 2.10. If boundaries collapse faster, decisions will tend to be made on the basis of fewer features, before the evidence has matured.

As shown in rows 4–8 of Table 3.5 the preferred model (#8) allows both c_T and N^* to vary. This is in accord with Rae et al. (2014) and Starns et al. (2012): participants under speed instructions make decisions both with lower *quality* evidence—reflected in decreased boundary separation and increased sampling noise c_T —and lower *quantity* of evidence—reflected in faster boundary collapse. This situation is depicted graphically in Figure 3.4.

Quality of fit

As shown in Figure 3.5, the best-fitting version of our dynamic model (#8) does a good job accounting for both accuracy and RT quantiles. For comparison, we again looked to the diffusion model (Ratcliff, 1978) as a reasonable baseline. According to AIC, the preferred diffusion model was one that allowed both the mean drift rate for each item type (LF and HF targets and foils) and the standard deviation of the drift rate distributions to vary as a function of speed/accuracy focus. In addition, boundary separation, bias, mean

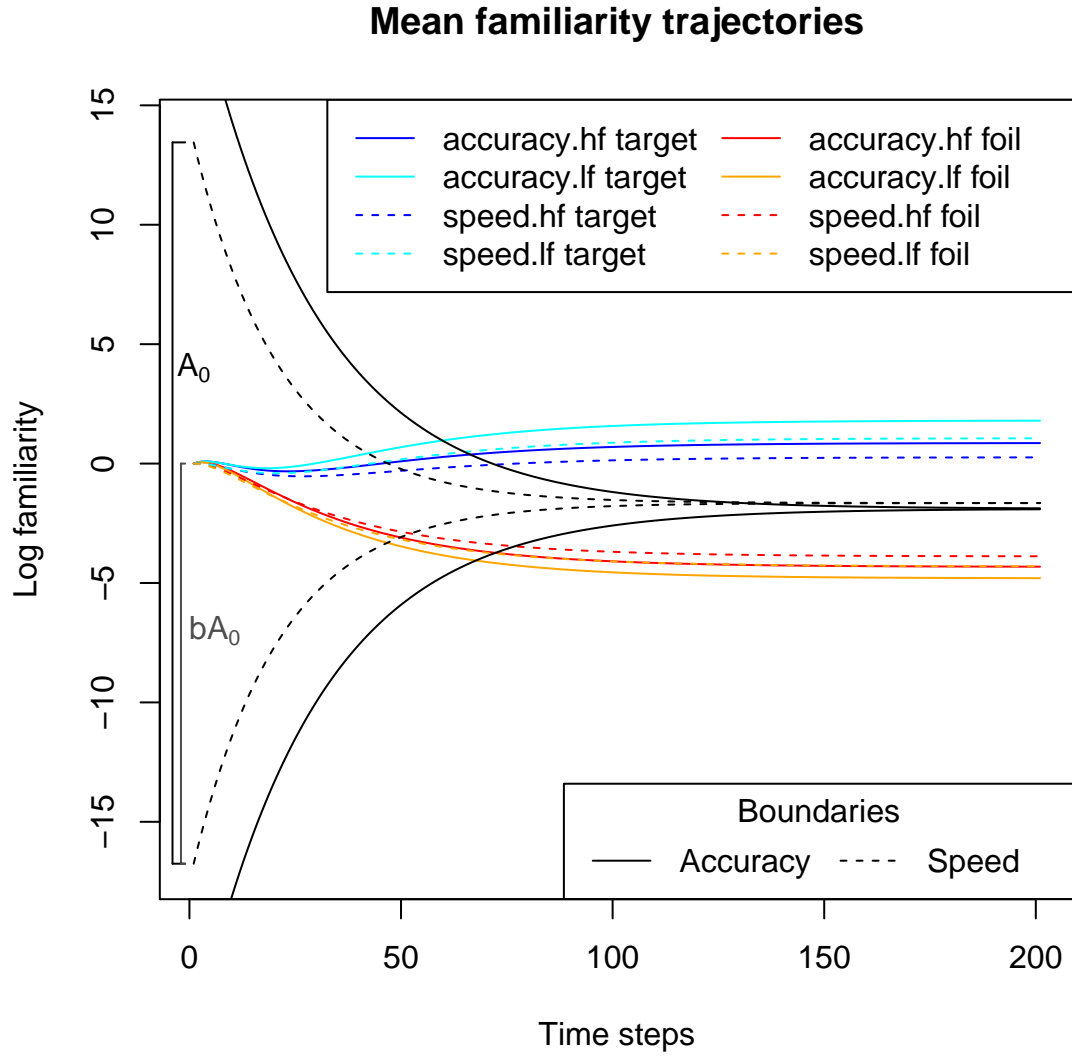


Figure 3.4: Mean familiarity trajectories and response boundaries for the preferred model fit (#8) to the data from Rae et al. (2014). The initial boundary separation (A_0) and bias (b) parameters are labeled for reference.

Table 3.6: Best-fitting parameter values and fit criteria of diffusion models fit to the data from Rae et al. (2014).

Parameter	Condition	Best AIC	Best BIC	No residual time variability
Target mean drift	HF-Accuracy	0.260	0.186	0.154
	HF-Speed	0.206	—	0.151
	LF-Accuracy	0.357	0.260	0.207
	LF-Speed	0.293	—	0.195
Foil mean drift	HF-Accuracy	-0.234	-0.166	-0.163
	HF-Speed	-0.164	—	-0.106
	LF-Accuracy	-0.300	-0.222	-0.205
	LF-Speed	-0.225	—	-0.153
Target drift s.d.	HF-Accuracy	0.325	0.217	0.166
	HF-Speed	0.270	—	0.154
	LF-Accuracy	0.347	0.222	0.148
	LF-Speed	0.284	—	0.113
Foil drift s.d.	HF-Accuracy	0.242	0.140	0.121
	HF-Speed	0.209	—	0.102
	LF-Accuracy	0.241	0.143	0.121
	LF-Speed	0.167	—	0.0811
Boundary separation	Accuracy	0.121	0.111	0.123
	Speed	0.0738	0.0691	0.0854
Bias	Accuracy	0.546	0.541	0.551
	Speed	0.528	0.539	0.510
Mean residual time	Accuracy	0.577	0.562	0.523
	Speed	0.455	0.452	0.410
Residual time range	Accuracy	0.206	0.191	—
	Speed	0.161	0.171	—
Start-point variability		0.0386	—	0.0326
NLL		505	531	1159
AIC		1061	1094	2364
BIC		1256	1219	2543

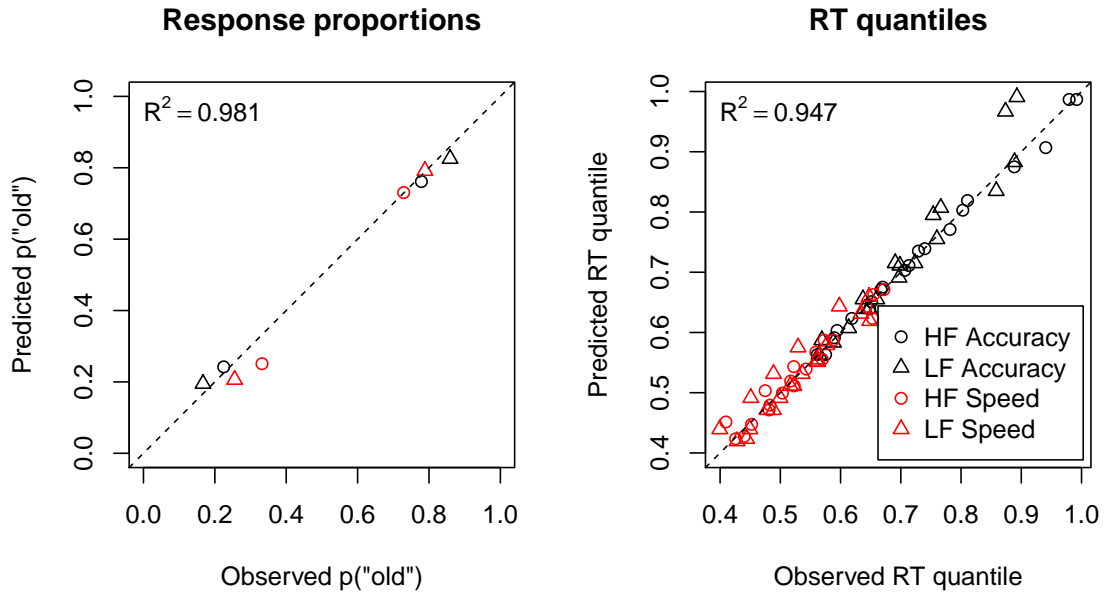


Figure 3.5: Diagnostic plots for the preferred model fit (#8) to the data from Rae et al. (2014), with proportion of variance accounted for (R^2) indicated in the upper left corner.

residual time, and residual time variability were allowed to vary between speed and accuracy instructions, although only a single start-point variability parameter was needed. The preferred diffusion model according to AIC, then, had 25 free parameters, achieving an AIC of 1061, considerably better than the best version of our dynamic model (#8, which had an AIC of 1366). According to BIC, which imposes a stronger penalty on free parameters, the preferred diffusion model fixed the mean and standard deviation of the drift rate distributions across instructions and had no start-point variability at all; this model allowed boundary separation, bias, mean residual time, and residual time variability to vary as a function of instructions. The BIC-preferred diffusion model had only 16 free parameters and achieved a BIC of 1219, again lower than our best model's BIC (1475).

For the most part, the best-fitting diffusion model parameters (Table 3.6) are sensible

and in accord with the explanation offered by our dynamic model, but offer a somewhat confusing picture of the effect of speed vs. accuracy instructions, in that BIC argues against an effect of instructions on mnemonic evidence. This could be an artifact of fitting to group data rather than individuals, as the full model (the one selected by AIC here) is preferred according to both AIC and BIC in individual fits (Rae et al., 2014). Still, it is interesting that not only mean residual time, but the range of the residual time distribution is still selected to vary according to BIC. This, combined with the fact that residual time variability is necessary for the diffusion model to fit these data at all (see the final column of Table 3.6), suggests that important decision-relevant processes are occurring during this “residual time” that are not accounted for by linear accumulation of noisy evidence. Such processes could include feature sampling, as posited by our model, reinforcing the importance using a dynamic model to dig deeper into the mechanisms underlying recognition.

3.3.3 Starns et al. 2012

In the experiment reported by Starns et al. (2012), four participants each completed 20 sessions of a recognition memory task for a total of 66,986 trials. Within each study/test block, study words were shown 1, 2, or 4 times and words at both study and test could be of either high or low frequency. Between blocks, participants were instructed to focus on either speed or accuracy. Unlike the Rae et al. (2014) study, test lists also varied in the proportion of targets, either .21, .32, .5, .68, or .79, thus inducing a bias (participants were informed about this proportion prior to testing). The reader is directed to their paper for further details of their experimental procedure.

We allowed u (the probability of encoding a content feature at study) to vary with number of repetitions; across all model fits (Table 3.7), u increases with the number of rep-

Table 3.7: Best-fitting parameter values and fit criteria for the data from Starns et al. (2012). Values fixed for each model are marked with an asterisk. Models 1-3 only vary frequency-related parameters while models 4-8 vary parameters related to speed/accuracy trade-off; the overall preferred model is indicated in boldface.

Parameter	Emphasis	Target proportion	Model							
			1	2	3	4	5	6	7	8
u_1			0.299	0.272	0.272	0.275	0.268	0.275	0.268	0.276
u_2			0.388	0.360	0.360	0.364	0.383	0.364	0.383	0.364
u_4			0.434	0.408	0.408	0.423	0.443	0.423	0.443	0.423
c_S			0.937	0.944	0.944	0.954	0.945	0.954	0.945	0.954
K_L			0	0*	0	0*	0*	0*	0*	0*
K_H			0	0*	0	0*	0*	0*	0*	0*
g_L			0.5*	0.472	0.472	0.473	0.473	0.477	0.473	0.481
A_0	Accuracy	0.21	16.2	15.4	15.4	15.1	15.1	15.5	15.1	16.4
		0.32	14.7	15.6	15.6	15.7	14.8	15.2	14.8	15.6
		0.50	15.1	15.8	15.8	15.4	15.4	15.4	15.4	15.8
		0.68	13.8	14.2	14.2	15.0	14.5	14.9	14.5	14.7
		0.79	16.0	15.5	15.5	15.8	15.6	15.7	15.6	16.9
	Speed	0.21	16.1	14.9	14.9	14.0	18.7	14.2	18.7	14.6
		0.32	12.6	12.9	12.9	12.1	15.5	12.0	15.5	12.7
		0.50	14.3	13.5	13.5	13.1	17.8	12.6	17.8	13.0
		0.68	13.7	13.2	13.2	12.1	16.4	12.1	16.4	13.1
		0.79	12.7	12.6	12.6	11.3	16.8	11.7	16.8	12.1
b	Accuracy	0.21	0.446	0.429	0.429	0.42	0.434	0.423	0.434	0.421
		0.32	0.498	0.488	0.488	0.48	0.489	0.474	0.488	0.476
		0.50	0.575	0.558	0.558	0.554	0.566	0.554	0.566	0.552
		0.68	0.67	0.655	0.655	0.649	0.663	0.647	0.663	0.645
		0.79	0.718	0.711	0.711	0.705	0.716	0.7	0.716	0.692
	Speed	0.21	0.494	0.471	0.471	0.507	0.474	0.503	0.475	0.5
		0.32	0.537	0.517	0.517	0.557	0.512	0.549	0.512	0.547
		0.50	0.566	0.552	0.552	0.582	0.539	0.582	0.539	0.576
		0.68	0.602	0.588	0.588	0.619	0.567	0.616	0.567	0.611
		0.79	0.669	0.654	0.654	0.697	0.613	0.687	0.613	0.68
T_R	Accuracy	0.21	321	329	329	338	333	333	333	329
		0.32	332	326	326	335	339	339	339	335
		0.50	334	331	331	344	340	344	340	340
		0.68	360	357	357	363	363	363	363	362
		0.79	346	354	354	358	355	358	355	349
	Speed	0.21	268	277	277	270	227	270	226	268
		0.32	299	296	296	288	252	292	252	287
		0.50	289	294	294	282	239	290	239	289
		0.68	293	294	294	287	243	291	243	285
		0.79	308	307	307	306	246	303	246	302
ρ	Accuracy		4.68	4.59	4.59	4.59	4.59	4.59	4.59	4.60
	Speed		3.99	3.96	3.96	3.96	6.47	3.96	6.47	3.98
N_C^*	Accuracy		30*	30*	30*	30*	30*	30*	30.0	30.0
	Speed		30*	30*	30*	30*	20.0	30*	20.0	30.0
c_T	Speed		1*	1*	1*	0.955	1*	0.962	1*	0.964
NLL			6095	5972	5972	5998	5796	5976	5795	5937
AIC			12265	12019	12023	12071	11667	12029	11669	11955
BIC			12611	12356	12378	12417	12014	12385	12024	12319

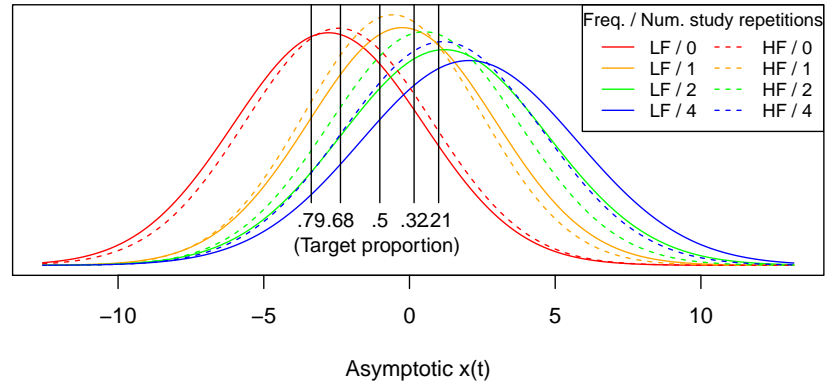


Figure 3.6: Asymptotic familiarity distributions and decision bounds (black) produced by model 7 of the data from Starns et al. (2012).

etitions. In all model fits, b increases with the proportion of targets, as in Figure 3.6 which shows the asymptotic boundaries ($A_0(\frac{1}{2} - b)$) for model 5. As above, the best account for word frequency effects assumes that HF/LF words differ in feature-frequency base rates rather than the number of episodic history traces (rows 1-3 of Table 3.7). Finally, speed instructions are found to result both in a smaller initial boundary separation and a faster rate of boundary collapse, but allowing for additional noise in probe feature sampling ($c_T < 1$) does not improve the model fit in this case. These results are consistent with Rae et al. (2014) in that speed instructions result in fewer probe features being sampled before a decision is made, but in this case, there is no evidence that the perception of those features is impaired.

Target/foil variance

One of the main arguments in Starns et al. (2012), expanded in Starns and Ratcliff (2014), is that the distribution of target evidence has greater variance than that for foils. Although

the evidence distributions in our model are non-stationary, the asymptotic familiarity distributions predicted by the AIC- and BIC-preferred model (#5; Figure 3.6) show greater target than foil variance in all but one case: HF targets that appear only once at study. In that case, the very low u value means that the single matching target trace cannot take many values even when the probe is saturated, thereby decreasing the variance of the familiarity distribution when this single target is averaged with the likelihoods of the many non-matching list traces (which have higher variance as their feature values bear only a chance resemblance to the target). Given the small size of this effect, we do not make a strong theoretical claim about it, but note that the ratio of target to foil variance is often estimated to be lower for weak HF words (Starns et al., 2012; Starns & Ratcliff, 2014).

Quality of fit

Once again, as shown in Figure 3.7, our model does a good job at accounting for both observed accuracy and RT quantiles, although it is not as good at predicting the longest (.9) quantiles. By comparison, the best-fitting diffusion model reported by Starns et al. (2012) achieved a BIC of 7886, considerably lower than our best BIC of 12014 (Table 3.7, albeit with 66 free parameters compared to the 38 free parameters in our preferred model (#5). As another point of model comparison, Starns et al. (2012) fit the RTCON model (Ratcliff & Starns, 2009) to these data; RTCON uses a bank of nonlinear leaky competing accumulators (similar to Usher & McClelland, 2001) rather than the perfect linear accumulator assumed in the diffusion model, and has been shown to fit joint confidence-RT data quite well (Ratcliff & Starns, 2009, 2013), so it stands to reason that it should be able to fit binary choice data as well. Nonetheless, they report that RTCON was only able to achieve a

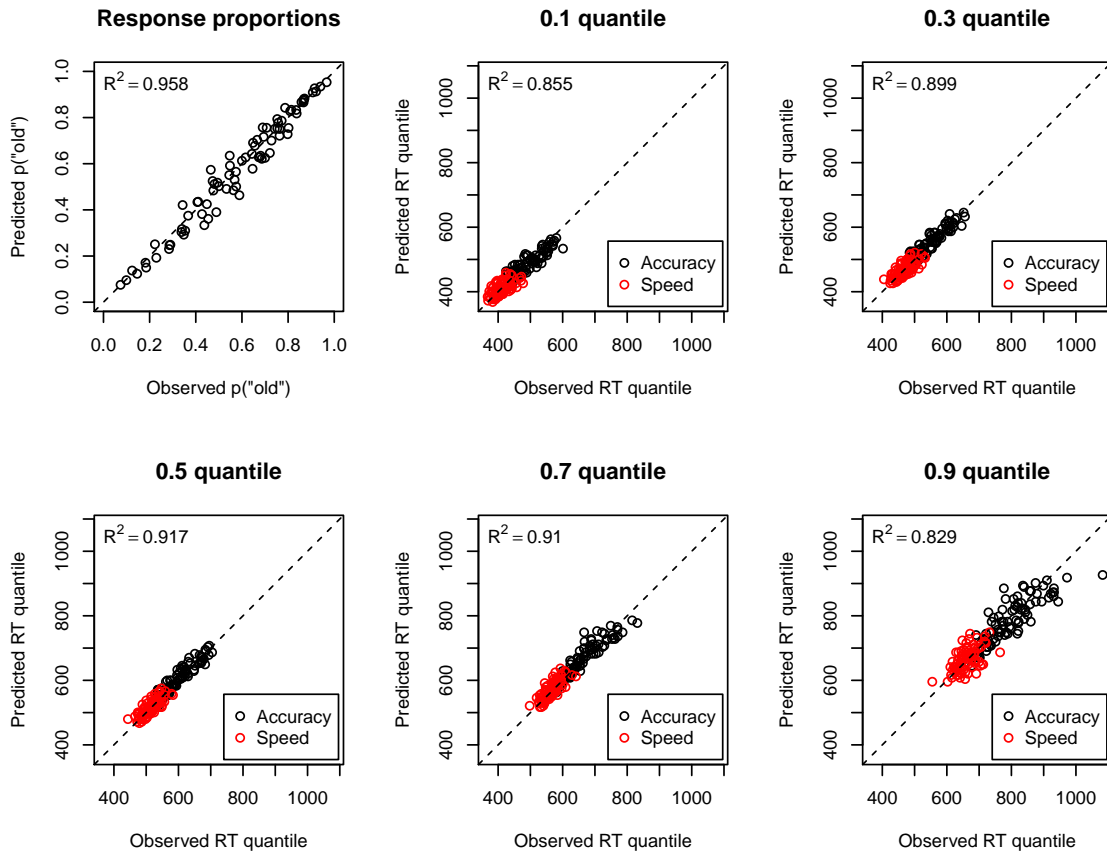


Figure 3.7: Diagnostic plots for the preferred model fit (#5) to the data from Starns et al. (2012), with proportion of variance accounted for indicated in the upper left corner.

χ^2 of 5533 on these data¹, which is worse than both the diffusion model ($\chi^2 = 2418$) and our BIC-preferred dynamic model (model 5, $\chi^2 = 5190$). Thus, while the diffusion model provides the best quantitative fit of the models under consideration, our dynamic model is competitive with other accounts like RTCON and represents a considerable savings in parameters (38 for our model versus 66 for the diffusion).

3.3.4 Discussion

Once again, we have shown that our dynamic model for recognition does a good job fitting free response data quantitatively. Although in these datasets, in contrast with those from (Nobel & Shiffrin, 2001), a diffusion model was always able to provide a better quantitative fit, it generally required considerably more parameters to do so. Furthermore, those parameters cannot be directly tied to memory processes, which is the primary goal of our modeling efforts: To develop a model that jointly describes how memory retrieval and decision evolve over time, and use this model to increase our understanding of memory. To that end, we have made considerable progress in discovering the source of word frequency effects in recognition.

Across all model fits, the best explanation for word frequency effects in recognition memory was that LF words contain more distinctive features that are less likely to match by chance (e.g., Shiffrin & Steyvers, 1997). In contrast, no support was found for a model in which word frequency effects arose from interference from prior occurrences of the word in life history (e.g., Nelson & Shiffrin, 2013). Indeed, no support was found for any account of word frequency effects that predict that word frequency effects are time-

¹Although we fit by minimizing NLL, for as large samples as we have here, this is essentially equivalent to minimizing χ^2 .

dependent. Regardless of response signal lag (Hintzman et al., 1994) or speed vs. accuracy focus (Starns et al., 2012; Rae et al., 2014), word frequency effects persist. Thus, our investigation appears to also rule out recall-based (Reder et al., 2000), criterion-setting (Glanzer & Adams, 1990; Hintzman, 1994), and time-specific attentional (Malmberg & Nelson, 2003) accounts of the word frequency effect in recognition, unless those secondary processes occur so rapidly as to be undetectable, at which point they cease to become useful theoretical constructs.

Although we have shown that interference from *specific* prior episodes does not underlie the word frequency effect in recognition, we do not claim that such interference does not exist (cf. Nosofsky, Cao, et al., 2014). Words, however, have been experienced over the course of a participants lifetime, albeit at different rates and in different contexts. Thus, we believe that there is, in fact, a deep relationship between our account of the word frequency effect and that offered by “context noise” models (Dennis & Humphreys, 2001). In particular, to the degree that semantic features are used to encode traces of a word at study—the content features of the trace—these features must have been learned over the lifetime. This prior learning, naturally, took place in contexts prior to that of the experiment.

A wealth of research into semantic memory has supported theories in which learned semantic features are dependent on the contexts in which a word occurs, whether that context is other nearby words (e.g., Lund & Burgess, 1996; M. N. Jones & Mewhort, 2007; Howard, Shankar, & Jagadisan, 2011), documents in which the word occurs (e.g., Landauer & Dumais, 1997; Griffiths, Steyvers, & Tenenbaum, 2007), or perceptual features of the environment in which the word is used (e.g., Barsalou, 1999; Rogers & McClelland, 2004)². As outlined in our model, when a word is encountered in a study episode, the con-

²Perceptual and linguistic context appear to carry a great deal of redundant information about word mean-

tent features of the trace that is formed derive from these learned semantic features, which in turn reflect the contexts in which the word is typically used in language. Because LF words tend to be used in more distinctive contexts, the semantic features by which they are encoded are correspondingly more distinctive, just as we have inferred from our model fits (M. N. Jones, Johns, & Recchia, 2012; Nelson & Shiffrin, 2013). One can even use representations derived from a context-based model of semantics to predict episodic recognition, enabling more fine-grained item-level predictions (Johns & Jones, 2010), so our model might be considered a more abstract “stand-in” for such an approach.

The key difference, then, between our model and a pure context noise model like BCD-MEM (Dennis & Humphreys, 2001) is that, in our model, we distinguish between two senses of “context”: Although representations in semantic memory are learned by aggregating over many prior contexts, episodic memory traces distinguish between semantic features and the environmental/task context. These two kinds of memory representations undoubtedly exist in a perpetual cycle, where each new event updates the semantic representation which is then used as a basis for encoding new events (e.g., Kumaran & McClelland, 2012; Nelson & Shiffrin, 2013).

ing (Riordan & Jones, 2010), although the two forms of experience can interact to support additional learning and inference (Johns & Jones, 2012).

Chapter 4

Secondary Processes in Recognition

Several theories of recognition posit the operation of two processes, in contrast to the model we have presented which involves only a single process by which familiarity evolves over time to lead to a recognition decision. In this section, we consider a variety of scenarios in which there is evidence for multiple processes operating at recognition and show how our model can be extended to account for them. In particular, we challenge the common claim that the secondary process is always a form of “recollection”, in which details of an event are retrieved from episodic memory and used to make a recognition decision. Recollection is typically contrasted with “familiarity”, which is assumed to depend only on the strength of the test item, and which might be identified with the model we have thus far proposed. The distinction between familiarity and recollection is referred to as “dual process” theory (for reviews of these theories and their empirical support, see Yonelinas, 2002; Mandler, 2008). Our attempt is not to argue for or against “dual process” models of recognition—indeed, the models we propose entail the operation of more than one process. Instead, our aim is to explicitly characterize the secondary processes involved, such that the resulting theory enhances understanding of how recognition operates.

A popular approach to distinguishing between familiarity-based and recollection-based responses is the “remember-know” paradigm (Tulving, 1985). In this paradigm, partici-

pants are instructed to give a “know” response if the test item only seems familiar, while they must give a “remember” response if they consciously recall having seen the test item before. It is unclear to what extent this subjective experience actually dissociates between two qualitatively different processes, or even if the two responses arise from different processes at all (W. Donaldson, 1996; Hirshman & Master, 1997; Dunn, 2004, 2008). Most critically for our purposes, correct response times for “remember” judgments are faster than those for “know” judgments (Dewhurst & Conway, 1994), even when confidence is controlled (Rotello & Zeng, 2008). This stands in contrast to the assumption that recollection—which is presumed to underlie “remember” responses—is slower than familiarity. We therefore believe that the remember-know procedure is not alone sufficient to characterize secondary processes in recognition.

Another technique used to determine the extent of the contribution of a secondary recollection process is the process dissociation procedure (Jacoby, 1991). This procedure begins by assuming that familiarity makes an equal contribution to recognition decisions across conditions, but that different test instructions can induce conscious recollective strategies. By comparing conditions in which certain information is relevant for the decision (e.g., accepting only items presented in an auditory rather than visual modality) with conditions in which that information is irrelevant for recognition, Jacoby’s procedure allows researchers to obtain an estimate of the relative contribution of recollection to recognition. However, this procedure will nonetheless produce high estimates of the contribution of recollection even when the data are simulated from a single-process model where it is known that recollection did not generate the data, and when a model with recollection is used to generate the data, the process dissociation estimates of recollection’s contributions are incorrect (Ratcliff, Van Zandt, & McKoon, 1995). The process dissociation procedure also relies on

the strong assumption that the two processes are stochastically independent of one another; when the assumption is violated, estimates obtained from this procedure are uninterpretable (Curran & Hintzman, 1995; Hillstrom & Logan, 1997). Indeed, it has long been established that there is a correlation between the probability of correct recall and that of correct recognition (Tulving & Wiseman, 1975; Kahana, Rizzuto, & Schneider, 2005), strongly arguing against the process dissociation assumption of independence (item and source memory performance are also correlated; Starns, Rotello, & Hautus, 2014). Finally, even if the assumption of the process dissociation procedure are satisfied, it is insufficient for identifying the relevant processing components (Humphreys, Dennis, Chalmers, & Finnigan, 2000). The process dissociation procedure cannot provide compelling evidence for or against the operation of secondary processes in recognition, let alone describe what those processes might be.

We believe that the debate over the relative contribution of secondary processes in recognition can be best advanced by attempting to develop explicit models of such processes, which we now proceed to do. As we have emphasized throughout, such a model should be dynamic in that it specifies how the state of the process evolves over time. Thus, for our purposes, accuracy data alone—which has been the primary source of evidence for dual-process theories—is not sufficient. Rather, we consider the clearest evidence for secondary processes to arise from speed-accuracy trade-off studies. When such a study finds a non-monotonic response curve, this is evidence that a secondary process or secondary source of information begins to contribute to the decision. Our aim in this section is to review studies that provide evidence for secondary processes in recognition and then to characterize these processes within our dynamic framework. In so doing, we cannot claim to be truly “predicting” the data, since in general we will need to tailor some aspect of the

model’s structure to each experiment. While some may then criticize our work for “re-describing” the data, this criticism is overly narrow. Our descriptions of the mechanisms needed to explain these data afford greater clarity and understanding by virtue of being embedded in our larger dynamic approach to recognition. In addition, by providing an explicit mechanistic account, our candidate explanations are made testable by experiment.

4.1 Highly similar foils

The operation of a recall-like process has been implicated in situations where participants must detect foils that are highly similar to studied items. It has been proposed that participants make use of a “recall-to-reject” strategy wherein the target memory trace is retrieved and compared to the current test item; a similar foil can then be rejected by direct comparison of the trace to the test item. The main evidence for this theory comes from speed-accuracy trade-off studies in which, at short signal lags, participants are likely to call similar foils “old”, with this tendency reversing at longer signal lags. We argue that, instead of recall, this finding is better explained by a delay in the availability of distinguishing features. That is, the features that would be needed to detect the mismatch between the test item and memory are not available until later in the trial. Such an explanation is more consistent with the finding that the time at which similar foils begin to be rejected is the same regardless of the strength of the similar target item. It is also easier to reconcile with the fact that only certain kinds of similarity lead to non-monotonic false alarm curves. We proceed to describe the evidence for secondary processes in the rejection of similar foils, and show how an extension of our dynamic recognition model can account for these findings.

Hintzman and Curran (1994) report a set of SAT studies in which participants studied a list of nouns. In their second and third experiments, test items comprised old words,

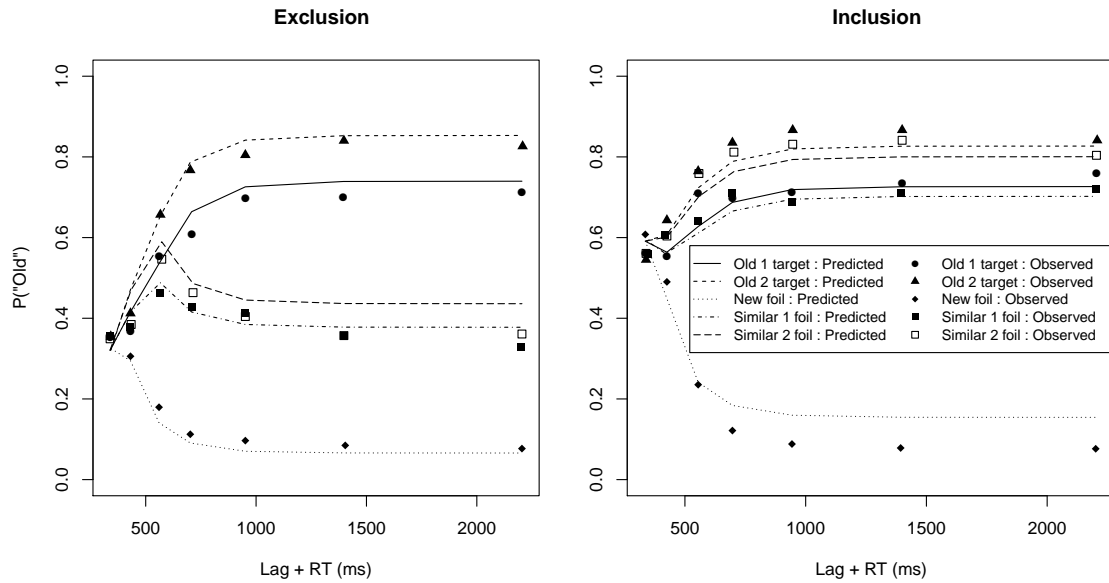


Figure 4.1: Observed and predicted probability of responding “old” as a function of response signal lag plus response time from Hintzman and Curran (1994, Exp. 3). See text for details.

new words, and words that were studied in the opposite plurality (e.g., if “apple” had been studied, “apples” might appear as a test item). When participants were told to accept only words in their exact studied form (Exp. 2 and the “exclusion” condition of Exp. 3), participants were initially more likely to endorse similar items that had been studied in the opposite plurality before reversing and correctly rejecting such items (Figure 4.1, “exclusion”). Hintzman and Curran proposed that, in order to correctly reject similar foils, participants had to explicitly retrieve the trace of the corresponding study item and compare the stored plurality to the plurality of the similar test item; if they failed to match, the participant would reject the item.

One explanation for these results that does not involve a secondary recollection process is that response bias changes over different signal lags. Rotello and Heit (1999) reanalyzed Hintzman’s and Curran’s data in terms of d_L , a discriminability measure similar to d'

that factors out the effect of bias. Under this analysis, d_L curves were strictly monotonic, suggesting that similar foils did not become more discriminable from completely novel foils later in the trial, as one might expect from a "recall-to-reject" strategy. However, it is not clear why bias should differ as a function of signal lag only for similar items; indeed, Hintzman et al. (1994) found that bias did not differ across lags. Without a compelling account of why response bias might change over signal lags for specific conditions in these experiments, we consider this explanation lacking.

Brockdorff and Lamberts (2000) offered a different explanation: Instead of needing a secondary retrieval process, plurality may simply be a feature that is not available until later in processing. That is, semantic features of a word might become active earlier, leading participants to endorse a foil regardless of its plurality, but only later does the discriminating plurality feature become available to be used to reject the item. In the context of our model, rather than the plurality feature "activating" at a later time, we would say that the features of the item pertaining to its plurality are only available for sampling later in the trial. In support of this explanation, the inflection point in the false alarm curves to reverse-plurality foils is the same regardless of whether the similar item had been studied once or twice. This follows directly from the assumption that plurality information isn't available until later in the trial, since the time at which plurality features are available to be sampled would not depend on memory. Recall, on the other hand, would depend on the strength of the corresponding memory trace, predicting that "similar 2" foils (foils similar to an item that had been studied twice) could be rejected earlier than "similar 1" foils, which is not seen in the data.

Under the assumption that plurality information is not available until later in a trial, we fit a version of our dynamic recognition model to Experiment 3 from Hintzman and Curran

(1994). In this experiment, participants studied lists consisting of 46 nouns, evenly split between singular and plural nouns. Half of the studied nouns were repeated on the study list. Test lists consisted of old items that appeared once or twice on the study list (labeled “Old 1” and “Old 2”, respectively), words that were the opposite plurality of a word that had been studied either once or twice (“Similar 1” and “Similar 2”, respectively), and completely novel items (“New”). Prior to the test lists, participants were told either to reject any item that did not exactly match a word on the study list (the “Exclusion” condition) or to accept words even if they did not match the studied plurality (the “Inclusion” condition). Thus, similar foils should be rejected in the exclusion condition, but accepted in the inclusion condition. Participants, then, need not attend at all to plurality in the inclusion condition, since it is no longer useful for distinguishing targets and foils.

We assume that a certain proportion of content features, p_P , represent the plurality of a word and that these features become available for sampling at time t_P ¹. Before that time, only non-plurality (e.g., semantic) features are available to be sampled, but after that time, all features are available for sampling. Although all plurality features are used in the exclusion condition, participants attempt to ignore them in the inclusion condition, since they are not needed to discriminate targets from foils. Notice, however, that the probability of calling a reverse-plurality item “old” in the inclusion condition is still lower than the probability of a hit to an actual old item; thus, we assume that a proportion p_I of plurality features are still available for sampling even in the inclusion condition, thus harming the match between a reverse-plurality test item and its corresponding memory trace. Because participants are not told what kind of test condition they are in until *after*

¹The proportion of features dedicated to representing plurality may be less in experiments where plurality is not critical for distinguishing targets from foils.

Table 4.1: Best-fitting parameters of the dynamic model to the data from Hintzman and Curran (1994, Exp. 3).

<i>Parameter</i>	<i>Value</i>
u_1	0.415
u_2	0.528
c_S	0.976
$\theta_{\text{Inclusion}}$	0.425
$\theta_{\text{Exclusion}}$	1.152
p_P	0.134
t_P	374.961
p_I	0.0585
t_0	87.138
ρ	5.034
R^2	0.958

study, attempting to ignore the plurality features of the test item in the inclusion condition means that participants cannot make use of the plurality features that were stored at study, effectively reducing the number of content features available on which to base a decision.

Besides p_P , t_P , and p_I , we fit separate decision criteria θ to each condition (owing to the different prior probabilities of a positive response—1/3 in the exclusion condition and 2/3 in the inclusion condition) and single values of c_S , t_0 , and ρ across both conditions. We also allowed u to vary with number of repetitions (u_1 , u_2); note that u applies equally to the semantic and plurality features of the trace, such that a repeated item has, on average, more of both types of features. The fitted values of these parameters are given in Table 4.1. As shown in Figure 4.1, the model captures the trends in the data both qualitatively and quantitatively. Obviously, our account was deliberately tailored to fit these data, so this

should not be surprising. Nonetheless, it shows that our account is reasonable and that the dynamic approach to recognition can be easily extended to apply to new paradigms.

A delay in sampling plurality features could be attributed to a strategy adopted by participants specifically for this task. It could also reflect the operation of lexical access, wherein both the singular and plural forms of the word are initially activate in the lexicon, with additional time needed to access a specific lexical trace. Although possible (New, et al., 2004), this explanation merely “kicks the can” from the dynamics of episodic retrieval to the dynamics of lexical/semantic retrieval; more work must be done to elucidate the relationships between these processes (e.g., Hintzman & Curran, 1997). Another explanation for the relative lateness of access to plurality features is that it reflects a narrowing of attention over time, from an initial broad distribution over the whole word to a subsequent focus on plurality-specific features. Such a narrowing of attention has been implicated in a variety of tasks, particularly the flanker task (e.g., White, Ratcliff, & Starns, 2011), and forms a critical component of theories of visual detection (Smith, Ratcliff, & Wolfgang, 2004; Ratcliff & Smith, 2010) and changes in preference over time (Busemeyer & Townsend, 1993; Diederich, 2003). An attentional explanation is also consistent with the Stroop task, wherein the semantic features of a word capture attention quickly, and it requires additional time and/or effort before finer details (like print color or plurality) of the word are available.

Finally, we note that Malmberg, Holden, and Shiffrin (2004) also used opposite-plurality foils in a recognition paradigm where the number of repetitions spanned a much larger range (0-12) than in Hintzman and Curran (1994, Exp. 3). They found that the false alarm rate to similar foils did not substantially increase with the number of repetitions of the corresponding target at study, arguing that the increase in familiarity was counteracted by the increased availability of recall with repetition. Because their study did not examine

response times or speed-accuracy trade-off, we will not address it here in full, although we acknowledge that, given the much larger number of study repetitions involved in their study, it may be that recall or some other process is playing a role in their results that is not apparent with smaller numbers of repetitions. We also note, however, that hits and false alarms to similar foils can rise at different rates if repetition increases the storage of *both* semantic and plurality information, as we have assumed throughout. The reason is that, in a likelihood-based model like ours, a mismatching feature value (say, in plurality) can count as stronger evidence *against* a match between the probe and trace than a matching feature (say, in semantics) counts in favor of a match. This can be clearly seen in the simplified likelihood computation in Equation 2.7 where, as $c_S \rightarrow 1$, the match term approaches $\frac{1}{1-g}$ while the mismatch term approaches $1 - c_S = 0$, overwhelming any positive evidence from matching features. This feature of the model is illustrated in the fit to the Exclusion condition (Figure 4.1), in which the asymptotic false alarm rate to “Similar 2” items (words of the opposite plurality to a repeated studied word) is lower than the asymptotic hit rate to “Similar 1” items.

4.1.1 Other kinds of similarity

Our account of the data of Hintzman and Curran (1994) depends on the assumption that plurality information is available late in the trial, while semantic features (that do not on their own convey any information about plurality) are available early, a view consistent with the early availability of semantic information in word recognition (Feldman, O’Connor, & Del Prado Martín, 2009). That is, the features that distinguish between a target and a similar foil are of different *kinds*. When similarity is defined in terms of a single type of feature, we would not expect any difference in the time at which distinguishing features would

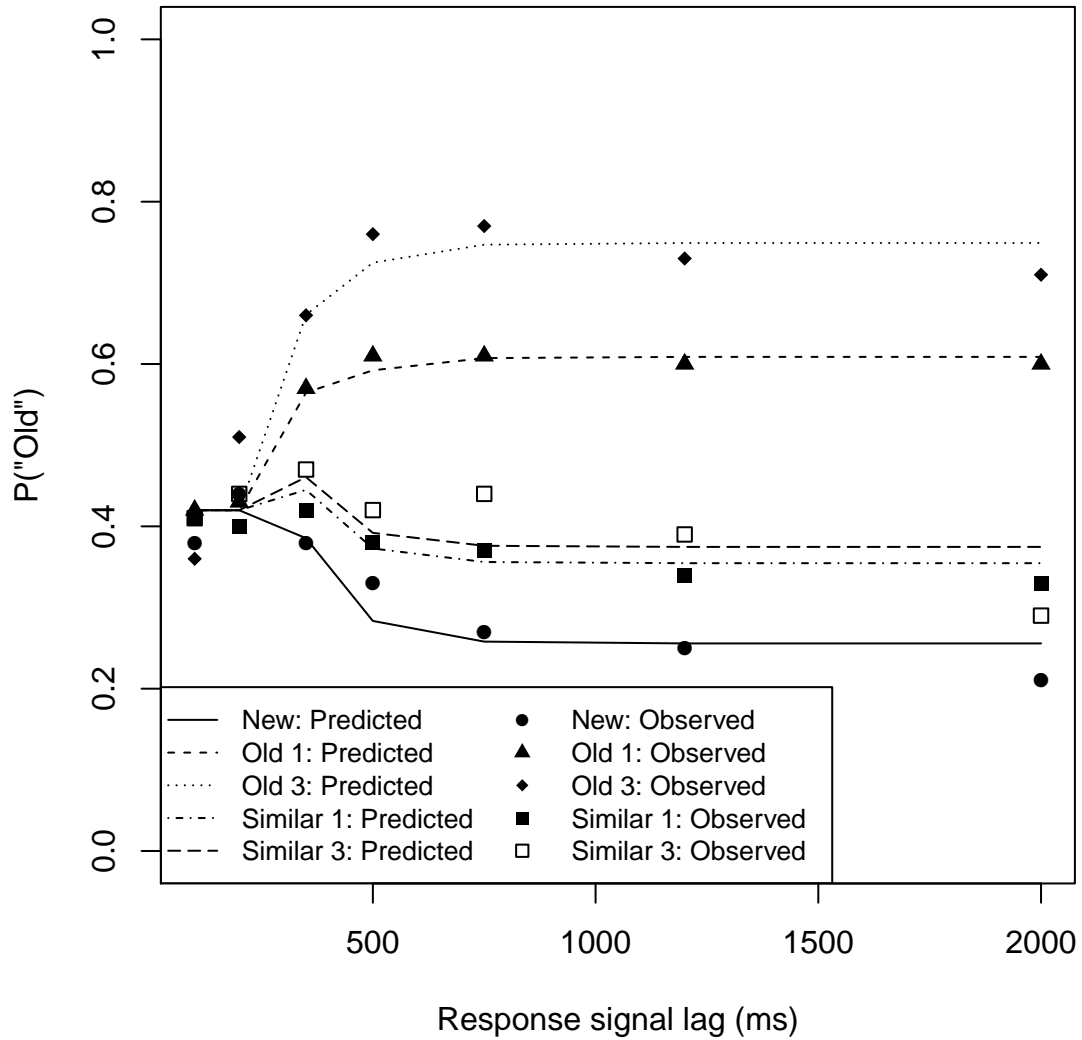


Figure 4.2: Observed and predicted probability of responding “old” as a function of response signal lag for the three “unpaired” condition in Rotello and Heit (1999, Exp. 2). Model parameters used to generate predictions are given in Table 4.2.

Table 4.2: Best-fitting parameters of the dynamic model to the data from the “unpaired” condition of Rotello and Heit (1999, Exp. 2).

<i>Parameter</i>	<i>Value</i>
u_1	0.294
u_3	0.383
c_S	0.974
s	0.711
θ	0.330
t_0	193.189
ρ	3.357
R^2	0.878

be available, and consequently we would not expect any strong nonmonotonicity in false alarm curves. Rotello and Heit (1999) replicated Hintzman and Curran (1994) using pseudowords, thus diminishing the role of semantic features and ensuring that “similarity” could depend only on orthography. Data from and model predictions for the “unpaired” condition of Rotello and Heit (1999, Exp. 2) are shown in Figure 4.2; this condition replicates the general design of the exclusion condition of Hintzman and Curran (1994, Exp. 3), but instead of unrelated words, participants studied unrelated pseudowords either once or three times (labeled “Old 1”, “Old 3”). Similar foils were created by altering some letters of one of the studied words. As shown in Figure 4.2, there is little suggestion of the kind of sharp nonmonotonicity observed by Hintzman and Curran (1994), and our model fits these data quite well without assuming any delay in the availability of orthographic features that distinguish old items from similar foils (model parameters are given in Table 4.2). The model still predicts a slight nonmonotonicity in the false alarm curves to similar foils, but this is

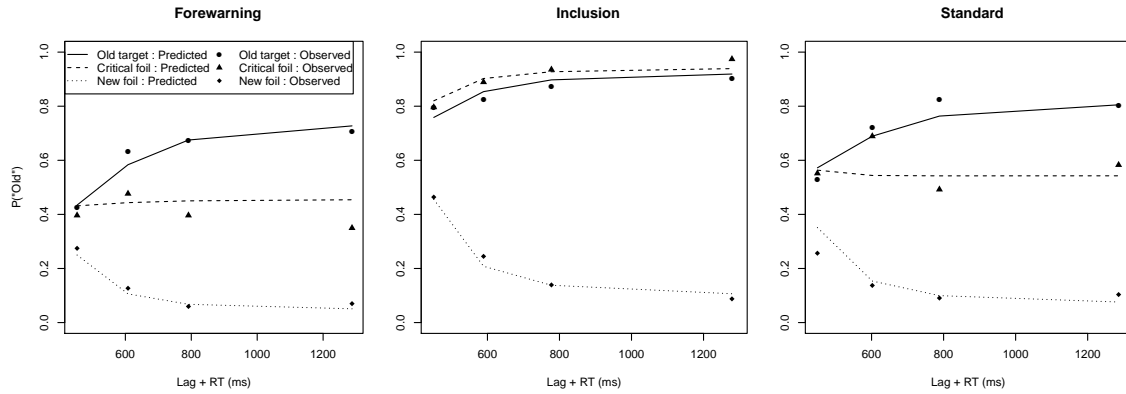


Figure 4.3: Observed and predicted probability of responding “old” as a function of response signal lag plus response time for the three instructional conditions in Heit et al. (2004, Exp. 2). Parameters used to produce model fits are given in Table 4.3.

a consequence purely of the normal feature sampling process: Given that the majority of features of similar foils match those of the corresponding target (represented by parameter s), the first few features sampled are more likely to increase the match between the probe and the target trace; only with the later sampling of additional mismatching features does the match begin to decrease.

Just as there is no difference in the time at which distinguishing features are available when similarity is defined purely in terms of orthography, there is no difference in the time at which distinguishing *semantic* features are available. Heit et al. (2004) conducted several response signal variants of the Deese-Roediger-McDermott (DRM; Deese, 1959; Roediger & McDermott, 1995) paradigm, in which participants study lists of words containing several close associates (e.g., bed, pillow, dream) of an unstudied critical word (e.g., sleep). In their second experiment (Figure 4.3), three sets of instructions were given to different groups of participants with regard to the recognition test: 1) a “standard” set of instructions in which participants were told to respond “old” only to words that were on the study list;

Table 4.3: Best-fitting parameters of the dynamic model to the data from Heit et al. (2004, Exp. 2).

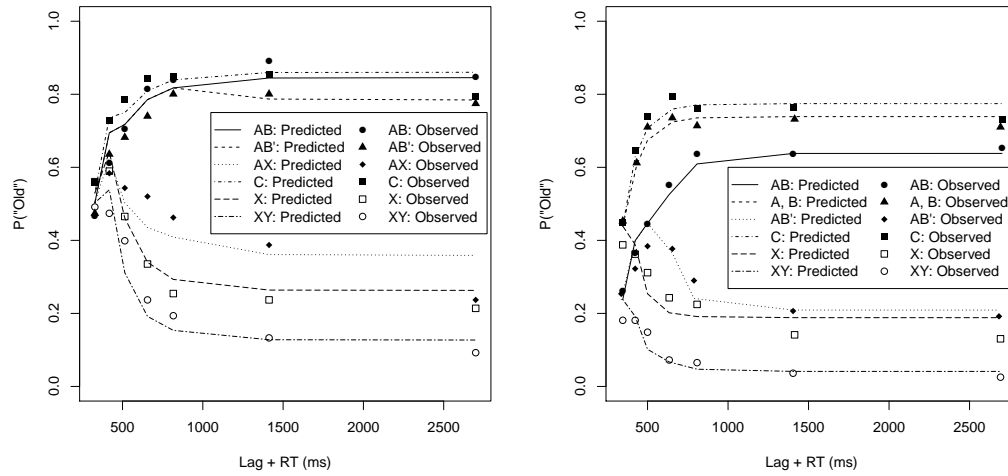
<i>Parameter</i>	<i>Value</i>
u	0.444
c_S	0.973
p_S	0.166
$p_{S,\text{Inc}}$	0.528
s_C	0.438
$\theta_{\text{Forewarning}}$	1.486
$\theta_{\text{Inclusion}}$	-0.138
θ_{Standard}	0.608
t_0	50.915
ρ	6.202
R^2	0.967

2) “forewarning” instructions that informed participants that the test list would contain foils that were related to the studied words; and 3) an “inclusion” condition in which participants were told to endorse not only studied words but words that were similar (i.e., the critical word). We assume that a certain proportion p_S of content features represent the semantic features shared among each group of associated words, such that these have equal values for the critical word and its studied associates; in the inclusion condition, this proportion is slightly higher (denoted $p_{S,\text{Inc}}$) reflecting the increased attention to similarity in response to instructions (which were given prior to study and hence could influence both encoding and retrieval). The critical word, in addition, shares other content features with its associates with probability s_C , reflecting its central status within the set of associates. Finally, we fit decision criteria for each condition (owing to the different instructions) and single values of u , c_S , t_0 , and ρ across all conditions.

The fitted values of these parameters are given in Table 4.3. As shown in Figure 4.3, the model captures the critical finding that, in the Standard and Forewarning conditions, participants are initially likely to false alarm to the critical word, with this tendency reversing as time passes. This is a direct consequence of large proportion of shared features between the critical word and its associates, which because $s_C > 0$ is higher, on average, than the degree of match between two studied associates. As more features are sampled, enough non-matching features enter the probe that the initial good match is overwhelmed, except in the Inclusion condition in which there are a sufficient number of shared features ($p_{S,Inc}$) to maintain a high rate of endorsement of the critical word. Our fits agree with the conclusions of Heit et al. (2004) by finding that the decision criterion is most lenient in the Inclusion condition and least in the Forewarning condition, where avoidance of false alarms to similar items was emphasized. Finally, we note that we have treated the notion of semantic similarity in a purely abstract way here, as a set of free parameters, but a more detailed account of these data is possible if the semantic features are derived from a model of semantics (Johns & Jones, 2010).

4.2 Associative recognition

In associative recognition, participants study sets of items simultaneously, such as pairs of words. They are then tested on both “intact” pairs—words that appeared together at study—and “rearranged” pairs—pairs formed from two studied words that were not presented together. Thus, associative recognition goes beyond single-item recognition in that one cannot discriminate intact from rearranged pairs on the basis of separate recognition judgments for each word. For that reason, many theorists assume that associative recognition requires a secondary process that is not present in single-item recognition. This



(a) Experiment 1 (inclusion instructions). (b) Experiment 2 (exclusion instructions).

Figure 4.4: Probability of endorsing a test item as a function of response signal lag and response time from Gronlund and Ratcliff (1989), along with predictions from our dynamic recognition model. Parameters used to generate these predictions are given in Table 4.4.

process has been described as a form of cued recall, wherein one pair member is used as a cue to retrieve the item that it had been studied with; if the recalled word matches the other word in the test pair, an “old” response is given, otherwise the pair is judged “new” (ANDERSON). Alternatively, the secondary process involved in associative recognition may be the formation of a compound probe (Doshier & Rosedale, 1989, 1997): Initial judgments might be based on separate assessments of the two members of the pair, but later judgments are made by comparing the two words jointly to the contents of memory. We argue that the dynamics of associative recognition are better described by a compound cue model than one based on recall. Before describing an extension of our dynamic recognition model for associative recognition, we explore the evidence in favor of secondary processes in associative recognition and show that they do not entail a recall-like process.

The critical evidence for a secondary process in associative recognition is exemplified

in a series of response signal experiments by Gronlund and Ratcliff (1989), the results of which are shown in Figure 4.4. In their Experiment 1, participants were instructed to endorse a test pair if both of its members had been studied, regardless of whether they had been studied together (labeled AB and AB', in contrast to novel pairs XY or pairs with one novel member AX); this decision involves only item-level information. In Experiment 2, participants were told to endorse a pair only if its two members had actually been studied together (only AB); this decision requires associative information. Under these exclusion instructions, participants initially endorsed rearranged pairs (AB') before this tendency reversed at later signal lags. Rotello and Heit (1999, 2000) did not find a similar nonmonotonic pattern when single words were tested against similar foils (orthographically similar words in 1999 and synonyms in 2000), suggesting that rearranged pairs are not simply "similar" versions of the intact studied pairs, but that correctly rejecting them requires the involvement of a secondary process or source of information that is not available early in associative recognition. Furthermore, as we shall see, not all studies of the dynamics of associative recognition find a nonmonotonic false alarm curve to rearranged foils (Wickelgren & Corbett, 1977; Doshier, 1984a), suggesting that this secondary process is flexible or is not always deployed.

Many researchers have described this secondary process as a form of recall. Early false alarms to AB' pairs result from the familiarity of both members of the pair, while later correct rejections arise from cued recall. One of the words, say word A, is used to cue recall of the item it had been studied with (call this word B). If the recalled word (B) fails to match the other word in the test pair (B'), the test pair AB' is correctly rejected. However, in their Experiment 4, Gronlund and Ratcliff (1989) attempt to test this "recall-to-reject" account by presenting the first word of the pair 200 ms before the second; if participants

were using a recall-to-reject strategy, they should use the first word to cue recall of the second, allowing them to correctly reject a rearranged pair much faster after the second word appeared and, perhaps, without any nonmonotonic false alarm curve. Instead, they found that, although participants could reject novel foils (XY) more quickly when given a 200 ms head start, that same time conferred no advantage in rejecting AB' foils. Rotello and Heit (2000) contend that 200 ms is quite brief and may not have been long enough to allow for participants to use the first word as a retrieval cue, since it would take participants roughly that amount of time to even read the word.

Nobel and Shiffrin (2001) directly compare the dynamics of cued recall and associative recognition in a set of signal-to-respond experiments (in recall, participants either generated a response word immediately after the signal or were allowed to make a "give up" response). They found that discriminability between intact (AB) and rearranged (AB') pairs, as measured by a variety of indices, rose to asymptote with dynamics comparable to that of the probability of correct cued recall. Of course, similarity in dynamics does not, on its own, entail that the same mechanisms give rise to those dynamics. In their third experiment, Nobel and Shiffrin (2001) tested the operation of recall in recognition in two ways: First, they manipulated both item strength (as a function of study time) and interference (as a function of the number of studied pairs per list), allowing us to assess whether retrieval dynamics vary with memory strength, which would be good evidence for recall. Second, in addition to usual associative recognition, they use a variant of cued recall developed by Wickelgren and Corbett (1977) which they called "wickelcall" in which only one word of a pair (A) is shown until the response signal is given, at which point the second word (B or B') appears and participants must immediately give an "old" (AB) or "new" (AB') response. Participants are explicitly instructed to attempt to recall the word that had been

studied (B) with the given cue (A); if participants are able to recall the correct associate by the time the second word appears, they should be able to easily decide whether the second word matches that retrieved word or not.

They found, first of all, that there was no reliable difference in the dynamics of either associative recognition or wickelcall as a function of study time or list length, with only asymptotic performance, not intercept or growth rate, varying significantly between conditions. This replicates the findings of Doshier (1984a), who also found that study time affected only asymptotic performance and not retrieval dynamics. This result fails to support the operation of recall during associative recognition, at least at signal lags of up to 4500 ms, since recall should operate more quickly not just on stronger items (those studied longer), but when there is less interference (shorter list in the case of Nobel & Shiffrin or smaller associative fan in Wickelgren & Corbett). Second, although the dynamics of wickelcall and associative recognition, as measured by the intercept and growth rate of the exponential functions fitted to the data, were statistically equivalent, asymptotic performance in wickelcall was considerably lower than that of associative recognition, replicating Wickelgren and Corbett (1977). One explanation, offered by Wickelgren and Corbett (1977) and echoed by Nobel and Shiffrin (2001), is that a cued recall process operates in both wickelcall and associative recognition, but the presence of two possible cues in associative recognition enables more to be retrieved. A second, closely related, possibility also mentioned by Nobel and Shiffrin (2001) is that increased performance in associative recognition results from the use of a compound cue, which is impossible to form when only a single pair member is present (e.g., Clark & Shiffrin, 1987; Doshier & Rosedale, 1989, 1997).

The relatively low performance in wickelcall and cued recall brings to light an issue

with the usual “recall-to-reject” account: what should a participant do if recall fails or provides an incorrect answer? Dual process theorists (Reder, Jacoby, etc.) often assume that, when recall occurs in recognition, it is veridical. When people are actually given a recall task, of course, they are far from perfect, not only failing to retrieve items, but producing incorrect items as well. For example, in their second experiment, Nobel and Shiffrin (2001) report an asymptotic probability of correct recall of .24, with a .10 probability of recalling an incorrect item. Consider a simple non-dynamic dual process account of associative recognition, similar to Jacoby’s process-dissociation model, in which responses at asymptote are a mixture of those produced by a familiarity process and those produced by cued recall. Familiarity produces a correct response with probability p_F (for simplicity, and because we do not seriously pursue this model, this probability is identical for both targets and foils). Recall is able to occur by the final signal lag with probability p_R , and if it does occur, it produces a correct response with probability p_C . For simplicity, we assume that if recall results in an intrusion, there is effectively zero chance that it matches the other word in the pair (in reality, this chance is probably not negligible, but it is not necessary for the present argument). Then, the probabilities of a hit, $p(H)$, and false alarm $p(FA)$ are given by

$$p(H) = p_F(1 - p_R) + p_R p_C \quad (4.1)$$

$$p(FA) = (1 - p_F)(1 - p_R) \quad (4.2)$$

and we can solve for p_F and p_R to find

$$p_R = \frac{p(H) + p(FA) - 1}{p_C - 1} \quad (4.3)$$

$$p_F = \frac{p(H) - p_R p_C}{1 - p_R}. \quad (4.4)$$

By plugging in the observed values of $p(H) = 0.75$, $p(FA) = 0.21$, and $p_C = 0.24/(0.24 + 0.1) \approx 0.71$ from Nobel & Shiffrin (2001; Exp.2, Table 11), we find $p_R = 0.14$ and $p_F = 0.76$. In other words, the probability that familiarity generates a correct response, p_F , is higher than the probability of correct recall (0.71), and recall is estimated to operate on only a small proportion (14%) of trials. Although this analysis is hardly exhaustive, it represents a direct test of the “recall-to-reject” idea and demonstrates that, by allowing recall to be error-prone, the probability that associative recognition arises from a cued recall-like process may be quite low. Indeed, given that recall is not perfect, it may not even be a good strategy for participants to use, assuming they have another source of information on which to make their decision.

4.2.1 Modeling associative recognition

Given the difficulties with a recall-based account of associative recognition, we decided to model associative recognition via the formation of a compound cue. In developing this model, we must first decide how associations should be represented. We must choose a representation that makes it straightforward to model the dynamics of associative recognition as reflected in the data, such that associative information can be clearly separated from item information, while maintaining the ability of the model to account for single-item recognition. Although space precludes us delving into the rich literature on associative learning and recognition, we begin by noting that many representations of association are possible in a vector-based model like ours. Shiffrin and Steyvers (1998) suggest that associations be represented by concatenating the feature vectors corresponding to the single items present in the association, like so:

$$AB \rightarrow \begin{bmatrix} f_1^A & f_2^A & \dots & f_N^A & f_1^B & f_2^B & \dots & f_N^B \end{bmatrix}$$

where f_i^j denotes feature i of item j . Criss and Shiffrin (2004b) suggest how this kind of representation can be augmented to represent different types of pairs. While this representation makes sense when all the items studied and tested are pairs, it is harder to justify when, as in the two studies under consideration, both single items and pairs can appear at either study or test. It is unclear how the features of a single item should be aligned for comparison with a pair, although several options exist: One could assume that a single item at study is encoded by concatenating its trace with itself, in order to allow it to be compared to the elements of a pair; at test, a single item is compared in parallel to both the left and right components of each stored trace. While this modeling scheme is certainly feasible from a technical standpoint, it muddles the relationship between single items and pairs, and does not offer a straightforward way to separate associative from single-item information, as is required to model these data.

In the TODAM model (Murdock, 1982), a pair AB is represented by the storage of three separate traces, one for each element of the pair (A, B) and another representing the compound of the two element ($A \otimes B$, where \otimes represents the convolution of the feature vectors representing A and B). The compound $A \otimes B$ can be thought of as configural information that goes beyond that present in either item alone. Doshier and Rosedale (1989, 1997) found evidence that such configural information is used in associative recognition, although their data do not necessarily entail a particular representation of such information. Criss and Shiffrin (2005) found that interference in single-item recognition increased when items were studied in pairs across different contexts, suggesting that item and associative (configural) information is stored separately (cf. Buchler, Light, & Reder, 2008).

We therefore adopt a representational assumption based on that used by Criss and Shiffrin (2005) and related to that of TODAM, as depicted in Figure 4.5. When a pair

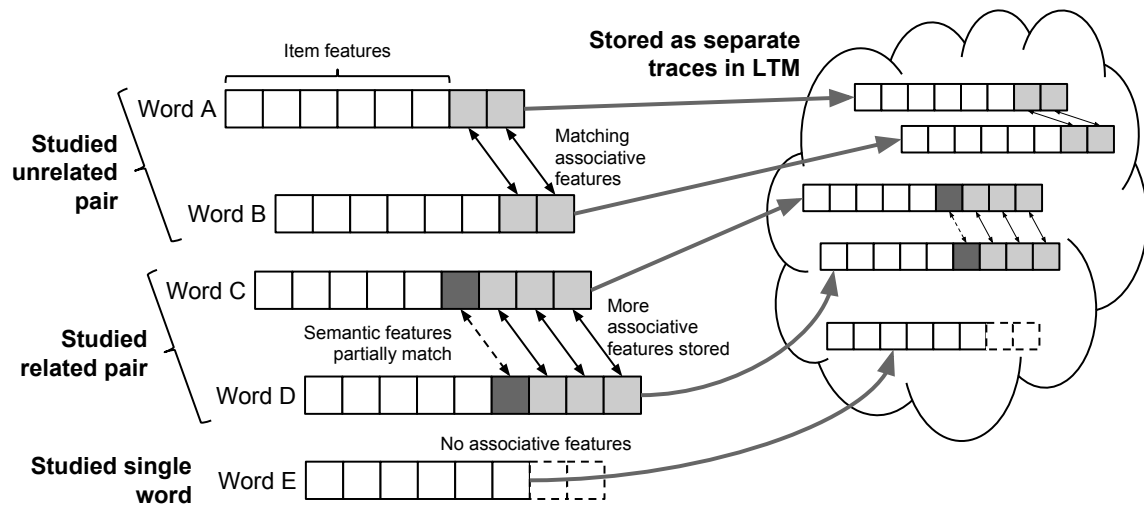


Figure 4.5: Schematic depiction of how studied pairs are stored as memory traces. See text for detail.

of words is presented at study, we assume that a separate trace is formed for each word. Each trace contains content features of the word itself as well as additional associative features. These associative features are “configural” in that they go beyond the features of either word alone. The true value of these features is equal between the two traces, although the features are still subject to noisy encoding (i.e., they may fail to be stored or have an incorrect value stored, as with any other feature). When the two items already have a prior association, such as the semantically-related word pairs used by Doshier (1984b) and Doshier and Rosedale (1991) discussed below, we assume that this allows for more associative features to be stored. This representation makes it clear how one could perform single-item recognition even when words were studied in pairs—the associative features are simply not used when a single item is tested.

Having decided on a way to represent the storage of word pairs in our model, we must now determine how to model recognition of multiple items. Shiffrin and Steyvers (1998); Criss and Shiffrin (2004b) suggest several ways a participant might make an associative

recognition decision. We begin by considering pair recognition, in which participants call a pair “old” if both of its members were studied, regardless of whether they were studied together (e.g., AB or AB’). Pair recognition performance is superior to single-item recognition performance, although the dynamics of single-item and pair recognition can differ, with pair recognition rising above chance at an earlier time than single-item recognition (Gronlund & Ratcliff, 1989; Nobel & Shiffrin, 2001). That pair recognition can occur slightly faster implies that participants are not taking extra time to read the two words, and the increased performance suggests that the two items can, in fact, facilitate one another. Shiffrin and Steyvers (1998) found, however, that the same parameters of the REM model used to predict single-item recognition would produce too high performance in pair recognition. To address this, they suggested two possibilities: 1) Although participants attempt to make a joint judgment about the pair, they are unable to make use of all the features of both items due to short-term memory capacity constraints (c.f. Gillund & Shiffrin, 1984). 2) Rather than making a judgment about the pair as a whole, participants assess each item individually and then combine those judgments. Indeed, Gronlund and Ratcliff (1989) found in their first experiment, where associative information was not relevant to the decision, performance is well characterized by a simple multiplicative model in which participants make independent assessments of the oldness of the two items and only respond “old” to the pair if each of its elements is thought to be old.

Thus, we assume that pair recognition proceeds by processing the two items in two parallel channels, using the same model for single-item recognition we have already proposed. Participants only make an “old” response if the accumulated change in familiarity for each item is sufficient to call *both* of them “old” (i.e., if both $x_A(t)$ and $x_B(t)$ are greater than θ at the time of the response signal). For each channel, we assume as we have thus far that the

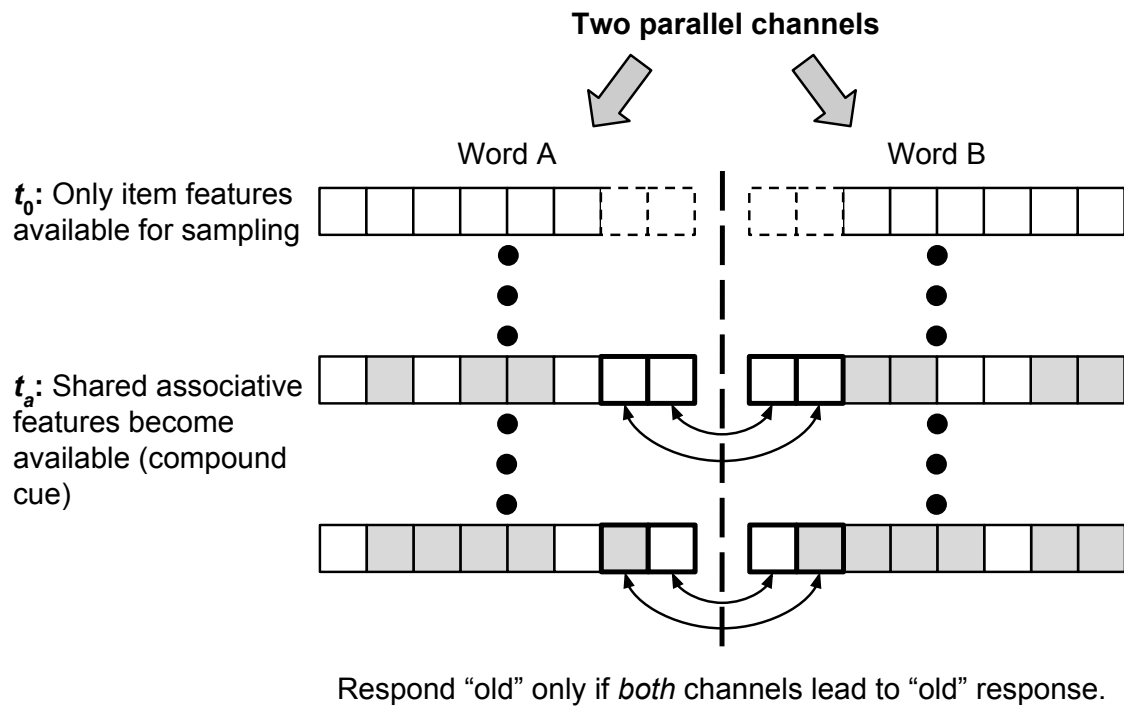


Figure 4.6: Schematic depiction of pairs are processed for recognition. In associative recognition, at time t_a , the participant begins to have access to associative features representing the compound of the two items. See text for detail.

probe begins with N_X context features and that there is available capacity for N_C content features. In pair recognition, a certain proportion of the N_C content features, denoted p_A , are given over to representing the association between the two items; the remaining features are used to represent the items. We allow for the possibility that, due to constraints on short-term memory capacity, not all item-specific features can be sampled into the two probes; we denote the proportion of the maximum number of item features (of which there are $(1 - p_A)N_C$) that can be sampled in pair recognition by p_S .

Associative recognition begins just like pair recognition, but at some time t_A after the start of processing, the $p_A N_C$ associative features become available for sampling. At that point, both item and associative features can be sampled into each probe. The only difference between item and associative features from the perspective of the retrieval process is that the values of the associative features match between the two probes. Just as in pair recognition, a participant only responds “old” if both probes result in sufficiently high accumulated change in familiarity. This process is depicted schematically in Figure 4.6. It is flexible enough to allow us to explain both single- and multiple-item recognition within the same modeling framework, and enables one to infer the degree to which associative information is used (p_A), when it is available (t_A), and to what extent it interferes with item-level processing (p_S). As we shall show, this flexibility is needed to account for the wide variety of data on associative recognition.

4.2.2 Model fits

We fit this model to the data from the first two experiments reported by Gronlund and Ratcliff (1989) and described above, where Experiment 1 involved only pair recognition while Experiment 2 required associative information. These experiments involved studying both

Table 4.4: Best-fitting parameters of the dynamic model for associative recognition to various datasets. In Doshier (1984a, Exp. 2), multiple study times were used, entailing multiple values of u_P ; study times were 1, 2, 4, and 6 seconds per pair.

<i>Parameter</i>	Gronlund and Ratcliff (1989)		Rotello and Heit (2000)		Doshier (1984a)	
	Exp. 1	Exp. 2	Exp. 1		Exp. 1	Exp. 2
u_S	0.316	0.414	—		—	—
u_P	0.339	0.388	0.439	0.356, 0.420, 0.437, 0.463		
c_S	0.988	0.980	0.984		0.969	
p_A	0.017	0.170	0.162		0.190	
t_A	751.655	411.847	437.489		447.302	
p_S	0.937	0.869	—		—	—
θ_S	-0.118	0.245	—		—	—
θ_P	-0.891	0.048	1.097		-0.228	
t_0	89.047	94.477	82.982		187.346	
ρ	5.880	4.096	7.899		9.639	
R^2	0.898	0.936	0.923		0.881	

pairs and single items; we allowed the probability of feature encoding to differ between these two classes (u_S for single items u_P for pairs). Even though participants did not need to use associative information in Experiment 1, we nonetheless fit values of p_A , the proportion of content features used to represent associations, and t_A , the delay in availability of associative features, for both experiments to verify the model's ability to detect this difference. Two criterion parameters were also required, one for single-item recognition θ_S and another that applied to both parallel channels in pair recognition θ_P . Finally, we fit values of p_S , c_S , t_0 , and ρ . Model predictions are shown in Figure 4.4, showing that the model closely matches the data, with the fitted parameter values in Table 4.4. Estimates of u_S and u_P are similar to each other within each experiment, consistent with the fact that participants were given twice as long to study a pair as they were a single item. As expected, p_A is much lower in Experiment 1, where associative information is not required, although the fact that it is non-zero suggests that some associative information is nonetheless encoded (this would account for the slight increase in asymptotic hit rate for AB pairs compared to AB'). It would also appear the increase in p_A in Experiment 2 is accompanied by a decrease in p_S , suggesting that a greater focus on associative features entails a smaller capacity for item features.

We also fit our model to two experiments that only tested pair recognition². In Rotello and Heit (2000, Exp. 1), participants had to distinguish between intact pairs (AB), rearranged pairs (AB'), and novel pairs (XY), rejecting both novel and rearranged pairs (data and predictions are shown in Figure 4.7). In Doshier (1984a, Exp. 2), participants only

²Accordingly, we assumed that $p_S = 1$ in these experiments since, without tests of both single items and pairs, this parameter becomes unidentifiable. It is still likely, of course, that participants are subject to capacity limitations on the number of item features that can be sampled into the two probes.

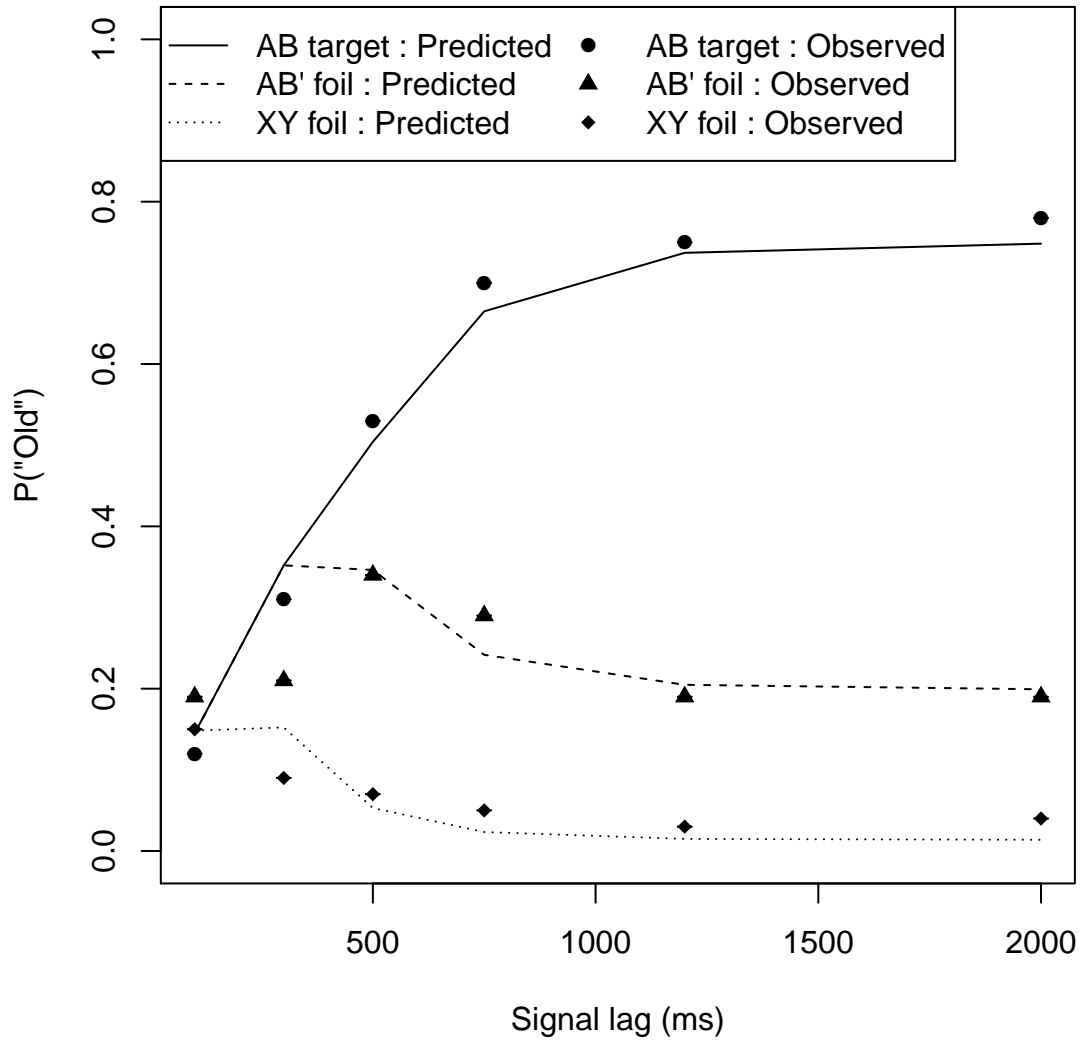


Figure 4.7: Probability of endorsing a test item as a function of response signal lag from Rotello and Heit (2000, Exp. 1), along with predictions from our dynamic recognition model. Parameters used to generate these predictions are given in Table 4.4.

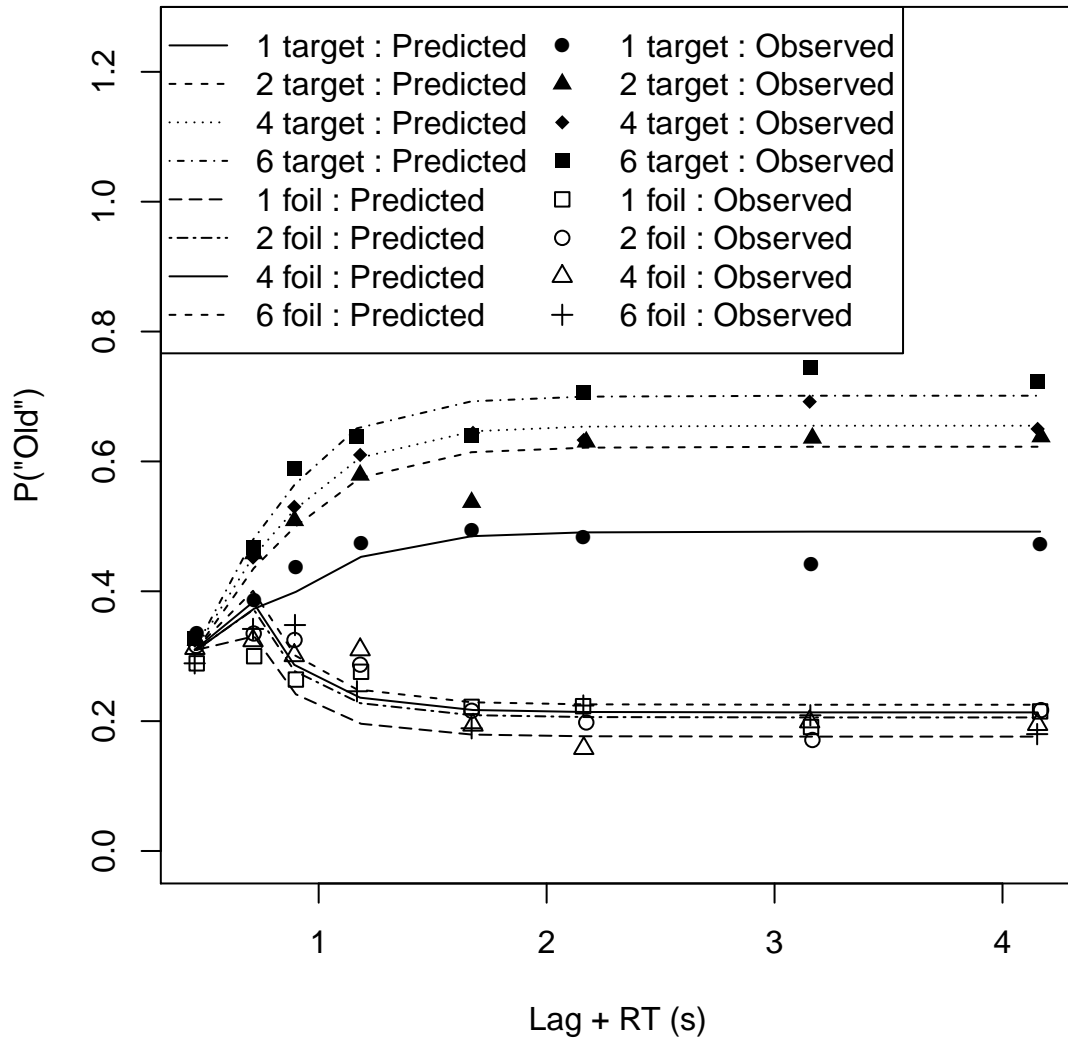
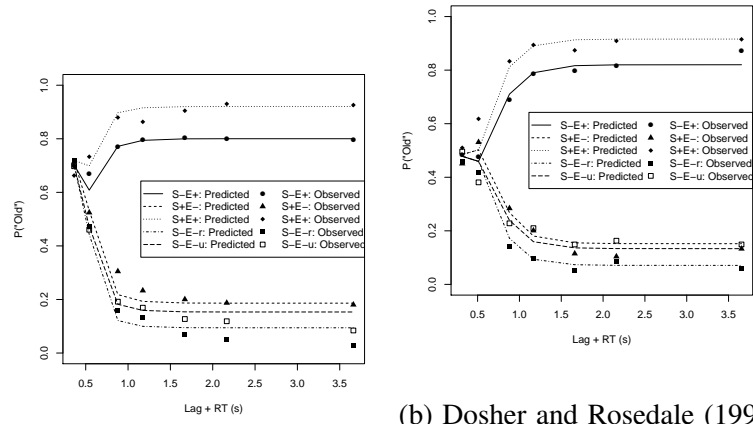


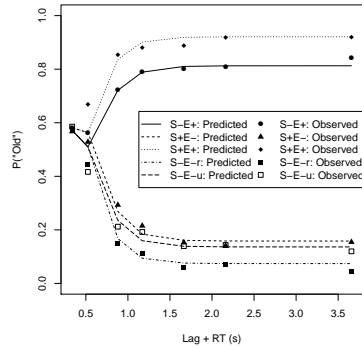
Figure 4.8: Probability of endorsing a test item as a function of response signal lag and response time from Doshier (1984a, Exp. 2), along with predictions from our dynamic recognition model. Parameters used to generate these predictions are given in Table 4.4.

had to distinguish between intact and rearranged pairs, such that item information alone would never be sufficient. Participants in Doshier (1984a, Exp. 2) studied pairs for varying amounts of time (1, 2, 4, or 6 seconds per pair) and foils were created by rearranging pairs within each level of study time. This experiment thus tests for the signature of recall-like processing: recall should operate faster on items studied longer, enabling participants to accept targets and reject foils more quickly. As shown in Figure 4.8, this is not the case; Doshier (1984a) found that an increase in study time did not affect the rate of approach to asymptotic performance, as would be predicted by a recall-based process.

Instead, our compound-cue formation model fits all these datasets quite well, suggesting that this is a better description of the processes underlying associative recognition. In further support of our model, we note that parameter estimates are consistent with the task demands in each experiment. Where associative information is not needed (Gronlund & Ratcliff, 1989, Exp. 1), the probe is given over to mostly item features and associative features, to the extent that they are involved at all, are not available until quite late. When associative information is required, it becomes available at roughly the same time across all experiments—between 400 and 450 ms after stimulus onset. Finally, when item information is not at all informative (Doshier, 1984a, Exp. 2), participants delay the onset of information accumulation (t_0) by 100 ms relative to experiments in which item information is informative. This effectively reduces the influence of item features by diminishing the time between the onset of processing and the time at which associative features become available.



(a) Doshier (1984b, Exp. 1). Exp. 1).



(c) Data pooled across both experiments.

Figure 4.9: Predicted and observed probabilities of endorsing a test item as a function of response signal lag and response time from two associative recognition experiments that used both semantically-related and unrelated word pairs. S-E+: Semantically unrelated intact pair; S+E-: semantically related rearranged pair; S+E+: semantically related intact pair; S-E-r: pair formed by rearranging words that had been studied with a related word; S-E-u: pair formed by rearranging words that had been studied with unrelated words. Parameters used to generate these predictions are given in Table 4.5.

4.2.3 Preexperimental associations

Finally, we examine associative recognition when the members of the pair have an existing pre-experimental association. Doshier (1984b) and Doshier and Rosedale (1991) had participants study word pairs, some of which had an existing pre-experimental association (e.g., EVIL-SIN). At test, participants had to distinguish between intact and rearranged pairs in a response signal paradigm. Some episodically-related pairs were also semantically (pre-experimentally) related (S+E+) while others were not (S-E+). Similarly, rearranged pairs might (S+E-) or might not (S-E-) be semantically related. Of the unrelated rearranged pairs, some were formed by rearranging that were themselves unrelated (S-E-u) while others “broke” one of the studied semantically-related pairs (S-E-r). As shown in Figure 4.9, which depicts data and model fits from Doshier (1984b, Exp. 1) and Doshier and Rosedale (1991, Exp. 1), intact semantically-related pairs (S+E+) had a higher asymptotic hit rate than unrelated intact pairs (S-E+). Related rearranged pairs (S+E-) are initially more likely to be endorsed than unrelated rearranged pairs, although this difference disappears at later lags. Finally, rearranged pairs that “break” a studied semantic relationship (S-E-r) have a lower asymptotic false alarm rate than other rearranged pairs.

To explain these results within the associative recognition framework we have developed, we make two additional assumptions: First, we assume that when two semantically related words are paired, some of their item features match by virtue of this relationship; the proportion of matching item features between semantically-related word pairs is denoted s . Since associative features also match, one could view these matching item features as a kind of associative feature that is available earlier in the trial. Second, we assume that a pre-existing semantic association allows for more associative features to be stored at study;

Table 4.5: Model parameters used to fit data from Doshier (1984b, Exp. 1) and Doshier and Rosedale (1991, Exp. 1).

<i>Parameter</i>	Doshier (1984b, Exp. 1)	Doshier and Rosedale (1991, Exp. 1)	Pooled
u	0.518	0.530	0.510
u_{AS}	0.727	0.748	0.732
c_S	0.954	0.967	0.967
s	0.041	0.019	0.025
p_A	0.292	0.245	0.263
t_A	268.623	486.032	476.931
θ	-2.127	-0.573	-1.064
t_0	0.964	0.303	0.000
ρ	6.692	8.161	7.726
R^2	0.966	0.950	0.958

the probability of storing an associative feature when semantically related words are studied together is denoted u_{AS} , as opposed to item features or associative features of unrelated pairs which are stored with probability u . The first assumption addresses the fact that S+E- and S+E+ pairs are more likely to be endorsed than unrelated pairs early in the trial; more matching features leads to a stronger match to the relevant memory traces. The second assumption deals with the lower asymptotic false alarm rate to S-E-r pairs relative to S-E-u pairs; more associative features in the traces of the rearranged words makes it easier to reject the S-E-r pair (both assumptions predict a higher hit rate for S+E+ pairs, of course). Note, however, that the difference between S-E-r and S-E-u pairs only manifests later in the trial, suggesting that semantically related pairs do not lead to greater storage overall, but especially for associative features which are only available later.

The parameters of our model used to fit the data from Doshier (1984b, Exp. 1) and

Dosher and Rosedale (1989, Exp. 1) are given in Table 4.5, with model predictions shown in Figure 4.9. Because the two experiments are near replications (in Dosher & Rosedale, 1989, study times were 1 s longer and blocks of recognition were intermixed with blocks of relatedness judgments, but other features of the design are identical), we also fit our model to data pooled across both experiments to increase power (there were four participants in Dosher, 1984b, Exp. 1 and five in Dosher & Rosedale, 1991, Exp. 1). Across both experiments, the estimated proportion of shared item features between semantic associates (s) is relatively small, but such pairs gain a large bonus to the proportion of associative features that can be stored ($u_{AS} > u$). In both experiments, sampling of item features begins almost immediately ($t_0 < 1$ ms) after stimulus onset, in contrast with previous fits in which the onset of sampling was slightly delayed. Oddly, when fitting just the data from Dosher (1984b, Exp. 1), the onset of the availability of associative features is abnormally fast relative to Dosher and Rosedale (1991, Exp. 1) and other associative recognition studies, but given that the estimate of t_A for the pooled data is closer to that of other experiments, this deviation is likely due to noise. In all cases, however, the model closely matches the observed SAT functions.

We note that our explanation of the nonmonotonic false alarm curve for S+E- pairs is different from the suppression mechanism suggested by Dosher (1984b) and Dosher and Rosedale (1991). They proposed a process whereby semantically related pairs were initially judged as old because participants initially failed to use the current context to focus on only recently learned associations. In other words, they argue that early responses to related pairs are really “relatedness” judgments and only later, when the current context is added to the memory probe, are participants actually engaging in an episodic judgment. According to our explanation, participants are always making an episodic judgment, since

all context features are present from the start of the trial, it is just that the presence of a semantic relationship influences how the pair is encoded. This is similar to our explanation for word frequency effects, where the effect results not from interference from traces stored in other contexts but from how prior experience influences episodic encoding. One advantage of our explanation is that a suppression mechanism would have to operate differently for targets (S+E+) and foils (S+E-), since hit rates for semantically related pairs are consistently higher than those for unrelated pairs (S-E+). Our explanation is also able to account for the lower false alarms to S-E-r pairs, which has no clear basis under a suppression account. Our assumption of enhanced storage of associative features for semantically related pairs explains both results—the consistent hit rate advantage for S+E+ pairs and the lower false alarm rate to S-E-r pairs.

4.2.4 Discussion

In this section, we have extended the dynamic model for recognition of single items to the recognition of pairs and showed how it can account for numerous results in both pair and associative recognition. Pair recognition results from the parallel operation of two single-item recognition processes in which changes in familiarity are accumulated as features of each item are sampled. If the accumulated changes for both items is sufficiently large, the pair is judged “old”, otherwise it is called “new”. In associative recognition, an additional set of features becomes available later in the processing of the two items; these features represent an associative binding between the two items. Our model can thus be considered a form of compound cue formation (Doshier & Rosedale, 1989, 1997).

The consistency of parameter estimates across experiments is remarkable and, coupled with the good quantitative fits, suggests that this model is capturing something important

about associative recognition. In particular, associative features, regardless of other aspects of the experiment, are generally available 400-500 ms after stimulus onset, which is intriguing because it lies on the border between the two temporal regions supposed to reflect “familiarity” (typically 300-500 ms) and “recollective” (typically 500-800 ms) processing in event-related EEG (Rugg & Curran, 2007). Even if associative features become available 400-500 ms after stimulus onset, on some trials they will all be sampled quickly and fall into the early “familiarity” bin, while on other trials they will be sampled more slowly and affect the later “recollective” ERP components, and this potential confusion is reflected in the ERP literature on associative recognition in which associative recognition performance sometimes preferentially related to late positive ERPs (D. I. Donaldson & Rugg, 1998) and sometimes with both early negative and late positive ERPs (Speer & Curran, 2007). Indeed, when test pairs contain a pre-existing (e.g., semantic) relationship—which we have assumed leads to shared features between the two probes from the beginning of sampling—those pairs exhibit different ERPs at early time windows (up to about 600 ms) but not later, consistent with our model that additional (experimental) associative features become available at the same time (t_A) for both related and unrelated pairs (Greve, van Rossum, & Donaldson, 2007).

4.3 List discrimination

In most recognition experiments, the recognition test occurs shortly after study, perhaps with a distractor task intervening between them. In such cases, the test context and study context are closely aligned, justifying our assumption that participants make recognition decisions by tracking changes in familiarity from a starting point determined primarily by context features. In a list discrimination paradigm, however, participants are only told to

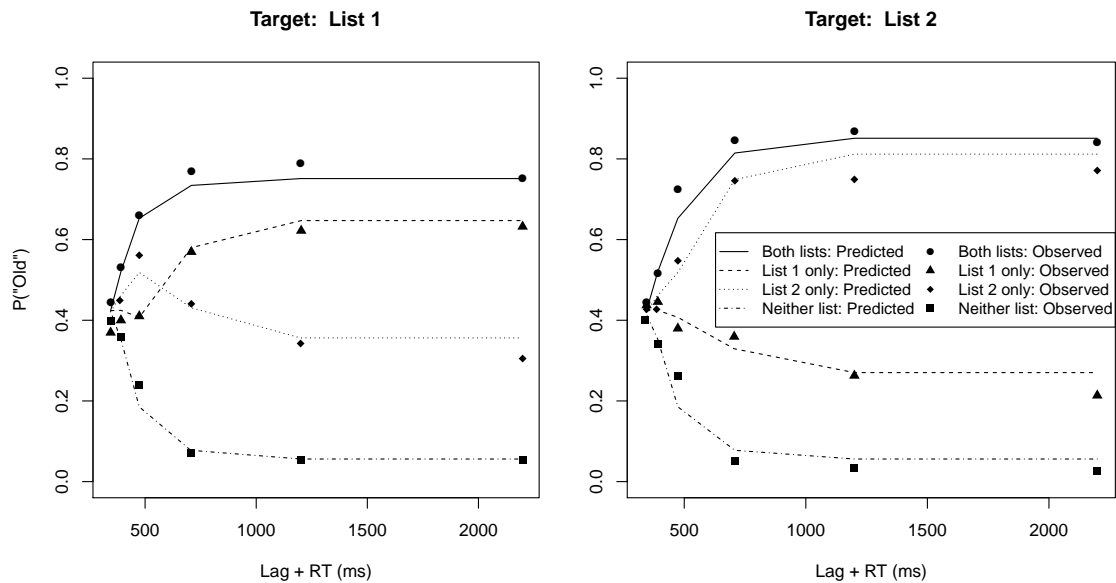


Figure 4.10: Data and model predictions for Hintzman et al. (1998, Exp. 1). Parameters used to generate model predictions are given in Table 4.6.

endorse an item if it appeared on a particular study list that may not be the most recent one. For example, Hintzman et al. (1998, Exp. 1) present results from a response signal study in which participants studied a series of word lists. At test, participants were told to respond “old” to items from either the most recent list (“List 2”) or the list prior to the current one (“List 1”). Thus, when List 2 is the target list, this is similar to the standard recognition situation; still, when participants were told to endorse only List 2 items, they would nonetheless produce false alarms to List 1 items. This false alarm curve, however, remains monotonic, as shown in Figure 4.10. However, when participants were to endorse only List 1 items—a departure from the usual recognition paradigm—they would initially false alarm to List 2 items and then reverse this tendency at longer lags.

This nonmonotonic false alarm curve provides evidence that a secondary process is involved when the target context is not the current one. One characterization of this secondary

process is a form of recall: The test item is used to cue recall of the context in which the item had been studied. If the retrieved context matches the target context—which is not the current context—then it is called “old”; otherwise, if the retrieved context does not match the target context, the item is rejected as “new”. Because this recall process takes longer to operate than the parallel matching process between the probe (with the current context) and memory traces, List 2 foils will initially and erroneously be called “old” before recall can be used to correctly reject such items.

An alternative account of this secondary process is a form of contextual reinstatement. The context used to probe memory is, initially, the current context, leading to a high false alarm rate to recent List 2 items. Over time, the context used to probe memory is changed as participants reinstate the target context. This reinstated context will grow to match the target context more and the current context less, thus producing a reversal in false alarms at later signal lags. A contextual reinstatement mechanism has received strong evidentiary support in both free and serial recall (Howard & Kahana, 2002; Jang & Huber, 2008), and has been implicated in recognition as well (Schwartz, Howard, Jing, & Kahana, 2005).

These two explanations differ in the driving force behind the observed changes in memory performance over time: Under the recall account, the change in false alarm rate over time is determined by recall of context from the trace of the test item. Under the contextual reinstatement account, changes are independent of the contents of memory and are driven entirely by the changes in the context cue used to probe memory. If the dynamics of list discrimination are affected by the strength of the trace of the test item, this would be evidence in favor of recall and against a pure contextual reinstatement account. This prediction was tested by Hintzman et al. (1998, Exp. 2), in which participants studied two lists of words where some words were studied once or twice on each list. Participants

subsequently engaged in one of two response signal tasks: a recognition test, in which participants were to endorse words from either studied list; or a list discrimination task which required participants to judge whether a word had been studied on either the first or second list. If list discrimination were based on recall, performance on the list discrimination task should rise above chance earlier for repeated words. In fact, there was no reliable difference between repeated and unrepeated words in terms of either recognition or list discrimination dynamics (both rose above chance performance at the same time and approached asymptotic performance at the same rate), failing to support the operation of recall in list discrimination.

We thus chose to model both experiments from Hintzman et al. (1998) by assuming that list discrimination depends on contextual reinstatement. We model contextual reinstatement in exactly the same way as we modeled a delay in plurality features and the delay of associative features, above. At time t_0 , the probe of memory consists of the N_X features of the current context and the N_C content features begin to be sampled. At a later time t_X , a certain number N_X^+ of additional context features becomes available and begin to be sampled into the probe. The original N_X context features will perfectly match the features of the most recent study list and partially match the features of the previous list (the features of List 1 match those of the current context with probability s_X), leading to an early tendency to endorse words from both lists. The N_X^+ additional features match *only* the target list context, thus “focusing” the set of activated traces on just those from the target context by simultaneously enhancing the match to the target context and decreasing the match to the non-target context, thus allowing participants to correctly reject words from the non-target list.

We note two important details of our approach to modeling contextual reinstatement:

Table 4.6: Model parameters used to fit data from Hintzman et al. (1998).

<i>Parameter</i>	Experiment 1	Experiment 2
u_1	0.403	0.401
u_2	—	0.514
c_S	0.959	0.937
s_X	0.966	0.970
N_X^+	7.774	8.438
t_X	370.024	416.258
$\theta_{\text{Recognition}}$	0.283	-2.079
$\theta_{\text{List discrim.}}$	—	-0.228
t_0	93.803	0.000
ρ	3.507	7.294
R^2	0.963	0.843

First, contextual reinstatement (the sampling of additional target context features) occurs whenever list discrimination is required, even if the target list is the most recent list. This reflects the idea that the “ambient” context features initially present in the probe are likely to overlap considerably between lists, given that they are experienced near in time and in the same location. Second, the N_X^+ additional context features are, effectively, added to the “pool” of N_C content features available for sampling at time t_X . Arguably, contextual reinstatement might occur independently of the sampling of content features, but we believe our approach is consistent with the idea that contextual reinstatement requires cognitive resources that might interfere with the ability to form a complete memory probe.

The fit of this model to the data from Hintzman et al. (1998, Exp. 1) is shown in Figure 4.10, with model parameters given in Table 4.6. Our proposed contextual reinstatement mechanism correctly reproduces the nonmonotonic false alarm curve found for List 2 words

(the most recent list) when the target is List 1, as well as the delayed rise in the hit curve to List 1 targets. The mechanism does so without sacrificing the ability of the model to correctly fit the curves for words repeated across lists or completely novel words.

In Hintzman et al. (1998, Exp. 2), participants had to make not just recognition judgments, but list discrimination judgments as well. Our contextual reinstatement mechanism suggests a way that participants could make list discrimination judgments by tracking changes in familiarity, in the same way that participants perform a recognition task. If instead of tracking changes from time t_0 , they tracked changes from time t_X , participants could use the change in familiarity over time as features of the target context entered the probe to gauge the degree to which a word matched that target context, such that the evidence for list discrimination $x(t)$ is given by

$$x(t) = \sum_{\tau=t_X+1}^t \phi(\tau) - \phi(\tau-1) = \phi(t) - \phi(t_X). \quad (4.5)$$

As features of the target context are sampled, they will tend to increase the match to a trace from that context and decrease the match to a trace encoded in a different context, so the accumulated change in familiarity can be used in exactly the same way as in normal recognition, only with a different start point. To maximize efficiency and minimize interference from non-target traces, the starting point for contextual reinstatement t_X should be sufficiently late that enough content features have been sampled to identify the relevant memory trace. Note, again, that this mechanism predicts no difference in retrieval dynamics as a function of item strength because sampling is driven by contextual reinstatement, which is independent of trace strength; strength only affects asymptotic performance under the assumption that additional storage time allows for more features to be stored.

We fit this model to both the recognition and list discrimination tasks from Hintzman et

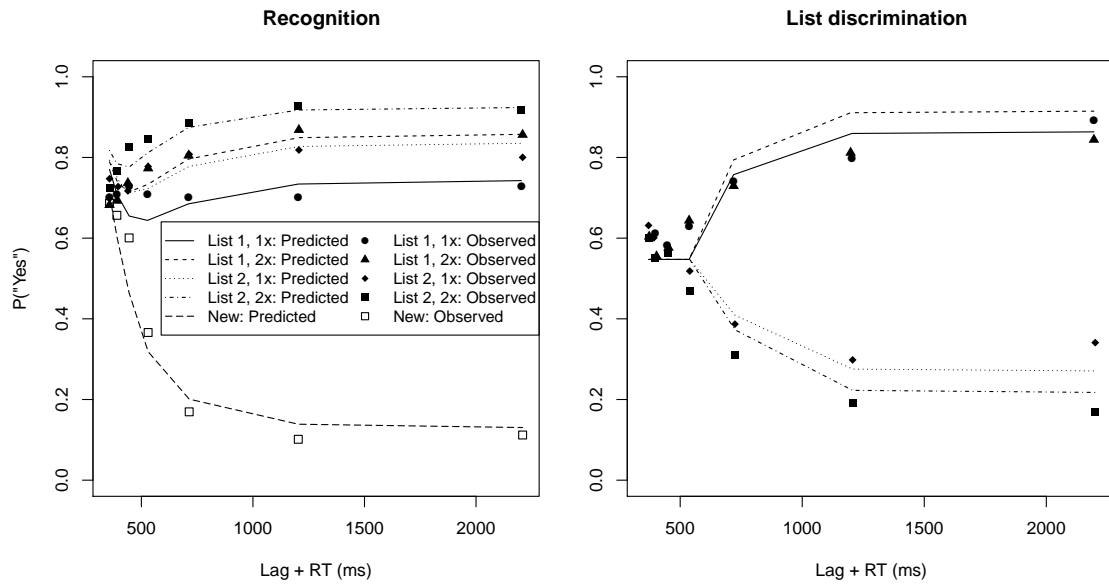


Figure 4.11: Data and model predictions for Hintzman et al. (1998, Exp. 2). In recognition, positive “yes” responses should be given to any studied item, regardless of list. In list discrimination, “yes” responses should be given only to words studied in List 1. Parameters used to generate model predictions are given in Table 4.6.

al. (1998, Exp. 2), as shown in Figure 4.11, with parameters given in Table 4.6. Participants were not told until test what kind of task they would be engaged in, so the storage parameters (u_1 , u_2 for the two levels of repetition, c_S , and s_X) were identical across tasks. We also assumed for simplicity that feature sampling dynamics (t_0 and ρ) were identical across the two tasks, although a better fit could undoubtedly be obtained by relaxing this assumption. Besides different criteria ($\theta_{\text{Recognition}}$ and $\theta_{\text{List discrim.}}$), the key difference between the two tasks is that contextual reinstatement only occurs in list discrimination. In recognition, the N_X features of the current context are the only context features that are present in the probe, and they are present throughout. Although these features perfectly match the storage context of List 2, they only match the context features of List 1 with probability s_X . Sampling of content features begins almost immediately after stimulus onset (t_0), but list discrimination does not begin until additional context features become available at time t_X . At this time—comparable to t_X in Experiment 1—the List 1 context is partially reinstated as N_X^+ features from that context begin to be sampled into the probe. This time corresponds to roughly 57 model “time-steps”, at which point there is an 86% chance that all content features have been sampled (Equation 2.1), consistent with the idea that list discrimination must wait until the target trace is effectively isolated.

At this point, a reasonable question to ask is why, if participants know the target context at the beginning of the trial (as in Hintzman et al., 1998, Exp. 1), do they not begin probing memory with the features of the target context? Why must reinstatement occur during the trial, rather than before? Doshier (1984b, Exp. 3) provides evidence that, even when participants could have used a simple rule to reject items from an irrelevant preexperimental context, they nonetheless fail to do so until later in retrieval. In Doshier (1984b, Exp. 1) and Doshier and Rosedale (1991), discussed above, both the study list and the test list

contained semantically related word pairs, so the presence of a preexperimental (semantic) association could not be used to diagnose whether a pair had been studied or not. In experiment 3 of Doshier (1984b), however, the study list contained no semantically related items; thus, the presence of a preexperimental association could, in principle, be used to immediately reject such pairs. Instead, three out of five participants showed the same pattern as Exp. 1, with initially rising, then falling false alarm rate to unstudied semantically related word pairs; the remaining two participants showed instead a longer non-decision time to semantically related pairs, effectively “smothering” the false alarms the other participants committed during the same time interval. Thus, it would appear that, at least under normal testing conditions, the initial context features present in participants’ memory probe are not under any kind of strategic control. When traces from outside the target context must be excluded, participants must engage in a process of contextual reinstatement that is delayed relative to the time at which content features begin to be sampled from the test item. Perhaps the need to expend resources maintaining context features that are not already present in the environment results in participants not maintaining them during the between-trial interval.

Finally, we note that our approach to modeling contextual reinstatement implies that participants can, in principle, store more than N_X context features in a memory trace. This does not alter the behavior of the model in any other setting, however, because what matters is that, without additional effort to reinstate the context, those features will never enter the memory probe and contribute to the likelihood ratio for the trace. When the current “ambient” context features are sufficient to pick out the target context, as they are in standard study-test recognition paradigms, only the original N_X context features are presumed to play a role. Consistent with this account, correct free response time distributions after a

four month retention interval—presumably entailing contextual reinstatement—are qualitatively similar to distributions after a ten minute interval (Reber, Alvarez, & Squire, 1997). Distributions after a longer delay are still unimodal, suggesting that a single retrieval process is involved, but are more spread out, as would occur if feature sampling were extended to allow for the reinstatement of target context features.

4.4 Source memory

We have thus far argued that many kinds of retrieval functions are best understood as reflecting the availability of different kinds of information at different times, rather than as a signal of recollection. The key claim we are making is that, to have evidence of recollection, retrieval dynamics must be affected by the strength of the trace that is being recalled. If retrieval dynamics are independent of trace strength, then the result is better explained by the dynamics of probe formation and feature sampling. In contrast to similar foils, associative recognition, and list discrimination, here we review a situation in which trace strength does affect retrieval dynamics, thus suggesting the involvement of an explicit recollection process. We develop a model of this process and show how it can be incorporated into our dynamic framework. The domain in which we find evidence for the involvement of recollection is source memory.

A source memory task requires participants to judge not just whether an item is present in memory or not, but to make a decision about the situation in which it was originally encountered—the “source” of the item. Thus, in source memory it is typically already known or assumed that a trace exists in memory corresponding to the test item, and the task requires extracting some additional information about the study event that is not present in the probe. Logically, then, it is no surprise that source memory typically requires the

operation of additional mnemonic processes beyond an assessment of familiarity (although we will discuss below situations in which a dynamic model of familiarity alone could still support source memory).

In their second experiment, Hintzman and Caulton (1997) had participants study a list of words, some of which were presented auditorily and some of which were presented visually. Afterwards, participants engaged in either a standard SAT recognition task or a modality judgment task, although they were not informed of the test format until after study, thus precluding any idiosyncratic study strategies. Test words were always presented visually, although, consistent with Mulligan and Hirshman (1995) modality match had only a minimal effect on recognition performance. Crucially for present purposes, however, when an item was repeated three times over the study list, not only did this improve asymptotic modality judgment performance, it also caused correct modality judgments to rise above chance more quickly (Figures 4.12 and 4.13), hence we consider their findings evidence that making a correct source judgment (in this case, modality) requires the recall of modality-related features from the study trace.

4.4.1 A model for recall dynamics

Our model of recall is based on a simplified search-and-recovery mechanism similar to SAM and to a variety of applications of similar models to recall (Raaijmakers & Shiffrin, 1981; Shiffrin & Steyvers, 1997, 1998; Diller et al., 2001; Malmberg, 2008). We assume that, as in our base model, features of the test item are gradually sampled into a memory probe that contains both content and context features. As usual, this probe is compared in parallel to each trace in memory, producing a likelihood ratio $\lambda_i(t)$, as described above (Equation 2.6). To recall a trace and examine its contents, it must first be sampled. The

probability $p_{Si}(t)$ that a trace i will be sampled at time t is based on its relative likelihood:

$$p_{Si}(t) = \frac{\lambda_i(t)}{\sum_{j=1}^{N_{Act}(t)} \lambda_j(t)}, \quad (4.6)$$

where $N_{Act}(t)$ is the number of active traces (those with $\lambda_j(t) > 1$) at time t . If a trace is sampled, it must then be recovered into working memory. The probability that a sampled trace i will be recovered depends on how complete both the probe and trace are:

$$p_{Ri}(t) = 1 - \exp[-\beta_R v(t) u_i], \quad (4.7)$$

where $v(t)$ is the completeness of the probe at time t , given in Equation 2.1, u_i is the proportion of content features stored in trace i , and β_R reflects the rate of recovery. As β_R increases, the probability of successfully recovering the trace increases.

To make a modality judgment, a participant must “read out” the modality features from the recovered trace. We assume that modality features are like other features in that they may or may not be stored and, if stored, may be stored incorrectly. For simplicity—and because the modality can be either visual or auditory—we assume that modality is represented by a single binary feature. Further, because modality may be particularly salient and because participants know they may later be tested on it, we assume that the modality feature might be stored at a greater rate than other kinds of features. To model this, we introduce a single scaling parameter m such that the probability of storing the modality feature for trace i is u_i^m . Thus, if $m = 1$, modality features are stored at the same rate as other features, but if $m < 1$, modality has a greater chance of being stored. The probability of giving an “auditory” source judgment to a word i that had, in fact, been presented auditorily is, then

$$p_{AA} = u_i^m [c_S + (1 - c_S)\frac{1}{2}] + (1 - u_i^m)b, \quad (4.8)$$

where b is an overall bias to say “auditory” in case the modality feature failed to be stored. Alternatively, the probability of incorrectly giving an “auditory” source judgment to a word i that had actually been presented visually is

$$p_{VA} = u_i^M (1 - c_S) \frac{1}{2} + (1 - u_i^m) b. \quad (4.9)$$

Notice that, in these equations, we assume that if the auditory feature is stored incorrectly (with probability $1 - c_S$), it is simply given a random equiprobable value (either visual or auditory).

The probability of giving an “auditory” judgment at time t depends on which trace has been sampled, whether it has been successfully recovered, whether the modality feature has been stored, and if so, whether it has been stored correctly. Because features continue to accumulate in the probe over time, the probability of sampling the correct (target) trace also increases over time. Furthermore, because a recovery attempt is made on each time step, the probability of successful recovery of a sampled trace also increases over time, beyond the increase in recovery probability that results as more features are sampled into the probe (i.e., as $v(t) \rightarrow 1$)³. Accuracy is limited, however, by the fact that the modality feature may not have been stored (with probability $1 - u_i^m$ for a given trace i) and even if it has, it may have been stored incorrectly (with probability $\frac{1}{2}(1 - c_S)$).

Crucially, however, the accuracy of modality judgments will evolve over time differently as a function of the strength of the target trace. This owes to the fact that both sampling and recovery depend on the strength of the trace, not just the features in the probe—the more features stored in the trace, the more quickly its likelihood ratio will grow relative

³Notice that, in the simplified model presented here, we assume that participants never “give up” after a recovery failure, although a mechanism that allows participants to “give up” would certainly be necessary in a more elaborate model of recall (e.g., Raaijmakers & Shiffrin, 1981; Diller et al., 2001).

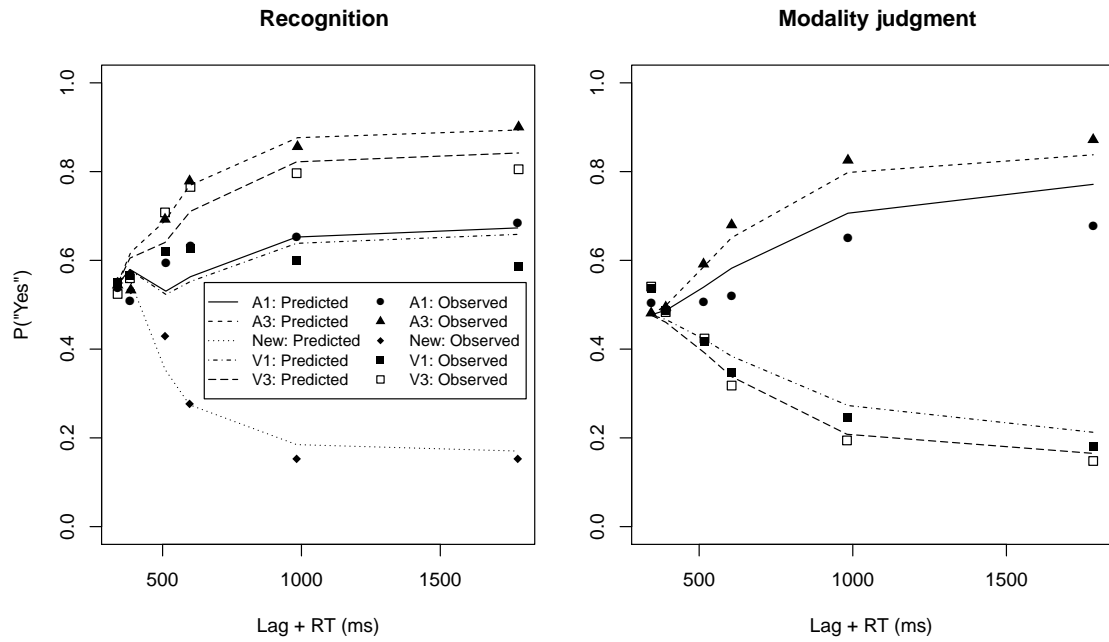


Figure 4.12: Probability of giving a positive response in the either the recognition or modality judgment tasks from Hintzman and Caulton (1997, Exp. 2). In recognition a “yes” response is given to studied items regardless of modality while in modality judgment, a “yes” response should be given to a word that had been studied auditorily. Conditions are labeled by modality (A for auditory, V for visual) and by the number of repetitions at study (1 or 3 times). Parameters used to generate model fits are given in Table 4.7.

Table 4.7: Parameters used to generate predictions for Hintzman and Caulton (1997, Exp. 2 and 3). Conditions are labeled by modality (A for auditory, V for visual) and by the number of repetitions at study (1 or 3 times).

<i>Parameter</i>	Exp. 2	Exp. 3
u_{A1}	0.287	0.341
u_{A3}	0.439	0.513
u_{V1}	0.279	0.167
u_{V3}	0.391	0.398
c_S	0.981	0.986
θ	0.558	—
m	0.424	0.515
b	0.477	0.440
β_R	0.073	0.061
t_0	78.575	145.657
ρ	7.554	5.343

to other traces and the more likely it is to be recovered on a given sampling attempt. In addition to varying study modality, Hintzman and Caulton (1997) also varied item strength by presenting words either once or three times within a particular modality. They found that, although the dynamic of episodic recognition were unaffected by strength or study modality (i.e., retrieval functions rose above chance at the same time and approached asymptote at the same rate), modality judgments rose above chance earlier for repeated items. Thus, our standard recognition model was able to fit the recognition conditions of their Experiment 2, but we assumed that recall was used to make recognition judgments, albeit operating on the same underlying memory structures. We fit separate values of u for the four types of item (auditory or visual and once- or thrice-presented), as well as values for c_S , t_0 , and p ; these applied to both the recognition and modality judgment conditions. Fitting the recognition condition required a single additional criterion parameter θ , while the modality judgments entailed a bias b , encoding factor m , and recovery rate β_R as described above. Fitted values of these parameters are given in Table 4.7 with model fits shown in Figure 4.12. As expected, u is higher for repeated than non-repeated items, reflecting greater storage of features with additional study. Storage probability u is also higher for auditorily presented (A) items than visually presented (V) items. The modality feature does, indeed, appear to be stored at a greater rate than other features, however, as reflected by $m < 1$.

We also fit our modality judgment model to Hintzman and Caulton (1997, Exp. 3), which did not include a response signal recognition task (recognition was still tested to ensure that participants were not using idiosyncratic strategies). Model fit, shown in Figure 4.13, was obtained using parameters very similar to those used to fit Experiment 2 (see Table 4.7). Once again, the modality feature is encoded at a greater rate than other features ($m < 1$), although the difference between auditory and visual words in overall feature stor-

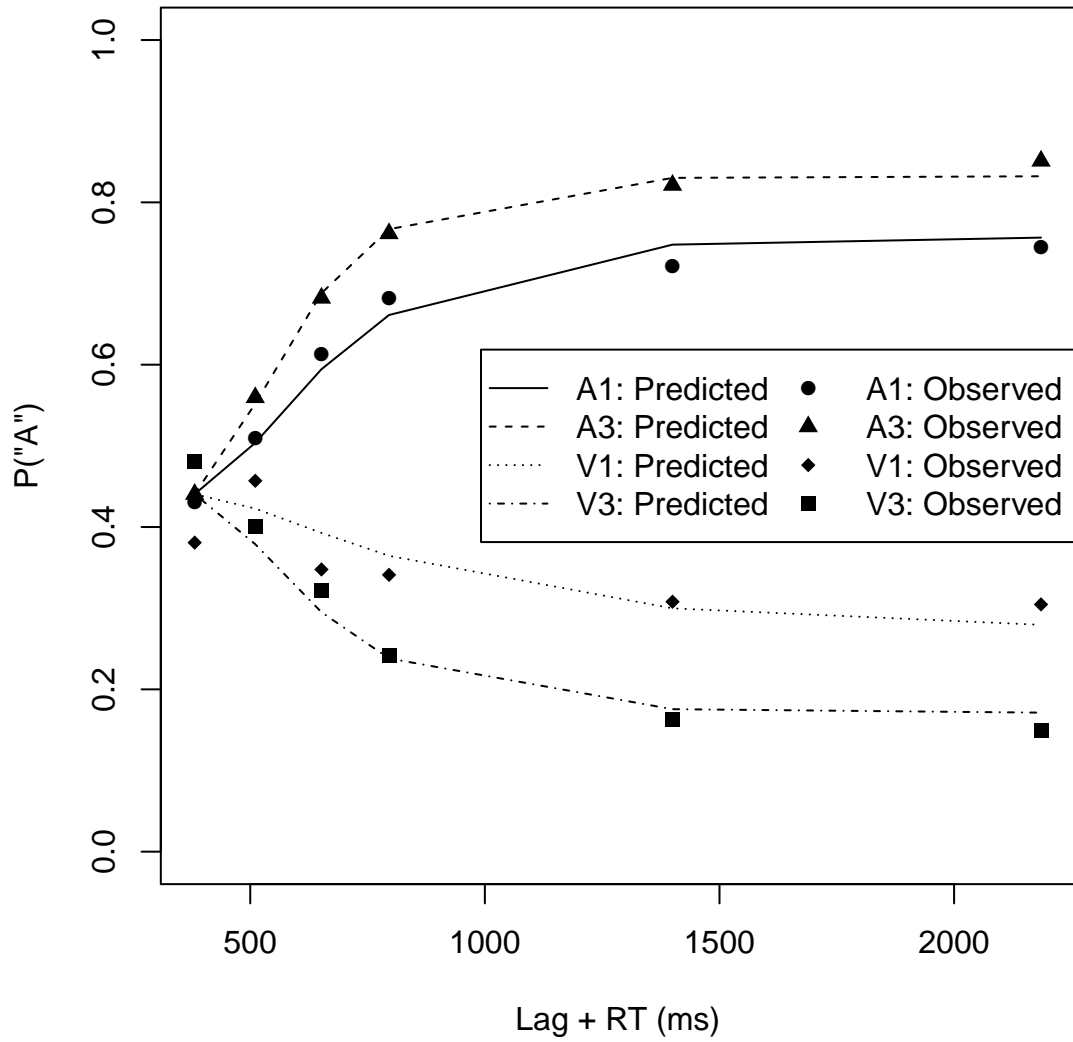


Figure 4.13: Probability of judging a word as having been presented auditorily in Hintzman and Caulton (1997, Exp. 3). Conditions are labeled by modality (A for auditory, V for visual) and by the number of repetitions at study (1 or 3 times). Parameters used to generate model fits are given in Table 4.7.

age probability u appears larger in Experiment 3. The only other notable difference is that participants in Experiment 3 appear to take longer to initiate feature sampling (t_0), which can be attributed to the longer lags used in Experiment 3, since participants would not feel as compelled to begin processing before very short lags.

4.4.2 Source memory vs. source exclusion

Our claim that source memory requires recall is consistent with the results showing that different regions of the medial temporal lobes are correlated with recognition confidence than with source judgments (Ranganath et al., 2003), although these regions partially overlap. This overlap, coupled with the correlation in performance between source judgments and recognition (Starns et al., 2014), reinforces the fact that the recall process used to make a source judgment operates on the same memory traces and relies on the same process of probe formation as in recognition. Unlike source memory—in which participants must retrieve information about source from a memory trace—source *exclusion* offers no evidence for the operation of recall. The participants in McElree, Dolan, and Jacoby (1999) studied two lists of words, one of which was presented auditorily and another was presented visually. In the subsequent recognition test, based on the exclusion procedure of Jacoby (1999), participants were to respond positively only to the heard words, giving negative responses to words that had been presented visually. Some of the visual words were presented multiple times (3 times in Experiment 1, 5 times in Experiment 2). Early in the trial, participants tended to false alarm to visually presented items, but this tendency reversed later, with a lower asymptotic false alarm rate to repeated visual words than to those that had been presented only once. Critically, however, the inflection point at which the false alarm curve began to reverse occurred at the same time for both once-presented and multiply-presented

visual words, and the rate at which these words approached their final asymptotes was the same. If recall were operating, as it appears to be in Hintzman and Caulton’s source memory task, multiply presented visual foils should show an earlier inflection point in their false alarm curves, reflecting the greater probability of recovering those traces.

Instead, these recognition results resemble those found for reversed-plurality foils in the exclusion condition of Hintzman and Curran (1994, Exp. 3). Just as in that previous experiment, we can best explain these results by assuming that modality features do not join the probe until later in the process of recognition, and that additional study time allows for greater storage of such features. We fit our dynamic model to the data from McElree et al. (1999, Exp. 2), shown in Figure 4.14, with a different value of u for each item type (auditory, once-presented visual, and five-times-presented visual), where a certain proportion p_M of features contain modality information which are not available for sampling into the probe until time t_M . Above, we made the simplifying assumption that a single modality feature was sufficient to make a modality judgment, but here we acknowledge that many features in a memory trace might give information about the modality in which the word had been presented (for example, a prosodic feature would probably only be encoded if a word had been heard rather than read). Thus, the p_M modality-related features might not be used solely to encode modality *per se*, but some other feature related to modality. As above, we assume that modality-related features may be stored at a different rate than other content features (like semantics), as represented by the parameter m ; the probability of storing a modality-related feature is u^m . Finally, we also fit values of c_S , θ , t_0 , and ρ .

The parameters used to generate predictions from the model are given in Table 4.8. Note that, in this application of the model, we assume that more than one feature is used to represent modality, and that modality-related features comprise a sizable proportion (p_M)

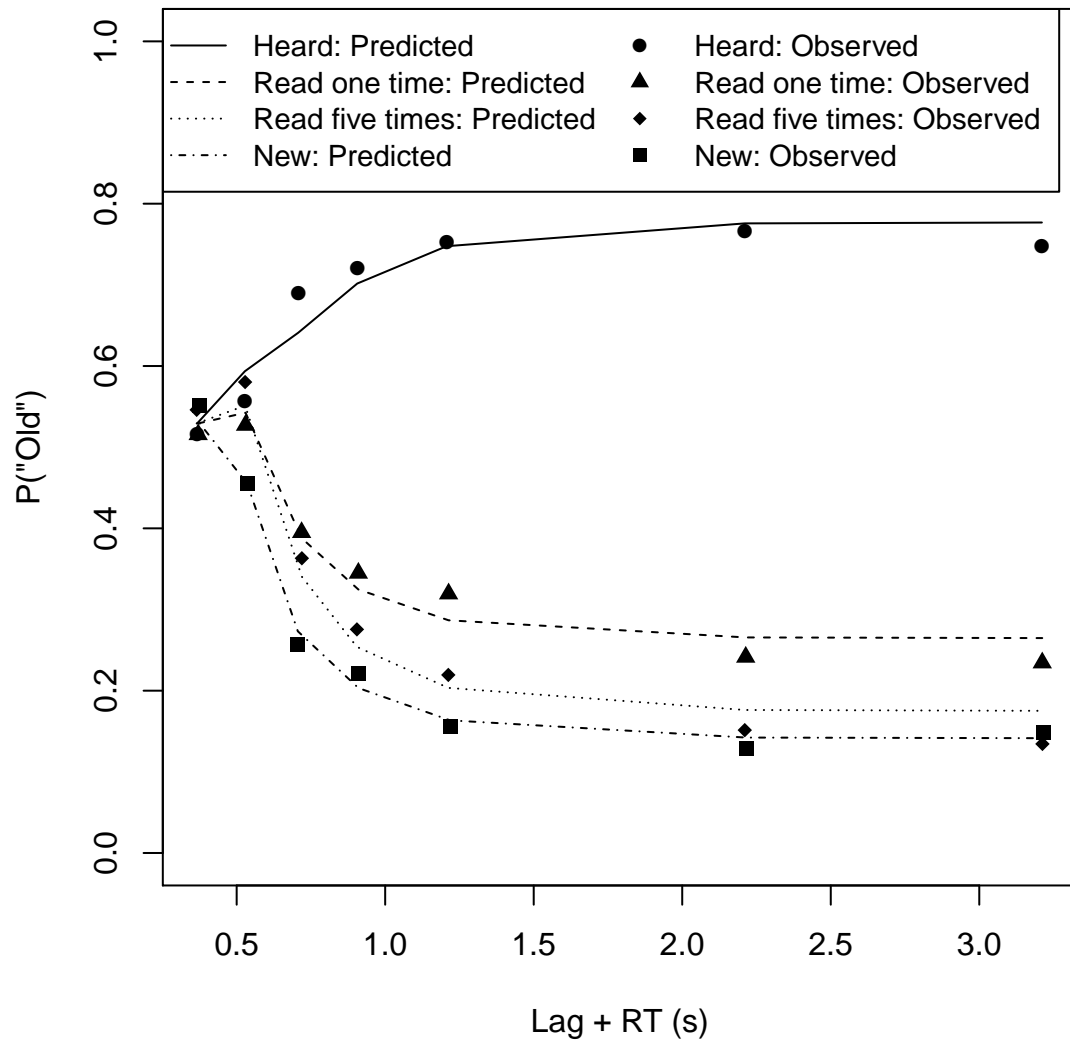


Figure 4.14: Probability of endorsing a word as a function of response signal lag and response time in McElree et al. (1999, Exp. 2). Parameters used to generate model predictions are given in Table 4.8.

Table 4.8: Parameters used to generate predictions for McElree et al. (1999, Exp. 2). Conditions are labeled by modality (A for auditory, V for visual) and by the number of repetitions at study (1 or 5 times).

<i>Parameter</i>	<i>Value</i>
u_A	0.483
u_{V1}	0.604
u_{V5}	0.987
c_S	0.972
p_M	0.466
t_M	182.986
θ	-0.117
m	1.801
t_0	87.229
ρ	10.444
R^2	0.973

of the content features. The rate at which modality-specific features are stored is, here, estimated to be less than the rate at which non-modality-related features are stored ($m > 1$), but given that the number of features that carry information about modality (p_M) is quite high, this is likely a case of diminishing returns. As reflected in the initial bias to respond “old” to visual words, their content features (u_{V1} and u_{V5}) are stored quite well; additional time also allows for greater storage of modality features, which allows repeated visual words to be correctly rejected at a greater rate once modality features become available. The time at which they become available (t_M) is only about 100 ms after feature sampling begins (t_0), suggesting that modality-related information can be brought to bear more quickly than associative, plurality, or context-related features.

The delayed availability of modality features may, as with plurality features, reflect the automatic deployment of attention to semantic features of a word first, followed later by perceptual aspects of the word (as in the Stroop effect). The later availability of modality features could also be strategic, since source-related differences in ERP only appear when source is relevant to the recognition decision (Senkfor & Van Petten, 1998), although this could depend on how “source” is defined (e.g., in terms of modality, speaker, location, color, etc.; Mollison & Curran, 2012). An interesting alternative possibility is that the delay results from the use of visual testing: because auditory features are not present in the percept of the test item itself, participants must supply them if they are to serve as a memory cue, as required by the exclusion instructions. This process is analogous to contextual reinstatement required in the list discrimination task—only the features of the current context are immediately available, while the participant must reinstate the target context themselves. Note that, if this hypothesis is correct, then matching the test and target modalities by instructing participants to respond “old” only to *visually* presented

items should attenuate or eliminate the nonmonotonicity in false alarm curves to words studied in the non-target modality because the target modality features would be available from the beginning of the trial. To our knowledge, this critical control has not yet been run.

4.5 Discussion

Throughout this section, we have explored the notion that recognition decisions involve multiple processes. We have shown that, in many cases, a secondary source of information must be invoked to explain certain kinds of recognition, like the rejection of certain kinds of foils, associative recognition, list discrimination, and source memory. We have formalized this secondary information and its evolution over time within our dynamic model for recognition, showing how our modeling framework can be extended to explore and understand new phenomena. In particular, we have argued that although the situations reviewed above might be considered “dual-process” explanations, the secondary process involved is not necessarily best described as “recollection”, the term most often identified with dual-process theory. Instead, a variety of processes seem to be involved, with only source memory providing any strong evidence for the involvement of a qualitatively different recall process, for which we also presented a preliminary model. In light of our model, it would appear that memory retrieval can involve a multitude of processes, operating concurrently, any of which might be used to support a particular memory decision.

Although we have been concerned primarily with behavioral data throughout this section, researchers investigating recognition memory via EEG have traditionally adopted a dual process perspective, with a “familiarity” component associated with an early frontal negativity peaking around 400 ms after stimulus onset (“FN400”; Curran, 1999) and a later “recollection” component associated with greater positivity in parietal areas between 400

and 800 ms after stimulus onset (see, e.g., Curran, 2000; Rugg & Curran, 2007). The rationale behind identifying these two components with processes of familiarity and recollection comes from two sources: First, the later parietal positivity is observed to be greater when participants make “remember” judgments than when they make “know” judgments (Düzel, Yonelinas, Mangun, Heinze, & Tulving, 1997); second, the later parietal positivity is selectively associated with correct source memory judgments (Wilding & Rugg, 1997). Other evidence supporting the identification of separate ERP components reflecting familiarity and recollection processes comes from assuming that recollection is necessary to, for instance, discriminate between targets and foils that vary in plurality, as discussed above (Hintzman & Curran, 1994; Curran, 2000); in such cases, only the later parietal component differentiates between targets and similar foils.

As should be clear, the ERP evidence supporting separate familiarity and recollection processes is of essentially the same form as the behavioral evidence discussed thus far: Both are predicated on the distinction between a fast and slow process, with the slower process making use of information not available to the faster one. We do not dispute the distinction between these two time-scales, but we maintain that this difference is not sufficient to characterize fast and slow processes as “familiarity” and “recollection”. We have shown above that recollection is not necessarily the best way to characterize secondary processes in recognition as detected in behavioral paradigms. In some cases, changes in memory dynamics may be attributed to the later availability of certain information (as with plurality), the formation of a compound cue (in associative recognition), or the need to reinstate context (in list discrimination). The late parietal positivity observed in such cases may well reflect these other secondary processes, and not that of recollection. To decide this question, explicit quantitative modeling of the putative processes must be done. We

have done so at the behavioral level, and similar work would need to be done to model neural responses. With such work, we believe that models of recognition dynamics like the one we have proposed could be a boon to the interpretation of neural data. In particular, the ability of EEG to probe ongoing cognitive processes at fine time scales provides a great opportunity for future research to develop and test dynamic models of recognition.

Chapter 5

Discussion

We have introduced a dynamic model for recognition memory and shown that it not only fits data quantitatively, it offers new and deeper insights into the processes behind recognition. It goes beyond most existing models of memory by specifying how retrieval evolves over time, and it goes beyond extant models of decision making by explicitly modeling how memory evidence is generated. By using this model, it is possible to take a dynamic approach to understanding memory, which provides insight about how instructions (such as speed/accuracy trade-off) influence mnemonic processes and allows one to test specific hypotheses about the source of word frequency effects that would be indistinguishable if one treated memory statically. The model can also be extended to explore phenomena like associative recognition, list discrimination, and source memory, showing that phenomena in these domains can be explained by the differential availability of certain kinds of information at retrieval, whilst source memory judgments entail a secondary recall process. Taken together, these theoretical advances demonstrate the utility of a dynamic approach to recognition and, we hope, will spur further empirical and theoretical development. In this final section, we discuss several ways in which a dynamic approach to recognition could be extended beyond the data and modeling presented here.

5.1 Other aspects of memory

We have demonstrated how our model of recognition dynamics can be extended to various tasks besides single-item recognition, including pair and associative recognition, list discrimination, and source memory, but episodic memory supports many other tasks, including free recall, serial recall, judgments of frequency, and judgments of recency. Just as a dynamic approach sheds light on recognition, we believe such an approach will be fruitful in these other domains, and has already done so in many cases. The simplified cued recall model we proposed to explain source memory judgments was, in turn, based on a sampling-and-recovery model of recall (Raaijmakers & Shiffrin, 1981; Diller et al., 2001) that was itself motivated by the dynamics of inter-retrieval times and stopping times. The temporal context model (Howard & Kahana, 2002) has also been extended to account for the dynamics of recall using a competing accumulator approach similar to the one we have used to explain recognition (Sederberg, Howard, & Kahana, 2008; Polyn, Norman, & Kahana, 2009). These extensions revealed, among other things, the importance of semantic associations between successively recalled items predicting inter-response latencies, which would not necessarily have been apparent otherwise. The dynamics of recency and frequency judgments have not been explored as fully as those of recall or recognition, although the response signal method indicates that the dynamics of frequency judgments are similar to those of normal recognition (Hintzman & Curran, 1994). Although we have modeled repetitions as strengthening only the content features of an existing memory trace, dissociations between the effects of study time and number of repetitions on confidence ratings and judgments of frequency and recency suggest that repetitions may introduce additional “reminding” features (Hintzman, 2004, 2010). These features could be a “tag” intention-

ally stored by the participant, or might be reflect the different temporal context in which the various repetitions occurred (e.g., Hintzman, 2002). Given the similar dynamics between frequency judgments and recognition, whatever new features are introduced might act like other content features and be available from the start of feature sampling. Curiously, neural evidence indicates that different processes occur when an item is repeated at study than when it is shown at test (which is, after all, another kind of repetition; de Zubicaray, McMahon, Eastburn, Finnigan, & Humphreys, 2005); this indicates that although recollection and strengthening—perhaps with additional “reminding” features—occurs at study, it does not do so at test. This is consistent with a model in which reminding features are just another kind of content feature available for sampling at test, and that otherwise recognition and frequency judgment are not qualitatively different, although more empirical work would be required to put any such theory on firm footing.

5.1.1 Short-term recognition and lag

Thus far we have not examined the effect of lag—the delay between study and test—within the context of our model. Lag has been explored extensively within the Sternberg (1966) paradigm, in which participants study a small set of items (typically within the span of working memory) immediately before a recognition probe is presented, but is also present in longer study-test paradigms of the kind we have addressed above. Although it was originally argued that a serial search process was necessary to explain these results (Sternberg, 1966; Atkinson & Juola, 1974), a variety of parallel matching models provide excellent fits to this paradigm as well (Ratcliff, 1978; Hockley & Murdock, 1987; Nosofsky et al., 2011; Donkin & Nosofsky, 2012b; Nosofsky, Cox, et al., 2014). At least at lags greater than one, there is no evidence for difference in retrieval dynamics, only for differences in asymptotic

performance, consistent with the idea that, for the most part, recognition even from short lists is based on a parallel matching process rather than serial search (McElree & Doshier, 1989).

To allow parallel matching models—a class which includes our dynamic model—to model lag effects, it is necessary to assume some sort of decay in the strength of the trace as a function of the number of items intervening between the target trace and the probe. There is no a priori reason why trace decay cannot be incorporated into our model in order to account for lag effects in both short and long lists. The question is what form this decay will take: Decay could be modeled as a loss of content features over time (decreasing u), which would imply a kind of forgetting. Alternatively, decay could result from context drifting over time, such that the context features at the time of test better match the context stored with more recent traces. A contextual drift explanation would be consistent with the idea that temporal information forms an integral part of “context” (particularly in judgments of recency or list discrimination; see Dennis & Humphreys, 2001; Klein et al., 2007). Within a particular parallel matching model—the Exemplar-Based Random Walk (EBRW; Donkin & Nosofsky, 2012a; Nosofsky, Cox, et al., 2014)—strength decay was well-modeled by a power function, but we leave it for future work to determine how contextual drift should operate in our modeling framework.

5.1.2 Encoding

Memory theorists have tended to focus on retrieval from memory rather than storage, a tradition we follow here. Nonetheless, it is clear that a complete theory of memory must include an account of how memory traces are formed and stored. We argue that, just as with retrieval, understanding such processes requires a dynamic approach, taking into account

the dynamics of the study event both within the study trial and in the ongoing encoding context. A dynamic approach to encoding as well as retrieval could also reveal similarities between the two processes. For example, it has been argued that memory representations themselves contain a dynamic component (e.g., Cooper & Shepard, 1973; Freyd, 1987) and that retrieval involves a kind of “simulation” of the original experience (Barsalou, 1999; Kent & Lamberts, 2008). Indeed, neural recordings from mice reveal that hippocampal activity when choosing which path to follow in a maze resembles a temporally-compressed version of the hippocampal activity that occurs when the mice actually traverse the maze (Jadhav, Kemere, German, & Frank, 2012).

5.1.3 Extended events

Our model describes how familiarity evolves over time to make recognition decisions, but the span of time we have considered is rather brief, on the order of a few seconds at most and typically must less. This timescale is entailed by the kinds of stimuli used in the experiments we have modeled, namely, visually presented words or images. Such stimuli, once they appear, are static and are not extended in time. Of course, in addition to the dynamics of probe formation and retrieval, most events have their own dynamics: a series of actions leads to a goal, a sequence of words forms a sentence, a sequence of notes forms a melody. Although we leave it as a topic for future investigation, we have no reason to believe that the kind of process we have described for recognition over short timescales might not apply to these kinds of extended events as well. As a longer event, e.g., a sentence or sequence of items (e.g., Farrell & McLaughlin, 2007), unfolds over time, aspects of the event—the incoming words, prosody, and momentary understanding of the semantics of the sentence—join those features already present in working memory

and could be used to probe memory in exactly the way our model describes. The dynamics of familiarity are influenced by what features are present in the probe over time. In the present applications, probe features arise from perception, lexical access, compound cue formation, and contextual reinstatement, all processes that tend to occur quickly. There is no reason to expect that process with longer timescales—like the unfolding of a sentence or action—would not also affect familiarity by adding features to the probe in their own characteristic way.

5.1.4 Categorization

Recognition can be viewed as a form of categorization, where the category labels are “old” and “new” (Brockdorff & Lamberts, 2000; Nosofsky et al., 2011). Both categorization and recognition rely on the comparison of a test item with the contents of memory, and rely on brain regions that tend to overlap (Nosofsky, Little, & James, 2012). Recognition decisions are typically binary (“old” vs. “new”), so we consider an extension of our model to binary categorization, where participants study items from either category A or category B. Each studied exemplar results in a trace stored in memory consisting of item features, context features, and category features. Context and item features would be the same as those in the usual recognition paradigms we have been considering; category features would be additional features representing the label (“A” or “B”) assigned to the item at study. At test, the features initially present in the probe are not just context features, but also include features of one of the categories—although the choice is arbitrary, say they are the features of category A. As a result, the initial value of familiarity represents the overall familiarity of category A (in the present context). Then, as item features are sampled, familiarity will evolve over time as it does in normal recognition, only now the changes in familiarity re-

flect the item's degree of match to category A, not just to all items in memory. Thus, an item from category A will, to the extent that it shares content features with other category A items, lead to a rise in familiarity over time, and conversely for an item from category B, and the accumulated change in familiarity over time can serve as the basis for a category decision. This account of "categorization as recognition" only applies in the binary case and would likely require further expansion, but represents a reasonable starting place, particularly given its resemblance to many extant successful accounts of categorization applied to recognition (Brockdorff & Lamberts, 2000; Nosofsky et al., 2011).

5.2 Prospects for neuroscience

Recently, much progress has been made in cognitive neuroscience by interpreting measures of neural function in terms of quantitative cognitive models, including in the domains of categorization and recognition (e.g., Nosofsky et al., 2012; Mack, Preston, & Love, 2013; Turner et al., 2013). By correlating model parameters with neural activity, it is possible to make stronger inferences about the role that a particular area plays while engaged in a particular task. Neural data can also be used to compare models that may be difficult to tell apart on the basis of behavior alone (Mack et al., 2013). The advantage of model-based analysis of neural data is strongly enhanced by using a dynamic model. Thus far, such analyses have been limited to models that, although they may have a dynamic component, only specify the internal state of the participant at the level of a whole trial (Bogacz, Wagenmakers, Forstmann, & Nieuwenhuis, 2010; Turner et al., 2013). A dynamic model like the one we have presented here makes predictions about the entire trajectory of internal states through which a participant is expected to pass on the way toward making a final recognition decision. Dynamic models that describe the evolution of cognitive states over time

have the potential to allow for the interpretation of neural states over time. The utility of this approach depends, of course, on the quality of the model—the level of detail at which it describes the hypothesized cognitive processes and how well it can predict behavioral data—and on the quality of the data. The kind of model we have developed here operates on the time-scale of milliseconds, which is far too fine-grained for functional magnetic resonance imaging, although it is possible to analyze BOLD signal change on the scale of hundreds of milliseconds with sufficiently focused regions of interest (Criss, Wheeler, & McClelland, 2012). Thus, it would seem the best prospects for increasing our understanding of cognitive processes like recognition lies with EEG and MEG. Although electrophysiological techniques have been used extensively in event-related (Rugg & Curran, 2007) and spectral (Jacobs, Hwang, Curran, & Kahana, 2006) analyses of recognition memory, their utility is limited by the fact that models of recognition have not been specified at a sufficiently fine grain of temporal detail. Under the dynamic approach to recognition that we advocate, it becomes possible to use functional electrophysiology to test specific hypotheses about the time-course of mnemonic processing, since our model makes predictions about the distribution of familiarity trajectories that any particular stimulus should follow. Whether the evolution of familiarity is most apparent in the time domain or frequency domain remains a target for future work (but see Polyn & Sederberg, 2014, for progress along these lines).

5.3 Model extensions

5.3.1 Modeling confidence judgments

In this paper, we have been concerned only with two-choice tasks, yet it is common and informative to collect multiple responses in the form of confidence ratings. Several recent

models of joint RT and confidence judgments in recognition memory have assumed that each possible rating response is assigned a separate accumulator, but where the competition between them is attenuated (Ratcliff & Starns, 2009, 2013) or eliminated entirely (Van Zandt, 2000). Such models have been successful at predicting a wide range of results, and it would be possible to extend our model with their decision mechanisms. For example, the RTCON model (Ratcliff & Starns, 2009, 2013) assumes that the range of evidence is divided into several regions corresponding to each possible confidence rating. When the evidence sample on a given trial falls within one of these regions, the accumulator for its corresponding rating is incremented while the other accumulators are decremented, in inverse proportion to the number of accumulators. In this model, then, the different confidence responses have a kind of global competition, but the competition is spread out over the possible competing responses, similar to the leaky competing accumulator model (LCA; Usher & McClelland, 2001). This kind of decision mechanism would be easy to graft onto our existing model, where changes in familiarity at each time step are divided into bins, with accumulators driven whenever the change on a given time step falls within its corresponding bin. Familiarity would still reach an asymptote, however: Accounting for this may require that boundaries for different responses collapse at different rates; alternatively, different response accumulators might themselves compete, à la the LCA, such that response that had received strong evidence before familiarity reached asymptote would be able to suppress the other accumulators and result in a final choice (this alternative model formulation was discussed earlier in regards to the need for collapsing boundaries in our current model).

A more challenging situation is presented when the number of response options becomes effectively infinite by allowing participants to make confidence judgments on a con-

tinuous scale (Province & Rouder, 2012). In such a setting, it would be very impractical to assume that there exist accumulators for responses for each pixel along the scale. Province and Rouder (2012) finds evidence that participants' confidence ratings nonetheless tend to cluster around various points along this scale, arguing that this indicates discrete-state processing in memory, in contrast to the continuous valued familiarity signal that we have argued underlies recognition. Nonetheless, their results are equally well accounted for by assuming not that memory evidence is discrete, but that participants discretize their own responses: If participants entertain only a few confidence ratings dispersed throughout the scale, their final responses would be a mixture of responses centered at each of these internally-generated accumulators, just as (Province & Rouder, 2012) found, regardless of whether the evidence being accumulated was discrete or continuous.

A different possibility is offered by our model's use of collapsing decision boundaries. According to this view, recognition decisions are not always made at the same level of accumulated change in familiarity; early decisions tend to be made on the basis of stronger evidence while later decisions are based on weaker evidence. If one assumes a monotonic mapping from the accumulated change in familiarity and a continuous confidence scale, our model can be directly applied to the setting studied by Rouder. Indeed, such a model could also apply to situations with a finite number of confidence ratings: Participants could partition accumulated change in familiarity into different regions, making confidence ratings according to where familiarity hit a boundary. Such a model is constrained to predict that high-confidence judgments would be faster than low-confidence judgments, which does not always occur (Ratcliff & Starns, 2009). It may be possible to ameliorate this by assuming that boundaries stay far apart for a time before quickly collapsing, thus reducing the correlation between confidence and response time, but we leave this for future work.

It is also not clear whether this model could predict the distributions of confidence ratings obtained by Province and Rouder (2012), although such predictions will depend strongly on the form of boundary collapse and mapping from accumulated change in familiarity to confidence.

5.3.2 Alternative decision mechanisms

In our model, once the memory probe is saturated with features, familiarity reaches a stable asymptote. We have chosen to model free response recognition decisions by assuming that the criterial level of familiarity needed to make an old or new decision decreases over time, such that a recognition decision will always be made, regardless of the asymptotic level of familiarity. Although, as we have noted above, it is possible to reformulate the model such that decision bounds are constant over time, we believe that a collapsing boundary mechanism is conceptually transparent, easily interpreted, and is well-motivated by the model architecture we propose. Nonetheless, other decision mechanisms might share these virtues without changing much of the underlying model.

The evolution of familiarity over time in our model resembles an Ornstein-Uhlenbeck process, which under certain conditions also reaches an asymptotic mean, albeit with continuing fluctuation around that mean that is not present in our model. Integrating an OU process over time results in a diffusion process that closely resembles the Wiener process, which often provides a good fit to behavioral data and bears resemblance to some neural systems (Smith, 2010), suggesting that instead of allowing decision boundaries to collapse over time, one might accumulate the accumulated change in familiarity from a starting point toward fixed decision bounds. Conceptually, this seems quite reasonable to us, although our own preliminary explorations of this model variant have not led to bet-

ter behavioral fits than those shown here, so this alternative might not be distinguishable from our current model on the basis of behavioral predictions alone. Finally, as a point of theoretical interest, we note that if no additional moment-to-moment variability in familiarity is present (as is the case in our current model), an integrated version of our model would, asymptotically, become a linear ballistic process, albeit one with a distribution of asymptotic rates. Given that our model is well-approximated by a Gaussian process, this asymptotic rate distribution is itself approximately Gaussian, just as assumed in the LBA model (S. Brown & Heathcote, 2008). The LBA, of course, does not involve changing drift rates over time as features are sampled into the probe, which we argue is essential to understanding recognition; the LBA must include such a process as part of its “non-decision” time. Nonetheless, this similarity points to a reason why both our model and the LBA can fit recognition data.

5.3.3 Alternative memory representations

For simplicity, we have assumed that the features used to encode items are binary and independent. However, alternative feature representations may be used that allow for much richer information to be encoded without altering the general structure of the model, such as the assumption of a geometric rather than Bernoulli distribution of feature values in the original REM model (Shiffrin & Steyvers, 1997). The SLiM model (McClelland & Chappell, 1998), like ours, assumes that probe features are binary, but allows for memory traces to contain continuous feature values in the range (0, 1). In a sense, the memory trace stores an estimate of the probability that the studied item had a “zero” or “one” feature value. Because trace features are not discrete, the degree of match between a probe and trace is more graded than in REM, leading to different predictions about similarity (Criss

& McClelland, 2006). It may also be possible, with continuous representation assumption, to more finely model changes in the memory probe over time, rather than assuming as we have here that probe features are represented perfectly in short-term memory. Such an extension might prevent familiarity from reaching a stable asymptote, obviating the need for collapsing boundaries to model free response times. Of course, this extension would have to be formulated in such a way as to retain the good fits we obtain between our current model and both response time distributions and speed-accuracy trade-off functions. We also note that the SLiM model makes a decision using the *maximum* probe/trace match, rather than the average likelihood that we employ, although in practice the distinction between a maximum and an average has turned out to be minor (Shiffrin & Steyvers, 1998; Criss & McClelland, 2006). Although we leave it for future work, it would be eminently possible to incorporate the SLiM model's representational assumptions and decision rule without our dynamic framework to investigate whether these assumptions lead to qualitatively different predictions¹.

Alternative representations may also be entailed by the design of the stimuli themselves: We have so far been agnostic as to the meaning of any one feature, although we allowed that certain features represented, say, plurality or semantic similarity. When stimuli are well controlled, with features/dimensions that have been discovered by multidimensional

¹We suspect any differences between our model and a SLiM-like variant will be minor, however. Criss and McClelland (2006) noted that SLiM was able to account for a violation of the word frequency mirror effect under the influence of Midazolam (Hirshman, Fisher, Henthorn, Arndt, & Passannante, 2002) by only varying the learning rate (u), while REM needed to assume a difference in encoding fidelity (c). Simulations with our model, which uses binary rather than geometrically-distributed features, show that we can obtain the same aberrant word frequency effect by varying only u , with very low values of u resulting in lower hit rates for low frequency than high frequency items.

scaling or by their design, it seems reasonable to assume that the features stored in memory traces and used in the probe correspond to these well-defined aspects of the stimuli (e.g., Nosofsky, 1992). It would still be possible to use a sufficient number of binary features to encode such information: For example, the shapes “circle” and “triangle” might lie on a psychological shape dimension, but are far apart, and could then be represented within the same “bank” of binary features as, respectively, 1, 1, 1, 0, 0, 0 and 0, 0, 0, 1, 1, 1. Assuming that participants can only draw one probe feature sample at a time, however, this representation would imply that more complex features/dimensions—those that require a larger “bank” of binary features to accurately reproduce their similarity relations—would take longer to enter the probe than less complex features. Consistent with this prediction, more complex stimuli take longer to encode, rotate, and compare (Bethell-Fox & Shepard, 1988), and features that take longer to perceive also take longer to be used in retrieval (Lamberts, Brockdorff, & Heit, 2002), although effects of complexity can diminish over training (Bethell-Fox & Shepard, 1988).

Distributing representations of both primitive and complex stimuli over a set of discrete (e.g., binary) features offers a way to represent continuous feature values within the current model. Color, for instance, may well be represented continuously along dimensions of hue, saturation, and lightness, for example. But so long as the number of binary features used to encode an apparently continuous feature is large enough, this does not pose a problem for the present model. It would still, however, be straightforward to allow for continuous rather than discrete features. Say that each feature used to encode an item was not binary, but could take on any real value. For any single feature j , a particular item i will be associated with a particular value of that feature, m_{ij} . For simplicity, assume that, across all items, the values of feature j are normally distributed with mean zero and variance σ^2 . Finally,

assume that, when a value is stored for feature j , an amount of Gaussian noise is added to the stored value with zero mean and variance η^2 . Then, assuming that the value of feature j in the probe, denoted P_j , is known exactly, one can specify the likelihoods $\Pr(D|\text{Same})$ and $\Pr(D|\text{Different})$ needed in Equation 2.6:

$$\Pr(D|\text{Same}) = \phi\left(\frac{t_{ij} - p_{ij}}{\eta}\right)$$

$$\Pr(D|\text{Different}) = \phi\left(\frac{t_{ij}}{\sqrt{\sigma^2 + \eta^2}}\right)$$

where $\phi(\cdot)$ is the standard normal density function. Because a sufficient number of binary features has, thus far, offered sufficient flexibility for our purposes, we have not yet pursued this model variant, but it may well become useful in the future.

One aspect of the Gaussian representation for features is that it affords a different approach to representing encoding noise and fidelity at both study and test. In the model we have proposed, we have assumed for simplicity that any one feature is either stored or not and that, if it is stored, it is either correct or not. Additional study time allows more features to be stored. One could imagine, instead, that all available Gaussian features are stored, but with fidelity η decreasing with study time, similar to SLiM (McClelland & Chappell, 1998). How a participant is able to adjust their internal estimate of η in response to such manipulations—as well as other exploration of the consequences of the Gaussian representation—is left for future work, although it is possible that a participant could store both the mean and variance (or other measure of uncertainty) of a feature in a memory trace.

Consideration of a Gaussian feature representation brings to light another avenue of further exploration: correlations between features. Thus far, we have assumed complete independence between features. This assumption is not as strong as it may first appear:

So long as features are being treated abstractly, they may be assumed to represent a high-dimensional embedding of the stimulus where the stimulus dimensions are chosen so as to maximize independence between features, as in independent components analysis (Hyvärinen & Oja, 2000). Apparent contingencies between features can, in principle, be represented by a different set of independent features. Nonetheless, there are situations where the assumption of feature independence may be untenable or the resulting representation will be difficult to interpret, in which case a representation that incorporates covariance between features might be employed (Mueller & Shiffrin, 2006).

5.4 Conclusion

A dynamic approach to recognition memory offers a variety of insights that would be impossible either by examining memory as a static system or by treating decision dynamics separately from memory dynamics. We believe that this approach not only deepens the understanding of memory, but invites new kinds of research questions and possibilities for synthesis.

Appendix A

Continuous approximation to the model

To derive predictions from the model in an efficient manner that does not require a prohibitive amount of simulation, we developed a continuous approximation. The layout of this approximation is as follows: First, each probe-trace comparison at any given time is approximated by a log-normal distribution. Then, familiarity is approximated as another log-normal that is the average of each probe-trace comparison distribution. Next, we take advantage of the fact that samples are independent to compute the mean vector and covariance matrix of familiarity over time. Finally, we adopt the renewal process approach described by Smith (2000) to derive first-passage time distributions to each boundary given the mean and covariance.

A.1 Likelihood as a log-normal

Recall that there are five possible results of a feature comparison between a probe and a trace: match (M), mismatch (N), a feature in the probe that is not in the trace (P), a feature in the trace that is not in the probe (T), and a feature that is in neither the probe nor the trace (O). At a given time, a particular probe/trace comparison will result in a vector of counts reflecting the number of comparisons that fall into each of these types, $\mathbf{N}(t) = [N_M(t), N_N(t), N_P(t), N_T(t), N_O(t)]$. Ignoring for the moment that these values can

change over time as sampling proceeds, we can consider this vector of counts to be a random variable drawn from a multinomial distribution, with the number of trials given by the number of content features N_C and the probabilities given as above (conditioned on whether the probe and trace match or not):

$$\begin{bmatrix} N_M(t) \\ N_N(t) \\ N_P(t) \\ N_T(t) \\ N_O(t) \end{bmatrix} \sim \text{Multinomial} \left(\begin{bmatrix} p_M(t) \\ p_N(t) \\ p_P(t) \\ p_T(t) \\ p_O(t) \end{bmatrix}, N_C \right) \quad (\text{A.1})$$

$$\mathbf{N}(t) \sim \text{Multinomial}(\mathbf{p}(t), N_C).$$

The actual likelihood $\lambda(t)$ is a function of these five counts:

$$\lambda(t) = \beta_M^{N_M(t)} \beta_N^{N_N(t)} \beta_P^{N_P(t)} \beta_T^{N_T(t)}$$

where each β_j is the likelihood ratio $\frac{p_{j|S}}{p_{j|D}}$. The logarithm of $\lambda(t)$ is, thus, a weighted sum, or dot product:

$$\begin{aligned} \log \lambda(t) &= N_M(t) \log \beta_M + N_N(t) \log \beta_N + N_P(t) \log \beta_P + N_T(t) \log \beta_T \\ \log \lambda(t) &= (\log \beta) \bullet \mathbf{N}(t). \end{aligned} \quad (\text{A.2})$$

It is clear, then, that the log-likelihood is a linear function of the vector of counts, $\mathbf{N}(t)$.

For sufficiently large N_C , a multinomial distribution can be approximated by a multivariate normal distribution. To approximate the distribution of $\mathbf{N}(t)$, the multivariate nor-

mal approximation has mean and covariance matrix

$$\mu(t) = \begin{bmatrix} p_M(t)N_C \\ p_N(t)N_C \\ p_P(t)N_C \\ p_T(t)N_C \\ p_0(t)N_C \end{bmatrix} \quad (\text{A.3})$$

$$\Sigma(t) = \begin{bmatrix} p_M(t)(1-p_M(t))N_C & -p_N(t)p_M(t)N_C & -p_P(t)p_M(t)N_C & -p_T(t)p_M(t)N_C & -p_0(t)p_M(t)N_C \\ -p_M(t)p_N(t)N_C & p_N(t)(1-p_N(t))N_C & -p_P(t)p_N(t)N_C & -p_T(t)p_N(t)N_C & -p_0(t)p_N(t)N_C \\ -p_M(t)p_P(t)N_C & -p_N(t)p_P(t)N_C & p_P(t)(1-p_P(t))N_C & -p_T(t)p_P(t)N_C & -p_0(t)p_P(t)N_C \\ -p_M(t)p_T(t)N_C & -p_N(t)p_T(t)N_C & -p_P(t)p_T(t)N_C & p_T(t)(1-p_T(t))N_C & -p_0(t)p_T(t)N_C \\ -p_M(t)p_0(t)N_C & -p_N(t)p_0(t)N_C & -p_P(t)p_0(t)N_C & -p_T(t)p_0(t)N_C & p_0(t)(1-p_0(t))N_C \end{bmatrix} \quad (\text{A.4})$$

Then, since the log-likelihood is a linear function of $\mathbf{N}(t)$, which is approximately multivariate normal, the log-likelihood is itself normal:

$$\log \lambda(t) \sim \mathcal{N} \left((\log \beta) \mu(t), (\log \beta)^T \Sigma(t) (\log \beta) \right). \quad (\text{A.5})$$

And thus it is seen how the likelihood ratio reflecting the match between a trace and the probe is approximately log-normally distributed.

A.2 Familiarity as an average of log-normals

The problem of finding the distribution of a sum (or average) of log-normal random variables has been treated extensively in the engineering literature, however no universally satisfying approximation has arisen. In many cases, the sum of several log-normal random variables itself resembles a log-normal. This is particularly useful for our purposes, since normal distributions are easier to work with. So, our first goal is to find a suitable approximation to the sum of several log-normals that is, itself, a log-normal.

A.2.1 Log-normal sum approximation

Since each trace is encoded independently, each likelihood ratio $\lambda_i(t)$ are independently distributed. The distribution function $F(x)$ of a convolution (sum) of independent random variables is less than or equal to the product of the distribution function of the convolved variates, i.e., $F(x) \lesssim \prod_{i=1}^N F_i(x)$. Intuitively, this is because for the sum to be less than or equal to a certain value, all of the summands must also be less than or equal to that value. In the case of a log-normal, where the distributions are strongly skewed, the product of the distribution functions of each component gives a good approximation to the true distribution.

The distribution function of an individual log-normal can be expressed in terms of the standard normal distribution function:

$$F_i(x) = \Phi\left(\frac{\log x - \mu_i}{\sigma_i}\right) \quad (\text{A.6})$$

where μ_i and σ_i are the logarithmic mean and standard deviation, respectively. Thus, the distribution function of a sum of N log-normals is approximated by

$$F(x) \lesssim \prod_{i=1}^N \Phi\left(\frac{\log x - \mu_i}{\sigma_i}\right) \quad (\text{A.7})$$

which is quite easy to compute, given the means and standard deviations as computed above for each likelihood ratio.

To approximate the sum distribution with another log-normal, we need to find its mean and standard deviation, denoted μ_Z and σ_Z . Under the assumption of normality, we can approximate these using the distribution function given in Equation A.7: The mean should be located at the median, where $F(\mu_Z) = 0.5$ and the standard deviation should be half the distance between x_u and x_l where $F(x_u) = \Phi(1) \approx .84$ and $F(x_l) = \Phi(-1) \approx .16$, since

this is the region that is one standard deviation on either side of the mean of a normal distribution.

We cannot directly invert the log-normal sum distribution function, but it is easy to numerically solve these three equations, which we do with a simple univariate root-finding routine. Now, we have approximated the sum of N log-normals with a single log-normal and found its mean and standard deviation.

A.2.2 Computing the average

Recall that familiarity is the *average* likelihood over the *activated* traces. A trace is considered “active” if its likelihood ratio exceeds one, or on a log scale, if its log-likelihood ratio exceeds zero. Thus, computing the probability of trace activation is easy with the log-normal approximation:

$$\Pr(\text{Trace } i \text{ active}) = 1 - \Phi\left(-\frac{\mu_i}{\sigma_i}\right). \quad (\text{A.8})$$

Strictly speaking, then, the distribution of familiarity is a *mixture* of those that would result from each combination of the N traces being active or not. Given that the number of such combinations grows factorially with N and that we are only interested in the first two moments of the distribution, rather than its exact shape, we can take a few shortcuts: First, because “deactivated” traces contribute very little to the resulting sum (since they must, by definition, be less than one), we simply compute the mean and s.d. of the sum using all N traces, regardless of activation status. Next, we divide the sum distribution by a

denominator S , which is the sum of the activation probabilities over all N traces:

$$\begin{aligned} S &= \sum_{i=1}^N \Pr(\text{Trace } i \text{ active}) \\ &= \sum_{i=1}^N \left[1 - \Phi \left(-\frac{\mu_i}{\sigma_i} \right) \right]. \end{aligned} \quad (\text{A.9})$$

Thus, S is the *expected* number of activated traces.

Because the sum distribution is log-normal, dividing by S simply amounts to subtracting $\log S$ from the logarithmic mean μ_Z . Thus, the mean and standard deviation of log-familiarity at time t are:

$$\mu_\phi(t) = \mu_Z - \log S \quad (\text{A.10})$$

$$\sigma_\phi(t) = \frac{1}{2} (x_u - x_l) \quad (\text{A.11})$$

where x_u and x_l are defined in the previous section.

A.2.3 Familiarity over time

So far, we have shown how to approximate the familiarity distribution at a single time t . However, a key feature of the full model is that familiarity evolves over time. Fortunately, another key feature of the model is that features are sampled independently at each time step. Thus, the change in familiarity from time t to time $t + 1$ is independent of the actual *value* of familiarity at time t . It is, however, dependent on elapsed time t , because the probe is eventually saturated with features until there are no new features available to be sampled.

Because of this dependence structure, each step of familiarity contributes unique variance to the value of familiarity. In other words, the covariance between familiarity at time t and familiarity at time s is the minimum variance at each of the two times:

$$\min \left(\sigma_\phi^2(t), \sigma_\phi^2(s) \right).$$

In this way, we can construct a covariance matrix Ω that describes how familiarity evolves over time. This, coupled with a vector of means $\mu_\phi(0 \dots t_{MAX})$, completely characterizes the distribution of familiarity across time.

From this, we can easily obtain the distribution of *changes* in familiarity over time.

Define a linear difference operator D , a matrix of the form

$$D = \begin{bmatrix} -1 & 1 & 0 & 0 & \dots \\ 0 & -1 & 1 & 0 & \dots \\ 0 & 0 & -1 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

and apply this operator to μ_ϕ and Ω to find the mean vector and covariance matrix of the differences (step sizes) in familiarity over time:

$$\mu_d = D\mu_\phi \tag{A.12}$$

$$\sigma_d^2 = \text{diag}(D^T \Omega D) \tag{A.13}$$

Because of the structure of Ω , $D^T \Omega D$ is diagonal (since each step is independent of the other steps), so really each step is distributed as a normal, independent of all other steps.

A.3 Decision probabilities and latency

Armed with the mean and variance of familiarity over time, there are several ways to compute the joint probability of familiarity reaching either the “old” or “new” decision bound after a number of steps t . We adopt a renewal-process approach based on the following logic (presented in terms of the first-passage time for the “old” boundary, but the reasoning applies to the “new” boundary as well): If $x(t)$ crosses the “old” boundary $B_O(t)$ at time t , there are three possibilities:

1. This is the first time $x(t)$ has crossed the “old” boundary, the probability of which is denoted $f_O(t)$.
2. $x(t)$ crossed the “old” boundary at some earlier time, moved back down, and is now crossing it again, the probability of which is denoted $r_{OO}(t)$.
3. $x(t)$ crossed the “new” boundary at some earlier time and has since moved back up to cross the “old” boundary, the probability of which is denoted $r_{NO}(t)$.

The probability that $x(t)$ crosses the “old” boundary at all at time t (regardless of its prior history) is the joint probability that $x(t) > B_O(t)$ and $x(t-1) < B_O(t-1)$:

$$p_O(t) = \Pr[x(t) > B_O(t) \wedge x(t-1) < B_O(t-1)]. \quad (\text{A.14})$$

This and all subsequent bivariate normal cumulative distribution functions were computed using high-precision algorithms by Genz (2004). Finding $f_O(t)$ is, then, a simple matter of rearranging terms:

$$\begin{aligned} p_O(t) &= f_O(t) + r_{OO}(t) + r_{NO}(t) \\ f_O(t) &= p_O(t) - r_{OO}(t) - r_{NO}(t). \end{aligned} \quad (\text{A.15})$$

The probability that $x(t)$ crossed the “old” boundary earlier is the “renewal” probability

$$r_{OO}(t) = \sum_{\tau=0}^{t-1} f_O(\tau) \Pr[x(t) > B_O(t) | \phi(\tau) > B_O(\tau)]. \quad (\text{A.16})$$

Similarly, the “renewal” probability of first crossing the “new” boundary and then crossing the “old” boundary at time t is

$$r_{NO}(t) = \sum_{\tau=0}^{t-1} f_N(\tau) \Pr[x(t) > B_O(t) | \phi(\tau) < B_N(\tau)]. \quad (\text{A.17})$$

Given the time-varying boundaries $B_O(t)$ and $B_N(t)$ and the mean μ_ϕ and variance Σ_ϕ , the results of Equations A.14, A.16, and A.17 can all be computed and used to solve for the

first-passage time density (Equation A.15). The densities thus found, $f_O(t)$ and $f_N(t)$, are defective in that they add to the total probability of crossing their respective boundary at any time, rather than 1, however they jointly determine the total probability of ϕ reaching either boundary, allowing a decision to be made.

Finally, we convert model “time-steps” to real time via a linear transformation via Equation 2.13, as described in the main text. The final result is a stepwise (in time) joint probability density of a response at a particular time, thus allowing one to compute the likelihood of a particular response or to integrate this density over a region of time to obtain predicted RT quantiles or quantile likelihoods, as used for model-fitting in the main text.

References

- Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity not word frequency determines word naming and lexical decision times. *Psychological Science*, 17, 814–823.
- Anderson, J. R. (1983). Retrieval of information from long-term memory. *Science*, 220(4592), 25–30.
- Atkinson, R. C., & Juola, J. F. (1974). Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology: I. learning, memory and thinking*. Oxford, England: W. H. Freeman.
- Balakrishnan, J. D., & Macdonald, J. A. (2011). Performance measures for dynamic signal detection. *Journal of Mathematical Psychology*, 55, 290–301.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577–660.
- Becker, C. A. (1976). Allocation of attention during visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 2(4), 556–566.
- Bellman, R. (1952). On the theory of dynamic programming. *Proceedings of the National Academy of Sciences*, 38(8), 716–179.
- Bethell-Fox, C. E., & Shepard, R. N. (1988). Mental rotation: Effects of stimulus com-

- plexity and familiarity. *Journal of Experimental Psychology: Human Perception and Performance*, 14(1), 12–23.
- Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, 113(4), 700–765.
- Bogacz, R., Wagenmakers, E.-J., Forstmann, B. U., & Nieuwenhuis, S. (2010). The neural basis of the speed–accuracy tradeoff. *Trends in Neurosciences*, 33(1), 10–16.
- Brockdorff, N., & Lamberts, K. (2000). A feature-sampling account of the time course of old-new recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(1), 77–102.
- Brown, J., Lewis, V., & Monk, A. (1977). Memorability, word frequency and negative recognition. *The Quarterly Journal of Experimental Psychology*, 29(3), 461–473.
- Brown, S., & Heathcote, A. (2005). A ballistic model of choice response time. *Psychological Review*, 112(1), 117–128.
- Brown, S., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, 57, 153–178.
- Brunton, B. W., Botvinick, M. M., & Brody, C. D. (2013). Rats and humans can optimally accumulate evidence for decision-making. *Science*, 340, 95–98.
- Buchler, N. G., Light, L. L., & Reder, L. M. (2008). Memory for items and associations: Distinct representations and processes in associative recognition. *Journal of Memory and Language*, 59, 183–199.
- Busmeyer, J. R., & Rapoport, A. (1988). Psychological models of deferred decision making. *Journal of Mathematical Psychology*, 32, 91–134.
- Busmeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic–cognitive

- approach to decision making in an uncertain environment. *Psychological Review*, 100(3), 432–459.
- Cisek, P., Puskas, G. A., & El-Murr, S. (2009). Decisions in changing conditions: The urgency-gating model. *Journal of Neuroscience*, 29(37), 11560–11571.
- Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review*, 3(1), 37–60.
- Clark, S. E., & Shiffrin, R. M. (1987). Recognition of multiple-item probes. *Memory & Cognition*, 15(5), 367–378.
- Cohen, A. L., & Nosofsky, R. M. (2003). An extension of the exemplar-based random-walk model to separable-dimension stimuli. *Journal of Mathematical Psychology*, 47, 150–165.
- Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology*, 33A(4), 497–505.
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108(1), 204–256.
- Cooper, L. A., & Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. In W. G. Chase (Ed.), *Visual information processing* (pp. 75–176). New York: Academic Press.
- Cox, G. E., & Shiffrin, R. M. (2012). Criterion setting and the dynamics of recognition memory. *Topics in Cognitive Science*, 4(1), 135–150.
- Criss, A. H., & McClelland, J. L. (2006). Differentiating the differentiation models: A comparison of the retrieving effectively from memory model (REM) and the subjective likelihood model (SLiM). *Journal of Memory and Language*, 55, 447–460.

- Criss, A. H., & Shiffrin, R. M. (2004a). Context noise and item noise jointly determine recognition memory: A comment on Dennis and Humphreys (2001). *Psychological Review*, 111(3), 800–807.
- Criss, A. H., & Shiffrin, R. M. (2004b). Pairs do not suffer interference from other types of pairs or single items in associative recognition. *Memory & Cognition*, 32(8), 1284–1297.
- Criss, A. H., & Shiffrin, R. M. (2005). List discrimination in associative recognition and implications for representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(6), 1199–1212.
- Criss, A. H., Wheeler, M. E., & McClelland, J. L. (2012). A differentiation account of recognition memory: Evidence from fMRI. *Journal of Cognitive Neuroscience*, 25(3), 421–435.
- Curran, T. (1999). The electrophysiology of incidental and intentional retrieval: ERP old/new effects in lexical decision and recognition memory. *Neuropsychologia*, 37, 771–785.
- Curran, T. (2000). Brain potentials of recollection and familiarity. *Memory & Cognition*, 28(6), 923–938.
- Curran, T., & Hintzman, D. L. (1995). Violations of the independence assumption in process dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(3), 531–547.
- Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of experimental psychology*, 58(1), 17.
- Dennis, S., & Chapman, A. (2010). The inverse list length effect: A challenge for pure exemplar models of recognition memory. *Journal of Memory and Language*, 63(3),

416–424.

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, 108(2), 452–478.

Dewhurst, S. A., & Conway, M. A. (1994). Pictures, images, and recollective experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(5), 1088–1098.

de Zubicaray, G. I., McMahon, K. L., Eastburn, M. M., Finnigan, S., & Humphreys, M. S. (2005). fMRI evidence of word frequency and strength effects during episodic memory encoding. *Cognitive Brain Research*, 22, 439–450.

Diana, R. A., Reder, L. M., Arndt, J., & Park, H. (2006). Models of recognition: A review of arguments in favor of a dual-process account. *Psychonomic Bulletin & Review*, 13(1), 1–21.

Diederich, A. (2003). Decision making under conflict: Decision time as a measure of conflict strength. *Psychonomic Bulletin & Review*, 10(1), 167–176.

Diller, D. E., Nobel, P. A., & Shiffrin, R. M. (2001). An ARC-REM model for accuracy and response time in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(2), 414–435.

Ditterich, J. (2006). Evidence for time-variant decision making. *European Journal of Neuroscience*, 24, 3628–3641.

Donaldson, D. I., & Rugg, M. D. (1998). Recognition memory for new associations: electrophysiological evidence for the role of recollection. *Neuropsychologia*, 36(5), 377–395.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition*, 24(4), 523–533.

- Donkin, C., Brown, S., Heathcote, A., & Wagenmakers, E.-J. (2011). Diffusion versus linear ballistic accumulation: Different models but the same conclusions about psychological processes? *Psychonomic Bulletin & Review*, *55*, 140–151.
- Donkin, C., & Nosofsky, R. M. (2012a). A power-law model of psychological memory strength in short- and long-term recognition. *Psychological Science*, *23*(6), 625–634.
- Donkin, C., & Nosofsky, R. M. (2012b). The structure of short-term memory scanning: an investigation using response time distribution models. *Psychonomic Bulletin & Review*, *19*, 363–394.
- Dosher, B. A. (1984a). Degree of learning and retrieval speed: Study time and multiple exposures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(4), 541–574.
- Dosher, B. A. (1984b). Discriminating preexperimental (semantic) from learned (episodic) associations: A speed-accuracy study. *Cognitive Psychology*, *16*, 519–555.
- Dosher, B. A., & Rosedale, G. (1989). Integrated retrieval cues as a mechanism for priming in retrieval from memory. *Journal of Experimental Psychology: General*, *118*(2), 191–211.
- Dosher, B. A., & Rosedale, G. (1991). Judgments of semantic and episodic relatedness: Common time-course and failure of segregation. *Journal of Memory and Language*, *30*, 125–160.
- Dosher, B. A., & Rosedale, G. (1997). Configural processing in memory retrieval: Multiple cues and ensemble representations. *Cognitive Psychology*, *33*, 209–265.
- Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N., & Pouget, A. (2012). The cost of accumulating evidence in perceptual decision making. *Journal of Neuroscience*, *32*(11), 3612–3628.

- Dunn, J. C. (2004). Remember-know: A matter of confidence. *Psychological Review*, 111(2), 524–542.
- Dunn, J. C. (2008). The dimensionality of the remember-know task: A state-trace analysis. *Psychological Review*, 115(2), 426–446.
- Düzel, E., Yonelinas, A. P., Mangun, G. R., Heinze, H.-J., & Tulving, E. (1997). Event-related brain potential correlates of two states of conscious awareness in memory. *Proceedings of the National Academy of Sciences*, 94(11), 5973–5978.
- Edwards, W. (1965). Optimal strategies for seeking information: Models for statistics, choice reaction times, and human information processing. *Journal of Mathematical Psychology*, 2, 312–329.
- Farrell, S., & McLaughlin, K. (2007). Short-term recognition memory for serial order and timing. *Memory & Cognition*, 35(7), 1724–1734.
- Feldman, L. B., O'Connor, P. A., & Del Prado Martín, F. M. (2009). Early morphological processing is morphosemantic and not simply morpho-orthographic: A violation of form-then-meaning accounts of word recognition. *Psychonomic Bulletin & Review*, 16(4), 684–691.
- Frazier, P. I., & Yu, A. J. (2008). Sequential hypothesis testing under stochastic deadlines. In *Advances in neural information processing systems* (Vol. 20, pp. 465–472). Cambridge, MA: MIT Press.
- Freyd, J. J. (1987). Dynamic mental representations. *Psychological Review*, 94(4), 427–438.
- Genz, A. (2004). Numerical computation of rectangular bivariate and trivariate normal and t probabilities. *Statistics and Computing*, 14, 251–260.
- Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall.

- Psychological Review*, 91(1), 1–67.
- Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, 13(1), 8–20.
- Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(1), 5–16.
- Gold, J. I., & Shadlen, M. N. (2000). Representation of a perceptual decision in developing oculomotor commands. *Nature*, 404, 390–394.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: John Wiley & Sons, Inc.
- Greve, A., van Rossum, M. C. W., & Donaldson, D. I. (2007). Investigating the functional interaction between semantic and episodic memory: Convergent behavioral and electrophysiological evidence for the role of familiarity. *NeuroImage*, 34, 801–814.
- Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, 114(2), 211–244.
- Gronlund, S. D., & Ratcliff, R. (1989). Time course of item and associative information: Implications for global memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(5), 846–858.
- Gronlund, S. D., & Ratcliff, R. (1991). Analysis of the Hockley and Murdock decision model. *Journal of Mathematical Psychology*, 35, 319–344.
- Heath, R. A. (1992). A general nonstationary diffusion model for two-choice decision-making. *Mathematical Social Sciences*, 23, 283–309.
- Heathcote, A., Brown, S., & Mewhort, D. J. K. (2002). Quantile maximum likelihood estimation of response time distributions. *Psychonomic Bulletin & Review*, 9(2),

394–401.

- Heit, E., Brockdorff, N., & Lamberts, K. (2003). Adaptive changes of response criterion in recognition memory. *Psychonomic Bulletin & Review*, 10(3), 718–723.
- Heit, E., Brockdorff, N., & Lamberts, K. (2004). Strategic processes in false recognition memory. *Psychonomic Bulletin & Review*, 11(2), 380–386.
- Hillstrom, A. P., & Logan, G. D. (1997). Process dissociation, cognitive architecture, and response time: Comments on Lindsay and Jacoby (1994). *Journal of Experimental Psychology: Human Perception and Performance*, 23(5), 1561–1578.
- Hintzman, D. L., & Curran, T. (1994). Retrieval dynamics of recognition and frequency judgments: Evidence for separate processes of familiarity and recall. *Journal of Memory and Language*, 33, 1–18.
- Hintzman, D. L. (1988). Judgements of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, 95(4), 528–551.
- Hintzman, D. L. (1994). On explaining the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(1), 201–205.
- Hintzman, D. L. (2002). Context matching and judgments of recency. *Psychonomic Bulletin & Review*, 9(2), 368–374.
- Hintzman, D. L. (2004). Judgment of frequency versus recognition confidence: Repetition and recursive reminding. *Memory & Cognition*, 32(2), 336–350.
- Hintzman, D. L. (2010). How does repetition affect memory? evidence from judgments of recency. *Memory & Cognition*, 38(1), 102–115.
- Hintzman, D. L., & Caulton, D. A. (1997). Recognition memory and modality judgments: A comparison of retrieval dynamics. *Journal of Memory and Language*, 37, 1–23.
- Hintzman, D. L., Caulton, D. A., & Curran, T. (1994). Retrieval constraints and the mirror

- effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(2), 275–289.
- Hintzman, D. L., Caulton, D. A., & Levitin, D. J. (1998). Retrieval dynamics in recognition and list discrimination: Further evidence of separate processes of familiarity and recall. *Memory & Cognition*, 26(3), 449–462.
- Hintzman, D. L., & Curran, T. (1997). Comparing retrieval dynamics in recognition memory and lexical decision. *Journal of Experimental Psychology: General*, 126(3), 228–247.
- Hirshman, E., Fisher, J., Henthorn, T., Arndt, J., & Passannante, A. (2002). Midazolam amnesia and dual-process models of the word-frequency mirror effect. *Journal of Memory and Language*, 47, 499–516.
- Hirshman, E., & Master, S. (1997). Modeling the conscious correlates of recognition memory: Reflections on the remember-know paradigm. *Memory & Cognition*, 25(3), 345–351.
- Hockley, W. E., & Murdock, B. B. (1987). A decision model for accuracy and response latency in recognition memory. *Psychological Review*, 94(3), 341–358.
- Hockley, W. E., & Murdock, B. B. (1992). Speed-accuracy trade-off and item recognition: A reply to Gronlund and Ratcliff. *Journal of Mathematical Psychology*, 36, 461–467.
- Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, 46, 269–299.
- Howard, M. W., Shankar, K. H., & Jagadisan, U. K. K. (2011). Constructing semantic representations from a gradually-changing representation of temporal context. *TopiCS in Cognitive Science*, 3(1), 48–73.

- Huk, A. C., & Shadlen, M. N. (2005). Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. *Journal of Neuroscience*, 25(45), 10420–10436.
- Humphreys, M. S., Dennis, S., Chalmers, K. A., & Finnigan, S. (2000). Dual processes in recognition: Does a focus on measurement operations provide a sufficient foundation? *Psychonomic Bulletin & Review*, 7(4), 593–603.
- Hyvärinen, A., & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks*, 13, 411–430.
- Jacobs, J., Hwang, G., Curran, T., & Kahana, M. J. (2006). EEG oscillations and recognition memory: Theta correlates of memory retrieval and decision making. *NeuroImage*, 32, 978–987.
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, 30, 513–541.
- Jacoby, L. L. (1999). Ironic effects of repetition: Measuring age-related differences in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(1), 3–22.
- Jadhav, S. P., Kemere, C., German, P. W., & Frank, L. M. (2012). Awake hippocampal sharp-wave ripples support spatial memory. *Science*, 336, 1454–1458.
- Jang, Y., & Huber, D. E. (2008). Context retrieval and context change in free recall: Recalling from long-term memory drives list isolation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(1), 112–127.
- Johns, B. T., & Jones, M. N. (2010). Evaluating the random representation assumption of lexical semantics in cognitive models. *Psychonomic Bulletin & Review*, 17(5), 662–672.

- Johns, B. T., & Jones, M. N. (2012). Perceptual inference through global lexical similarity. *Topics in Cognitive Science*, 4, 103–120.
- Jones, M., & Dzhafarov, E. N. (2014). Unfalsifiability and mutual translatability of major modeling schemes for choice reaction time. *Psychological Review*, 121(1), 1–32.
- Jones, M. N., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology*, 66(2), 115–124.
- Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, 114(1), 1–37.
- Kahana, M. J. (2012). *Foundations of human memory*. New York, NY: Oxford University Press.
- Kahana, M. J., Rizzuto, D. S., & Schneider, A. R. (2005). Theoretical correlations and measured correlations: Relating recognition and recall in four distributed memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(5), 933–953.
- Kent, C., & Lamberts, K. (2008). The encoding–retrieval relationship: Retrieval as mental simulation. *Trends in Cognitive Sciences*, 12(3), 92–98.
- Kiani, R., Hanks, T. D., & Shadlen, M. N. (2008). Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *Journal of Neuroscience*, 28(12), 3017–3029.
- Kılıç, A., & Öztekin, I. (2014). Retrieval dynamics of the strength based mirror effect in recognition memory. *Journal of Memory and Language*, 76, 158–173.
- Klein, K. A., Shiffrin, R. M., & Criss, A. H. (2007). Putting context in context. In J. S. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger III*. New York: Psychology Press.

- Kumaran, D., & McClelland, J. L. (2012). Generalization through the recurrent interaction of episodic memories: A model of the hippocampal system. *Psychological Review*, 119(3), 573–616.
- Kwantes, P. J. (2005). Using context to build semantics. *Psychonomic Bulletin & Review*, 12(4), 703–710.
- LaBerge, D. (1962). A recruitment theory of simple behavior. *Psychometrika*, 27(4), 375–396.
- Lamberts, K. (2000). Information-accumulation theory of speeded categorization. *Psychological Review*, 107(2), 227–260.
- Lamberts, K., Brockdorff, N., & Heit, E. (2002). Perceptual processes in matching and recognition of complex pictures. *Journal of Experimental Psychology: Human Perception and Performance*, 28(5), 1176–1191.
- Lamberts, K., Brockdorff, N., & Heit, E. (2003). Feature-sampling and random-walk models of individual-stimulus recognition. *Journal of Experimental Psychology: General*, 132(3), 351–378.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2), 211–240.
- Link, S. W. (1975). The relative judgment theory of two choice response time. *Journal of Mathematical Psychology*, 12, 114–135.
- Link, S. W., & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika*, 40, 77–105.
- Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, and Computers*, 28(2),

203–208.

- Mack, M. L., Preston, A. R., & Love, B. C. (2013). Decoding the brain's algorithm for categorization from its neural implementation. *Current Biology*, 23(20), 2023–2027.
- Malmberg, K. J., Holden, J. E., & Shiffrin, R. M. (2004). Modeling the effects of repetitions, similarity, and normative word frequency on old–new recognition and judgments of frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(2), 319–331.
- Malmberg, K. J. (2008). Recognition memory: A review of the critical findings and an integrated theory for relating them. *Cognitive Psychology*, 57, 335–384.
- Malmberg, K. J., & Nelson, T. O. (2003). The word frequency effect for recognition memory and the elevated-attention hypothesis. *Memory & Cognition*, 31(1), 35–43.
- Malmberg, K. J., & Shiffrin, R. M. (2005). The “one-shot” hypothesis for context storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(2), 322–336.
- Malmberg, K. J., Steyvers, M., Stevens, J. D., & Shiffrin, R. M. (2002). Feature frequency effects in recognition memory. *Memory & Cognition*, 30(4), 607–613.
- Mandler, G. (2008). Familiarity breeds attempts: A critical review of dual-process theories of recognition. *Perspectives on Psychological Science*, 3(5), 390–399.
- McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, 86(4), 287–330.
- McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review*, 105(4), 724–760.
- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are com-

- plementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102(3), 419–457.
- McElree, B., Dolan, P. O., & Jacoby, L. L. (1999). Isolating the contributions of familiarity and source information to item recognition: A time course analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(3), 563–582.
- McElree, B., & Doshier, B. A. (1989). Serial position and set size in short-term memory: The time course of recognition. *Journal of Experimental Psychology: General*, 118(4), 346–373.
- Metcalf, J. (1982). A composite holographic associative recall model. *Psychological Review*, 89(6), 627–661.
- Meyer, D. E., Irwin, D. E., Osman, A. M., & Kounios, J. (1988). The dynamics of cognition and action: Mental processes inferred from speed-accuracy decomposition. *Psychological Review*, 95(2), 183–237.
- Mollison, M. V., & Curran, T. (2012). Familiarity in source memory. *Neuropsychologia*, 50, 2546–2565.
- Mueller, S. T., & Shiffrin, R. M. (2006). REM—II: A Bayesian model of the organization of semantic and episodic memory systems. In *International conference on learning and development*. Bloomington, IN.
- Mullen, K., Ardia, D., Gil, D., Windover, D., & Cline, J. (2011). DEoptim: An R package for global optimization by differential evolution. *Journal of Statistical Software*, 40(6), 1–26.
- Mulligan, N., & Hirshman, E. (1995). Speed-accuracy trade-offs and the dual process model of recognition memory. *Journal of Memory and Language*, 34, 1–18.

- Munoz, D. P., Dorris, M. C., Paré, M., & Everling, S. (2000). On your mark, get set: Brainstem circuitry underlying saccadic initiation. *Canadian Journal of Physiological Pharmacology*, 78, 934–944.
- Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, 89(3), 609–626.
- Nelder, J. A., & Mead, R. (1965). A simplex method for function optimization. *Computer Journal*, 7(4), 308–313.
- Nelson, A. B., & Shiffrin, R. M. (2013). The co-evolution of knowledge and event memory. *Psychological Review*, 120(2), 356–394.
- Nobel, P. A., & Shiffrin, R. M. (2001). Retrieval processes in recognition and cued recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(2), 384–413.
- Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review*, 110(4), 611–646.
- Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 115(1), 39–57.
- Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(4), 700–708.
- Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology*, 43, 25–53.
- Nosofsky, R. M., Cao, R., Cox, G. E., & Shiffrin, R. M. (2014). Familiarity and categorization processes in memory search. *Cognitive Psychology*, 75, 97–129.

- Nosofsky, R. M., Cox, G. E., Cao, R., & Shiffrin, R. M. (2014). An exemplar-familiarity model predicts short-term and long-term probe recognition across diverse forms of memory search. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(6), 1524–1539.
- Nosofsky, R. M., Little, D. R., Donkin, C., & Fific, M. (2011). Short-term memory scanning viewed as exemplar-based categorization. *Psychological Review*, 118(2), 280–315.
- Nosofsky, R. M., Little, D. R., & James, T. W. (2012). Activation in the neural network responsible for categorization and recognition reflects parameter changes. *Proceedings of the National Academy of Sciences*, 109(1), 333–338.
- Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, 104(2), 266–300.
- Nosofsky, R. M., & Stanton, R. D. (2006). Speeded old-new recognition of multidimensional perceptual stimuli: Modeling performance at the individual-participant and individual-item levels. *Journal of Experimental Psychology: Human Perception and Performance*, 32(2), 314–334.
- Nosofsky, R. M., & Zaki, S. R. (2003). A hybrid-similarity exemplar model for predicting distinctiveness effects in perceptual old–new recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(6), 1194–1209.
- Pazzaglia, A. M., Staub, A., & Rotello, C. M. (2014). Encoding time and the mirror effect in recognition memory: Evidence from eyetracking. *Journal of Memory and Language*, 75, 77–92.
- Pike, R. (1973). Response latency models for signal detection. *Psychological Review*, 80(1), 53–68.

- Plate, T. A. (2003). *Holographic reduced representations*. Stanford, CA: CSLI Publications.
- Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, 116(1), 129–156.
- Polyn, S. M., & Sederberg, P. B. (2014). Brain rhythms in mental time travel. *NeuroImage*, 85, 678–684.
- Price, K. V., Storn, R. M., & Lampinen, J. A. (2006). *Differential evolution: A practical approach to global optimization*. Berlin Heidelberg: Springer-Verlag.
- Province, J. M., & Rouder, J. N. (2012). Evidence for discrete-state processing in recognition memory. *Proceedings of the National Academy of Sciences*, 109(36), 14357–14362.
- Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, 88(2), 93–134.
- Rae, B., Heathcote, A., Donkin, C., Averell, L., & Brown, S. (2014). The hare and the tortoise: Emphasizing speed can change the evidence used to make decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(5), 1226–1243.
- Ranganath, C., Yonelinas, A. P., Cohen, M. X., Dy, C. J., Tom, S. M., & D’Esposito, M. (2003). Dissociable correlates of recollection and familiarity within the medial temporal lobes. *Neuropsychologia*, 42, 2–13.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85(2), 59–108.
- Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, 86(3), 446–461.
- Ratcliff, R. (1988). Continuous versus discrete information processing: Modeling accu-

- mulation of partial information. *Psychological Review*, 95(2), 238–255.
- Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). List-strength effect: I. data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*(2), 163–178.
- Ratcliff, R., & Frank, M. J. (2012). Reinforcement-based decision making in corticostriatal circuits: Mutual constraints by neurocomputational and diffusion models. *Neural Computation*, 24, 1186–1229.
- Ratcliff, R., & Murdock, B. B. (1976). Retrieval processes in recognition memory. *Psychological Review*, 83(3), 190–214.
- Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, 9(5), 347–356.
- Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, 111(2), 333–367.
- Ratcliff, R., & Smith, P. L. (2010). Perceptual discrimination in static and dynamic noise: The temporal relation between perceptual encoding and decision making. *Journal of Experimental Psychology: General*, 139(1), 70–94.
- Ratcliff, R., & Starns, J. J. (2009). Modeling confidence and response time in recognition memory. *Psychological Review*, 116(1), 59–83.
- 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), 697–719.
- Ratcliff, R., & Tuerlinx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, 9(3), 438–481.

- Ratcliff, R., Van Zandt, T., & McKoon, G. (1995). Process dissociation, single-process theories, and recognition memory. *Journal of Experimental Psychology: General*, 124(4), 352–374.
- Reber, P. J., Alvarez, P., & Squire, L. R. (1997). Reaction time distributions across normal forgetting: Searching for markers of memory consolidation. *Learning & Memory*, 4, 284–290.
- Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember-know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(2), 294–320.
- Reed, A. V. (1973). Speed-accuracy trade-off in recognition memory. *Science*, 181(4099), 574–576.
- Reed, A. V. (1976). List length and the time course of recognition in immediate memory. *Memory & Cognition*, 4(1), 16–30.
- Ricciardi, L. M. (1976). On the transformation of a diffusion process into the wiener process. *Journal of Mathematical Analysis and Applications*, 54, 185–199.
- Riordan, B., & Jones, M. N. (2010). Redundancy in perceptual and linguistic experience: Comparing feature-based and distributional models of semantic representation. *Topics in Cognitive Science*, 3(2), 1–43.
- Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(4), 803–814.
- Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed*

- processing approach*. Cambridge, MA: MIT Press.
- Roitman, J. D., & Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of Neuroscience*, 22(21), 9475–9489.
- Rotello, C. M., & Heit, E. (1999). Two-process models of recognition memory: Evidence for recall-to-reject? *Journal of Memory and Language*, 40, 432–453.
- Rotello, C. M., & Heit, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory & Cognition*, 28(6), 907–922.
- Rotello, C. M., & Zeng, M. (2008). Analysis of RT distributions in the remember-know paradigm. *Psychonomic Bulletin & Review*, 15(4), 825–832.
- Rouder, J. N., Lu, J., Speckman, P. L., Sun, D., & Jiang, Y. (2005). A hierarchical model for estimating response time distributions. *Psychonomic Bulletin & Review*, 12(2), 195–223.
- Rouder, J. N., & Speckman, P. L. (2004). An evaluation of the Vincentizing method of forming group-level response time distributions. *Psychonomic Bulletin & Review*, 11(3), 419–427.
- Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences*, 11(6), 251–257.
- Schmidt, S. R. (1991). Can we have a distinctive theory of memory? *Memory & Cognition*, 19(6), 523–542.
- Schwartz, G., Howard, M. W., Jing, B., & Kahana, M. J. (2005). Shadows of the past: Temporal retrieval effects in recognition memory. *Psychological Science*, 16(11), 898–904.
- Sederberg, P. B., Howard, M. W., & Kahana, M. J. (2008). A context-based theory of

- recency and contiguity in free recall. *Psychological Review*, 115(4), 893–912.
- Senkfor, A. J., & Van Petten, C. (1998). Who said what? an event-related potential investigation of source and item memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(4), 1005–1025.
- Shiffrin, R. M., Ratcliff, R., & Clark, S. E. (1990). List-strength effect: II. theoretical mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(2), 179–195.
- Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review*, 4(2), 145–166.
- Shiffrin, R. M., & Steyvers, M. (1998). The effectiveness of retrieval from memory. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 73–95). Oxford, UK: Oxford University Press.
- Smith, P. L. (2000). Stochastic dynamic models of response time and accuracy: A foundational primer. *Journal of Mathematical Psychology*, 44(3), 408–463.
- Smith, P. L. (2010). From Poisson shot noise to the integrated Ornstein–Uhlenbeck process: Neurally principled models of information accumulation in decision-making and response time. *Journal of Mathematical Psychology*, 54, 266–283.
- Smith, P. L., Ratcliff, R., & Wolfgang, B. J. (2004). Attention orienting and the time course of perceptual decisions: Response time distributions with masked and unmasked displays. *Vision Research*, 44, 1297–1320.
- Smith, P. L., & Van Zandt, T. (2000). Time-dependent Poisson counter models of response latency in simple judgment. *British Journal of Mathematical and Statistical Psychology*, 53, 293–315.
- Smith, P. L., & Vickers, D. (1988). The accumulator model of two-choice decision. *Journal*

- of Mathematical Psychology*, 32, 135–168.
- Speer, N. K., & Curran, T. (2007). ERP correlates of familiarity and recollection processes in visual associative recognition. *Brain Research*, 1174, 97–109.
- Starns, J. J., & Ratcliff, R. (2014). Validating the unequal-variance assumption in recognition memory using response time distributions instead of ROC functions: A diffusion model analysis. *Journal of Memory and Language*, 70, 36–52.
- Starns, J. J., Ratcliff, R., & McKoon, G. (2012). Evaluating the unequal-variance and dual-process explanations of zROC slopes with response time data and the diffusion model. *Cognitive Psychology*, 2012(1), 1–34.
- Starns, J. J., Rotello, C. M., & Hautus, M. J. (2014). Recognition memory zROC slopes for items with correct versus incorrect source decisions discriminate the dual process and unequal variance signal detection models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(5), 1205–1225.
- Starns, J. J., White, C. N., & Ratcliff, R. (2010). A direct test of the differentiation mechanism: REM, BCDMEM, and the strength-based mirror effect in recognition memory. *Journal of Memory and Language*, 63(1), 18–34.
- Sternberg, S. (1966). High speed scanning in human memory. *Science*, 153, 652–654.
- Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, 30, 276–315.
- Steyvers, M., & Malmberg, K. J. (2003). The effect of normative context variability on recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(5), 760–766.
- Stone, M. (1960). Models for choice-reaction time. *Psychometrika*, 25, 251–260.
- Swensson, R. G., & Green, D. M. (1977). On the relations between random walk models

- for two-choice response times. *Journal of Mathematical Psychology*, 15, 282–291.
- Teodorescu, A. R., & Usher, M. (2013). Disentangling decision models: From independence to competition. *Psychological Review*, 120(1), 1–38.
- Thura, D., Beauregard-Racine, J., Fradet, C.-W., & Cisek, P. (2012). Decision making by urgency gating: Theory and experimental support. *Journal of Neurophysiology*, 108, 2912–2930.
- Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. New York: Cambridge University Press.
- Tsetsos, K., Usher, M., & McClelland, J. L. (2011). Testing multi-alternative decision models with non-stationary evidence. *Frontiers in Neuroscience*, 5, 1–18.
- Tuerlinx, F. (2004). The efficient computation of the cumulative distribution and probability density functions in the diffusion model. *Behavior Research Methods, Instruments, & Computers*, 36(4), 702–716.
- Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, 26(1), 1–12.
- Tulving, E., & Wiseman, S. (1975). Relation between recognition and recognition failure of recallable words. *Bulletin of the Psychonomic Society*, 6(1), 79–82.
- Turner, B. M., Forstmann, B. U., Wagenmakers, E.-J., Brown, S. D., Sederberg, P. B., & Steyvers, M. (2013). A Bayesian framework for simultaneously modeling neural and behavioral data. *NeuroImage*, 72, 193–206.
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, 108(3), 550–592.
- Van Zandt, T. (2000). ROC curves and confidence judgments in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(3), 582–600.

- Vickers, D. (1970). Evidence for an accumulator model of psychophysical discrimination. *Ergonomics*, 13(1), 37–58.
- Wagenmakers, E.-J., Steyvers, M., Raaijmakers, J. G. W., Shiffrin, R. M., van Rijn, H., & Zeelenberg, R. (2004). A model for evidence accumulation in the lexical decision task. *Cognitive Psychology*, 48, 332–367.
- Wald, A. (1945). Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics*, 16(2), 117–186.
- Wald, A. (1947). *Sequential analysis*. New York: Wiley.
- White, C. N., Ratcliff, R., & Starns, J. J. (2011). Diffusion models of the flanker task: Discrete versus gradual attentional selection. *Cognitive Psychology*, 63, 210–238.
- Wickelgren, W. A., & Corbett, A. T. (1977). Associative interference and retrieval dynamics in yes-no recall and recognition. *Journal of Experimental Psychology: Human Learning and Memory*, 3(2), 189–202.
- Wilding, E. L., & Rugg, M. D. (1997). Event-related potentials and the recognition memory exclusion task. *Neuropsychologia*, 35(2), 119–128.
- 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), 1341–1354.
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, 46, 441–517.
- Zhang, S., Lee, M. D., Vandekerckhove, J., Maris, G., & Wagenmakers, E.-J. (2014). Time-varying boundaries for diffusion models of decision making and response time. *Frontiers in Psychology*, 4, 1364.

Curriculum vitae

Education

2009–2015 Indiana University, Bloomington

Ph.D., Psychological & Brain Sciences, Cognitive Science

2005–2009 University of Maryland, College Park

B.M., Music Composition

B.S., Psychology

Publications

Refereed journal articles

Nosofsky, R. M., Cao, R., **Cox, G. E.**, & Shiffrin, R. M. (2014). Familiarity and categorization processes in memory search. *Cognitive Psychology*, 75, 97–129.

Nosofsky, R. M., **Cox, G. E.**, Cao, R., & Shiffrin, R. M. (2014). An exemplar-familiarity model predicts short-term and long-term probe recognition across diverse forms of memory search. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(6), 1524–1539.

Nokes-Malach, T. J., VanLehn, K., Belenky, D. M., Lichtenstein, M., & **Cox, G. E.** (2013). Coordinating principles and examples through analogy and

self-explanation. *European Journal of Psychology of Education*, 28(4), 1237–1263.

Cox, G. E., & Shiffrin, R. M. (2012). Criterion setting and the dynamics of recognition memory. *Topics in cognitive science*, 4(1), 135–150.

Cox, G. E., Kachergis, G., Recchia, G., & Jones, M. N. (2011). Toward a scalable holographic word-form representation. *Behavior Research Methods*, 43(3), 602–615.

Refereed conference proceedings

Cox, G. E., & Shiffrin, R. M. (2015). A dynamic approach to secondary processes in associative recognition. In D. Noelle et al. (Eds.), *Proceedings of the 37th annual conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Cox, G. E., & Shiffrin, R. M. (2014). Linking memory models and decision models: Insights on frequency and speed/accuracy trade-off. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th annual conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Cox, G. E., Lewis, N. J., & Shiffrin, R. M. (2013). On the dynamics of information accumulation in recognition. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Cox, G. E., Kachergis, G., & Shiffrin, R. M. (2013). Similarity and strategic effects in recognition memory. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Kachergis, G., **Cox, G. E.**, & Shiffrin, R. M. (2013). The effects of repeated sequential context on recognition memory. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Cox, G. E., Kachergis, G., & Shiffrin, R. M. (2012). Gaussian process regression for trajectory analysis. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th annual conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Cox, G. E., & Shiffrin, R. M. (2012). Exploring decision rules and sampling dynamics in recognition memory. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th annual conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Cox, G. E., & Shiffrin, R. M. (2011). Criterion setting and the dynamics of recognition memory. In L. Carlson, C. H. ölscher, & T. Shipley (Eds.), *Proceedings of the 33rd annual conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Kachergis, G., **Cox, G. E.**, & Jones, M. N. (2011). OrBEAGLE: Integrating orthography into a holographic model of the lexicon. In *Artificial Neural Networks and Machine Learning—ICANN 2011*. Berlin Heidelberg: Springer.

Cox, G. E. (2010). On the relationship between entropy and meaning in music: An exploration with recurrent neural networks. In S. Ohisson and R. Catrambone (Eds.), *Proceedings of the 32nd annual conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Cox, G. E., Harbison, J. I., & Davelaar, E. J. (2008). From list learning to semantic knowledge: Search and learning of associative memory. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Awards and honors

2015 Robert J. Glushko and Pamela Samuelson Foundation Student Travel Grant,
Cognitive Science Society

2015 Outstanding Research Award, Cognitive Science Program, Indiana University,
Bloomington

2014 William K. Estes Summer Research Award, Department of Psychological & Brain
Sciences, Indiana University, Bloomington

2012 Robert J. Glushko and Pamela Samuelson Foundation Student Travel Grant,
Cognitive Science Society

2011 Computational Modeling Prize for Higher-Level Cognition, Cognitive Science
Society

2009 Mark S. Harper Award for Excellence in Psychology by an Undergraduate,
University of Maryland, College Park

2005–2009 Banneker/Key Scholarship, University of Maryland, College Park

2005–2009 National Merit Scholarship, University of Maryland, College Park

2005–2009 Maryland Distinguished Scholar, University of Maryland, College Park

2005–2009 Dean's List, University of Maryland, College Park

2007–2008 Keith Steyer Scholarship, University of Maryland, College Park

2005 American Society of Mechanical Engineers Citation, Gaithersburg, Maryland

Grants and fellowships

2010–2013 Graduate Research Fellowship, National Science Foundation

2009 Graduate Fellowship, Indiana University, Bloomington

Invited talks

2015 A Dynamic Approach to Recognition Memory. University of Massachusetts, Amherst, February 25.

2014 Unpacking Models of Response Time: A Case Study from Recognition Memory (in Workshop on Sequential Sampling Models for Cognitive and Perceptual Decision Making). Annual Conference of the Cognitive Science Society, July 23–26.

2014 An Introduction to Gaussian Processes in Psychology (with Joseph W. Haupt and Richard M. Shiffrin). Annual Meeting of the Society for Mathematical Psychology, July 18–21.

Conference presentations

2015 A Dynamic Approach to Secondary Processes in Associative Recognition. Annual Conference of the Cognitive Science Society, July 23–25.

- 2015** The Dynamics of Single Item and Associative Recognition. Annual Meeting of the Society for Mathematical Psychology, July 17–21.
- 2015** Hierarchical Bayesian Modeling of Item-Level Effects in Item and Associative Memory. Annual Meeting of the Society for Mathematical Psychology, July 17–21.
- 2015** A Dynamic Approach to Item and Associative Recognition. Annual Summer Interdisciplinary Conference, July 10–15.
- 2015** A Dynamic Approach to Item and Associative Recognition. Context and Episodic Memory Symposium, May 7–8.
- 2014** Linking Memory Models and Decision Models: Insights on Frequency and Speed/Accuracy Trade-off. Annual Conference of the Cognitive Science Society, July 23–26.
- 2014** Mechanisms of Recognition Memory: Insights from a Dynamic Approach. Annual Meeting of the Society for Mathematical Psychology, July 18–21.
- 2014** Modeling Information Accumulation in Recognition Memory: Speed-Accuracy Functions and Response Time Distributions. Annual Summer Interdisciplinary Conference, June 23–28.
- 2014** Insights from a Dynamic Approach to Recognition: Word Frequency, Speed/Accuracy Trade-off, and Decision Mechanisms. Context and Episodic Memory Symposium, May 8–9.
- 2013** Effects of Perceptual Similarity and Encoding Strategy in Recognition Memory. Annual Meeting of the Society for Mathematical Psychology, August 4–7.

- 2013** Similarity and Strategic Effects in Recognition Memory. Annual Conference of the Cognitive Science Society, July 31–August 3.
- 2013** Dynamics of Information Accumulation in Recognition. Annual Summer Interdisciplinary Conference, July 24–30.
- 2013** Effects of Perceptual Similarity and Encoding Strategy in Recognition Memory. Midwest Cognitive Science Conference, May 18.
- 2013** Information Accumulation for Recognition: Dynamic Presentation and Diagnosticity. Context and Episodic Memory Symposium, May 9–10.
- 2012** The Dynamics of Sentence (In)comprehension. Annual Conference of the Cognitive Science Society, August 1–4.
- 2012** Dynamic Effects of Perceptual and Categorical Similarity on Recognition Memory. Annual Conference of the Cognitive Science Society, August 1–4.
- 2012** Gaussian Process Regression for Trajectory Analysis. Annual Conference of the Cognitive Science Society, August 1–4.
- 2012** Exploring Decision Rules and Sampling Dynamics in Recognition Memory. Annual Conference of the Cognitive Science Society, August 1–4.
- 2012** Priming and Fluency within a Dynamic Model of Recognition. Annual Meeting of the Society for Mathematical Psychology, July 21–24.
- 2012** Analyzing Trajectories with Gaussian Process Regression. Annual Summer Interdisciplinary Conference, June 7–12.

- 2012** A Dynamic Model for Fluency. Annual Summer Interdisciplinary Conference, June 7–12.
- 2012** Modeling of Recognition Decision Making. Context and Episodic Memory Symposium, May 10.
- 2012** Gaussian Process Modeling of Behavioral Trajectories. Midwest Cognitive Science Conference, May 7.
- 2011** Criterion Setting and the Dynamics of Recognition Memory. Annual Conference of the Cognitive Science Society, July 20–23.
- 2011** Modeling Recognition of Different Stimulus Classes with a Dynamic Activation Model. Annual Meeting of the Society for Mathematical Psychology, July 15–18.
- 2011** Modeling Recognition of Different Stimulus Classes with a Dynamic Activation Model. Annual Summer Interdisciplinary Conference, July 7–12.
- 2011** A Dynamic Model for Recognition Memory: Towards a Solution to the Problem of Criterion Setting. Midwest Cognitive Science Meeting, April 16.
- 2010** Towards a Scalable Holographic Word-Form Representation. Annual Meeting of the Society for Computers in Psychology, November 18.
- 2010** On the Relationship Between Entropy and Meaning in Music: An Exploration With Recurrent Neural Networks. Annual Conference of the Cognitive Science Society, August 11–14.
- 2010** Developing Knowledge and Probing Its Structure with Primed Lexical Decision. Annual Meeting of the Society for Mathematical Psychology, August 7–10.

- 2010** On the Formation of Associative Structure in Knowledge. Annual Summer Interdisciplinary Conference, July 31–August 5.
- 2009** The Role of Information Ecology in Judgment: A Computational Account of Expertise Bias. Annual Meeting of the Society for Mathematical Psychology, August 1–4.
- 2008** From List Learning to Semantic Knowledge: Search and Learning of Associative Memory. Annual Conference of the Cognitive Science Society, July 23–26.

Departmental talks

- 2014** Insights from a Dynamic Approach to Recognition Memory. Cognitive Lunch talk series, March 26.
- 2014** Different Windows on the Hypothesized Bilingual Advantage. Grey Matters talk series, February 10.
- 2012** Gaussian Process Regression for Dummies, or, Rich's Revenge: How I Learned to Stop Worrying and Love the Kernel. Cognitive Lunch talk series, November 7.
- 2012** The Varieties of Musical Experience: Music and Information Theory. Grey Matters talk series, February 13.
- 2011** A Dynamic Model for Recognition Memory. Cognitive Lunch talk series, October 5.

Teaching Experience

- 2012** Experimental Methods in Psychology, Associate Instructor.

Professional Service

2015–Present Reviewer, *Canadian Journal of Experimental Psychology*

2014–Present Reviewer, *Psychological Review*

2014–Present Reviewer, *Psychonomic Bulletin and Review*

2014–Present Reviewer, *Quarterly Journal of Experimental Psychology*

2011–Present Reviewer, Proceedings of the Annual Conference of the Cognitive Science Society.

2012–2013 Organizer, Indiana University Cognitive Lunch talk series.

Professional Employment

2009–2014 Indiana University

Graduate Research Assistant

2007–2009 University of Maryland, Decision, Attention, and Memory Laboratory

Research Assistant

2008 University of Pittsburgh, Cognitive Science Learning Laboratory

Research Assistant (via Pittsburgh Science of Learning Center internship)