

A Dynamic Approach to Recognition Memory

Gregory E. Cox
Syracuse University

Richard M. Shiffrin
Indiana University

Abstract

We present a dynamic model of memory that integrates the processes of perception, retrieval from knowledge, retrieval of events, and decision making as these evolve from one moment to the next. The core of the model is that recognition depends on tracking changes in familiarity over time from an initial baseline generally determined by context, with these changes depending on the availability of different kinds of information at different times. A mathematical implementation of this model leads to precise, accurate predictions of accuracy, response time, and speed-accuracy trade-off in episodic recognition at the levels of both groups and individuals across a variety of paradigms. Our approach leads to novel insights regarding word frequency, speeded responding, context reinstatement, short-term priming, similarity, source memory, and associative recognition, revealing how the same set of core dynamic principles can help unify otherwise disparate phenomena in the study of memory.

Keywords: Recognition memory; response time; episodic memory; mathematical modeling.

Introduction

Memory for specific events—typically called “episodic” memory—is a crucial component of cognition that is engaged in a variety of situations. Arguably one of the simplest expressions of episodic memory is “recognition”, the feeling and/or realization that one has experienced a particular event, where an event consists of some set of perceptual and/or conceptual *content* occurring in a particular *context*, that is, a specific situation and span of time. Despite the apparent simplicity

This work was supported by the National Science Foundation through a Graduate Research Fellowship to Gregory E. Cox and by Air Force Office of Scientific Research grant FA9550-12-1-0255 to Richard M. Shiffrin.

Portions of these ideas in development were presented in meetings of the Cognitive Science Society and Society for Mathematical Psychology, as well as the Annual Summer Interdisciplinary Conference and Context and Episodic Memory Symposium in each of the years 2011–2015, inclusive.

Correspondence concerning this article should be addressed to Gregory E. Cox, Department of Psychology, 430 Huntington Hall, Syracuse University, Syracuse, NY 13244-2340. E-mail: gregcox7@gmail.com.

of recognition and the fact that it is usually achieved without considerable effort (compare the ease with which one recognizes the face of someone one has seen before with the difficulty of recalling that person's name), it nonetheless involves a variety of processes, including detecting the relevant perceptual information in the environment, retrieving conceptual information from knowledge, joining content features with those of the target context in working memory, and selecting and executing an appropriate response. Much prior work has focused on treating these processes in isolation, focusing solely on their outputs. In this article, we focus instead on how these processes and their interactions unfold over time—a so-called “dynamic” approach. We develop a model of episodic memory tasks requiring the retrieval of events, focusing primarily on recognition. The dynamic aspects of recognition are especially important given the relative ease and rapidity with which it is usually accomplished, and they are revealed by experimental tasks in which there is incentive to retrieve quickly, such as response time or response signal experiments. Before laying out our approach in detail, we briefly review some key theoretical issues with which any model of memory must contend.

Representation and interference

Memory theories must specify, whether in a formal way (our focus in this article) or at the neural level, exactly how event memories are represented. In other words, they must answer the questions, what is the information stored in memory about an event? and how does this information relate to or overlap with that stored for other events? Addressing these questions is often framed in terms of interference between different memory “traces”, where traces contain information about the features or aspects of an event (Underwood, 1969). Events are, in general, defined by their content—the perceptual and conceptual features of the episode—and their context—their time and location, as well as the relations (e.g., schematic or causal) with other events (e.g., Zacks & Tversky, 2001). If the presence of a particular trace in event memory leads to interference at retrieval, this is evidence that they are represented using similar features, leading the two traces to be confused with one another. These features may be related to either content or context. When a word is falsely recognized by virtue of its perceptual or semantic similarity to words that had actually been encountered, this is evidence that the *content* of the memory traces formed from studying those related words is encoded using the relevant semantic or perceptual features (Deese, 1959; Roediger & McDermott, 1995; Shiffrin, Huber, & Marinelli, 1995). When studying additional material in the same session yields impaired recognition for that material (Murnane & Shiffrin, 1991), this implies that the *context* for all of that material was encoded using the same features. In general, interference can arise as a result of overlap in either content or context (Dennis & Humphreys, 2001; Criss & Shiffrin, 2004a).

The features used to encode both the content and context of an event arise from both the perceptual aspects of the environment and prior knowledge. This is evident in the case of encoding a word in terms of both the contours that depict its visual form (a set of relatively low-level perceptual features) as well as its semantic features, which must be learned over time (Logan, 1988; Rogers & McClelland, 2004). Semantic knowledge is thus less tied to a particular context than are memories for specific episodes, yet that knowledge is crucial for encoding the content of an episode (McClelland, McNaughton, & O'Reilly, 1995; Norman & O'Reilly, 2003; Kumaran & McClelland, 2012; Nelson & Shiffrin, 2013). Conversely, although an individual's present context influences what knowledge is easily accessible (Anderson & Schooler, 1991), context plays a much more central role in event recognition (Klein, Shiffrin, & Criss, 2007).

Non-dynamic models of event recognition

Given that memory contains traces which contain information about the content and context of experienced events, a variety of non-dynamic models of event recognition have been proposed (for reviews from various perspectives, the reader is directed to Clark & Gronlund, 1996; Diana, Reder, Arndt, & Park, 2006; Malmberg, 2008; Kahana, 2012). These models can be broadly divided into two classes. One class assumes a single-valued metric for the recognition decision (e.g., Metcalfe & Shimamura, 1994; Murdock, 1982; Gillund & Shiffrin, 1984; Hintzman, 1988; Shiffrin & Steyvers, 1997; McClelland & Chappell, 1998; G. D. A. Brown, Preece, & Hulme, 2000; Dennis & Humphreys, 2001), akin to signal detection (Green & Swets, 1966). According to these theories, the content and context of a test event (e.g., a single word after having studied a list of words) is compared to what was stored in memory, leading to a value of “familiarity” or memory strength. If the value of familiarity on a particular test exceeds a criterion, the participant gives a positive response, indicating that they are willing to say they experienced that item in the target context. Otherwise, they reject the item. Different theories in this class specify how various factors lead to the value of familiarity experienced on a particular test trial.

A second subclass of recognition theories, such as dual-process theory (e.g., Atkinson & Juola, 1974; Mandler, 1980; Jacoby, 1991; Yonelinas, 1994; Reder et al., 2000; Yonelinas, 2002; Mandler, 2008), augments this familiarity-based process with a secondary process that is presumed to operate similarly to recall. Different models assume that the secondary process is at least partially correlated with the familiarity process and remains error-prone (Reder et al., 2000; Malmberg, 2008), or is independent of familiarity and leads to perfect accuracy when engaged (Yonelinas, 1994). Various efforts have been made to assess the presence or degree of recall-like processes in recognition tasks on the basis of receiver-operating characteristic functions (Yonelinas, 1994) and process-dissociation procedures (Tulving, 1985; Jacoby, 1991). However, evidence from these studies remains equivocal with respect to the need for a qualitatively different secondary process in normal recognition tasks (Curran & Hintzman, 1995; Ratcliff, Van Zandt, & McKoon, 1995; W. Donaldson, 1996; Hillstrom & Logan, 1997; Hirshman & Master, 1997; Humphreys, Dennis, Chalmers, & Finnigan, 2000; Dunn, 2004; Wixted, 2007; Dunn, 2008), at least in the absence of explicit models of the two processes (e.g., Reder et al., 2000; Malmberg, 2008). As we argue below, however, processes that are not easily dissociated in terms of their final outcomes may be better identified by the dynamics that lead to those outcomes.

Evidence accumulation

Although most memory models focus on the final outcome of retrieval, much research in decision making has focused on the dynamics of the processes leading to a final outcome. Specifically, they consider these processes to be a form of evidence accumulation in which information for or against different outcomes—like a positive or negative recognition decision—is collected and weighed. The cornucopia of models developed as part of this research program differ in the degree to which the accumulated evidence is noisy, in the number of separate accumulation processes, and in how evidence gets weighed (e.g., Stone, 1960; LaBerge, 1962; Edwards, 1965; Vickers, 1970; Pike, 1973; Link, 1975; Link & Heath, 1975; Townsend & Ashby, 1983; Smith & Vickers, 1988; Busemeyer & Townsend, 1993; Smith & Van Zandt, 2000; Van Zandt, 2000; Usher & McClelland, 2001). Two of the most common approaches, both of which have been applied to data from recognition memory paradigms, are the diffusion models of Ratcliff and colleagues (e.g., Ratcliff, 1978)

and the linear ballistic accumulator models of S. Brown and Heathcote (2005, 2008). Crucially, although these models account for the dynamics of *accumulating* evidence, they are agnostic with regard to the processes that *generate* evidence. Just like the recognition theories reviewed above, they treat only the final outcome of the retrieval processes that generate evidence for a decision, only now the final outcome is not (typically) a binary variable, but a—still static—distribution of memory evidence.

Overview

In this article, we develop a dynamic approach to recognition memory that explicitly models the structure of memory as well as the way retrieval evolves over time to lead to a decision. Our approach integrates the key theoretical aspects of recognition reviewed above into a continuous flow:

- Events are encoded in memory as traces that contain both content and context features.
- Content features are, in part, derived by accessing knowledge.
- The content features of a test item gradually accumulate as perception and knowledge access proceed and joint features of the current context to act as an evolving probe of memory.
- At any given time, the memory probe yields a value of familiarity that changes over time as features enter the probe.
- Participants track these changes in familiarity in order to make recognition decisions.

This approach serves to link concepts in memory with those of learning (via knowledge access and event trace representation) and of decision making (via the accumulation of evolving memory evidence over time). This approach offers a novel perspective on various phenomena in recognition memory, allowing for new ways to test theories of memory beyond those that would be available via a static approach that views the component processes in isolation. In addition, it leads to a mathematical model that accurately predicts the quantitative details of accuracy, response time, and speed-accuracy tradeoff in recognition memory. After providing a conceptual overview of our approach, we describe the technical details of a model that implements these ideas and use the core properties of this model as a starting point to explore the dynamics of recognition memory.

A Dynamic Model for Recognition Memory

The dynamic approach to recognition memory that we adopt is an outgrowth of our earlier work (Cox & Shiffrin, 2012) and shares many goals with previous efforts to link retrieval and decision making (Ratcliff, 1978; Hockley & Murdock, 1987; Nosofsky & Palmeri, 1997; Diller, Nobel, & Shiffrin, 2001), particularly with those that focus on the time-course with which perceptual and conceptual information become available (Brockdorff & Lamberts, 2000; Cohen & Nosofsky, 2003). We begin by describing the general framework for how perception and retrieval evolve over time to provide evidence for a recognition decision. We then formalize this within a mathematical model that is capable of predicting the details of response time and accuracy as well as speed-accuracy trade-off. Beyond the qualitative properties of our conceptual framework, the model also allows for quantitative tests of particular hypotheses. Although alternative formulations of the model

might be possible without altering the core of our framework (many possible variants are discussed later), we believe the present form makes clear the principles and mechanisms behind our dynamic approach to recognition.

Conceptual framework

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. We begin by providing a conceptual overview of our model, depicted schematically in Figure 1, to which we will refer in the subsequent description. We emphasize that while we distinguish between these various components of the recognition process, we do *not* assume that they operate in strictly serial fashion; instead, they operate in a continuous flow across time.

Structure of memory traces. Prior to the recognition test, studying a list of items leads to the formation of a set of traces in memory. One trace is formed for each different 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 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 explicitly model it in this work, 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).

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 can typically be increased by increasing the time available to study the item via either duration or repetition. 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). If a feature is stored in the trace, its value is

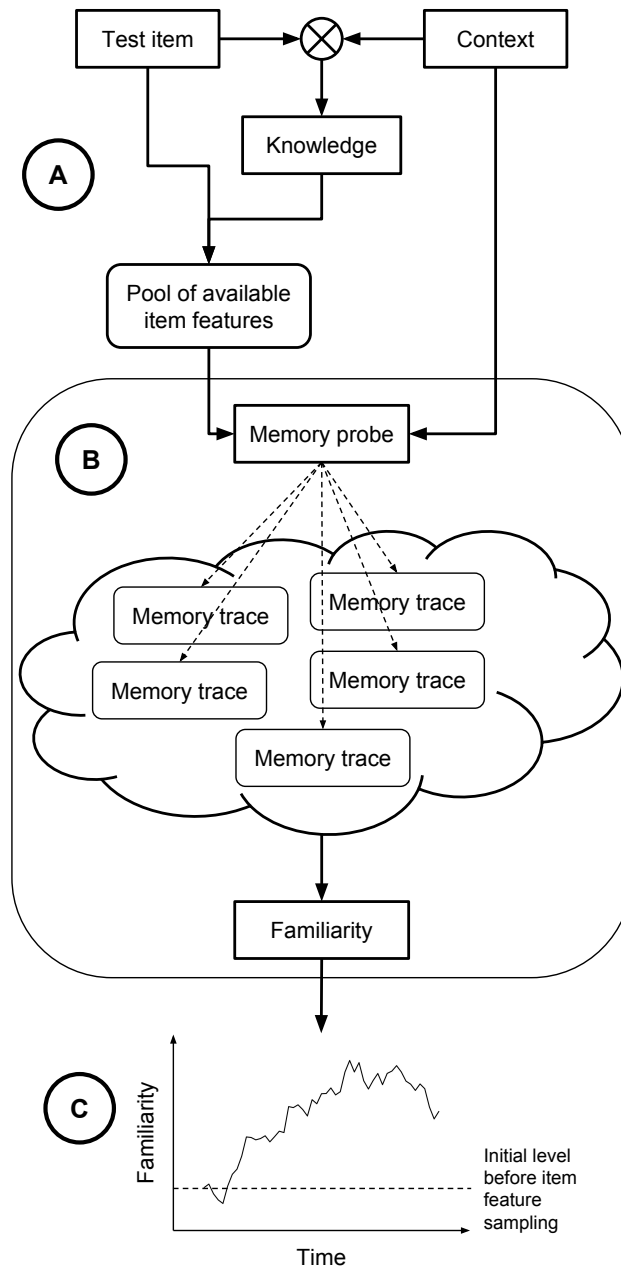


Figure 1. Schematic depiction of the recognition process. A memory probe is formed from a combination of the current context and both physical and semantic features of the test item (A). Item features are gradually sampled into the probe which is continually compared in parallel to all the traces in memory, each of which is activated to a degree depending on the similarity between the features in the probe and those in the trace (B). To make a recognition decision, a participant tracks changes in the overall level of activation of these traces—termed “familiarity”—from an initial level determined by context alone (C).

stored correctly in the trace with probability c_S . Otherwise, the stored value is sampled at random according to the environmental distribution of values for that feature. This distribution is presumed to reside in semantic memory, given that it represents information accrued over a long period of time.

Accessing knowledge. When a test item is presented, a process similar to that used to encode a memory trace is engaged to construct a memory probe in short-term memory, corresponding to region A in Figure 1. The physical form of the test item in conjunction with context are used to access knowledge (for a word, this would be a lexical trace in semantic memory). The physical features of the item arising from early perception as well as the semantic features of the item arising from knowledge thus form a pool of available features with which to probe memory. Features may enter the pool at different rates, with certain physical features (e.g., brightness or contour) entering early and other more conceptual features (e.g., configural aspects) entering late. We assume that the same capacity limits that apply during study also apply at test, with a maximum of N_X context and N_C content features able to be held in the probe.

Probing event memory. In typical recognition tests, the memory probe initially contains only the N_X features of the current context, which in most tasks is identical to the target study context (one setting where this is not true is in a list discrimination task, discussed later). After a test item has been presented, content features are gradually sampled from the available pool and join the context features to form an accumulating probe of memory. At any given time, the probe is compared in parallel to the traces in episodic memory, corresponding to region B in Figure 1, producing similarity-based activation levels for each trace characterized as likelihood ratios (see below). The overall degree of activation across event memory—which we refer to as “familiarity”—at any given time is the average likelihood ratio over all traces that are sufficiently activated at that time.

Most of the traces in memory will tend to match the probe very poorly, of course, so we presume that only those traces whose activation exceeds a threshold value actually contribute to the 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) historical 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.

Trace activation and familiarity over time. Familiarity evolves over time as item features enter the memory probe from the pool of available features. Participants track these changes in familiarity relative to an initial baseline level defined solely by context, before any item features had been sampled. 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. Recognition decisions are based on accumulating changes in familiarity from this initial value. 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 that the test stimulus is “new”. On any one trial, the way familiarity evolves 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. On average, however, each new feature that enters the probe provides more statistical evidence about

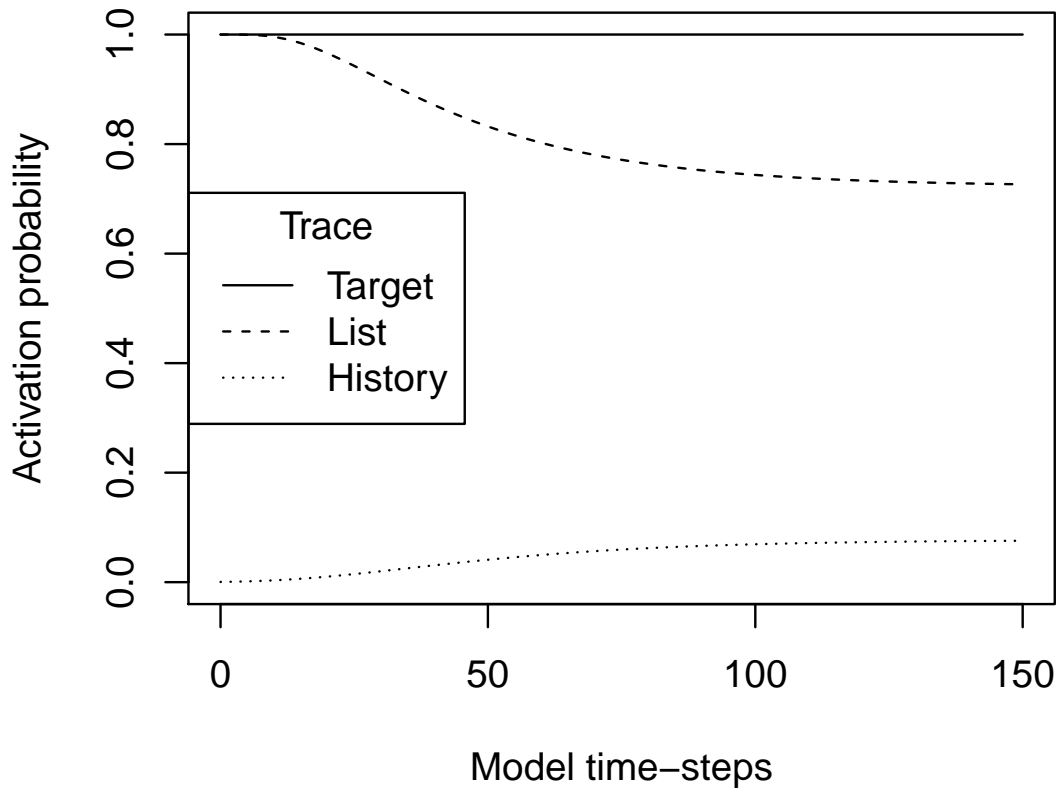


Figure 2. Illustration of how different kinds of memory traces can remain active over time (like target traces), be activated as more content features join the memory probe (historical traces of the test item from prior to the experiment), or be deactivated as an initial match to context is overwhelmed by mismatching content features (other traces from the recent study list).

whether there is a trace in memory that encodes the same event as the probe, depending on whether that feature leads to an increase or decrease in overall familiarity.

Some historical traces that match in content but not in context, such as traces formed of the test item prior to the experiment, may become active as their matching content features overcome their initially mismatching context features. On the other hand, list traces formed by other items than the one being tested—which match in context but not in content—will tend to deactivate as more and more content features fail to match. If the item tested is a target, there will usually be a single trace in memory that therefore matches in *both* content and context; this target trace will tend to get even more activated beyond its initial level based on context alone. The result of these changes in trace activation over time (Figure 2) is that, as shown in the central panel of Figure 4, the average familiarity for target tests will increase over time whereas that for foil tests will decrease over time. 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).

Table 1

A summary of the parameters of the model of recognition dynamics.

Parameter	Description
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.
\bar{T}_R	Mean residual time attributable to stimulus detection and response execution.
σ_R	Standard deviation of residual time.
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.

In typical free-response situations, a participant makes a "yes" or "no" recognition response when the accumulated change in familiarity on that trial reaches either an upper or lower criterion, respectively. When participants must instead wait for a signal before they respond, they track changes in familiarity over time until the signal and make a "yes" response if the accumulated change is above a criterion, otherwise they respond "no".

Mathematical model

We now proceed to describe the technical details of how an implementation of this model, with topic headings corresponding to that for the conceptual framework above. For reference, the parameters of our model are summarized in Table 1 while other model notation is summarized in Table 2 and we present a numerical example of these technical details in Figure 3.

Structure of memory traces. Each studied item forms a single trace in memory consisting of a maximum of N_C content features and N_X context features. For simplicity, we assume all features are binary and have an environmental distribution governed by parameter g which reflects the probability that any two independent draws will yield the same outcome, which is typically assumed to be $\frac{1}{2}$ for all features (this peculiar parametrization actually simplifies the resulting computations). All N_X context features in a memory trace have a stored value, but any given content feature is only stored with probability u . With probability c_S , a feature (whether content or context) is stored correctly, otherwise the stored value is a sample from the environmental base-rate distribution.

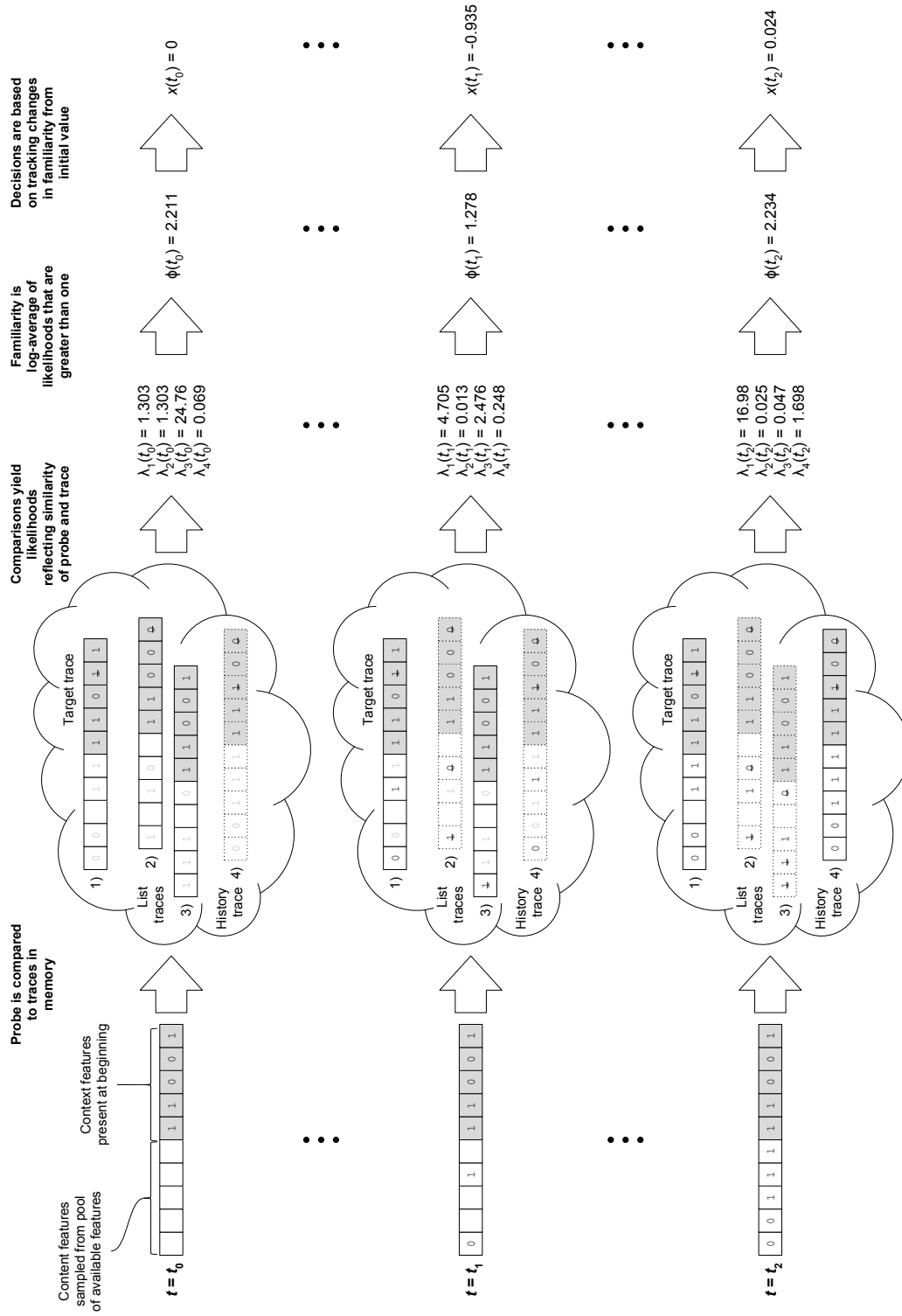


Figure 3. A numerical example showing how memory trace activations and familiarity are computed and tracked over time.

Table 2

Notational reference.

Notation	Description
$\lambda_i(t)$	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 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 8.
$x(t)$	Accumulated change in familiarity from starting time t_0 , given by Equation 9. This is the evidence used for recognition decisions.
$v(t)$	Probability that all content features have been sampled into the probe by time t , assuming they are all available at the start of the trial, given by Equation 1.
$v^*(t)$	Participant’s estimated probability that all desired content features have been sampled into the probe by time t , given by Equation 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 11. When $x(t)$ crosses one of these boundaries, the corresponding decision is made.

Accessing knowledge. The time at which a test stimulus is presented is called $t = 0$. At this time, the memory probe consists only of the N_X features of the current context. After some delay, content features are available to enter the probe from the pool of available features at time t_0 . Content features are sampled from the pool over time: At any given time after t_0 , one of the N_C content features is chosen uniformly at random. If that feature is not yet in the available pool, then nothing new enters the probe. If, however, the feature is available and the probe does not yet have a value for that feature, it enters the probe. With probability c_T , the correct value enters the probe; otherwise, a value is chosen at random according to the environmental base-rate distribution 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. In case all N_C content features are in the available pool from the beginning of the trial, the probability that the probe is complete at time t is

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

Probing event memory. At each time 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. For practical reasons, we only explicitly model comparisons with the N traces formed during list study 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 (the context features for history traces are sampled at random). Because the many other traces in memory tend not to match at all in either

content or context, they are unlikely to be activated and are not explicitly modeled.

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), when lists consist of items that are of widely different kinds that are unlikely to be encoded using the same features.

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

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

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

$$p_{M|D} = g, \quad (3)$$

where g is the base-rate probability of a match as described above. The probability of a mismatch (N) given that probe and trace encode the same item is

$$p_{N|S} = c_S(1 - c_T)(1 - g) + (1 - c_S)c_T(1 - g) + 2(1 - c_S)(1 - c_T)g(1 - g), \quad (4)$$

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, again, the base-rate

$$p_{N|D} = 1 - g. \quad (5)$$

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

$$\lambda_i(t) = \left(\frac{p_{M|S}}{p_{M|D}} \right)^{N_M(t)} \left(\frac{p_{N|S}}{p_{N|D}} \right)^{N_N(t)}, \quad (6)$$

which simplifies to

$$\lambda_i(t) = \left[1 + c_S \left(\frac{1-g}{g} \right) \right]^{N_M(t)} [1 - c_S]^{N_N(t)} \quad (7)$$

when $c_T = 1$.

A memory trace i is considered “active” at time t if its likelihood ratio $\lambda_i(t) > 1$. This fixed threshold means that an active trace matches sufficiently well in *either* content or context features to make it a plausible candidate to be related to the test item. It also means that only a relatively small subset of the many (effectively infinite) traces in event memory will take part in the recognition process. Familiarity $\phi(t)$ is then defined as the logarithm of the average likelihood ratio over these active traces:

$$\phi(t) = \log \langle \lambda_i(t) | \lambda_i(t) > 1 \rangle. \quad (8)$$

Taking the logarithm means that $\phi(t)$ can be considered a kind of “log-odds” that there exists a trace in the activated subset of memory that encodes the same event as the probe (Shiffrin & Steyvers, 1997). In addition, the logarithm yields the computational advantage that the dynamics of familiarity—how it changes over time—are linear rather than multiplicative, as we describe next (cf. Edwards, 1965). Although it would be possible to reformulate the model without the logarithmic transformation, in our opinion it would make it somewhat harder to understand the behavior of other model quantities (like response boundaries).

Trace activation and familiarity over time. Initially, the probe only contains context features, so the familiarity at time t_0 reflects the overall 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. Familiarity can decrease for both target and foil tests with the first few features sampled because, even when there is a well-matching target trace, a few content features are not sufficient to deactivate the many mismatching non-target traces (see Figure 4, the numerical example in Figure 3, and our discussion of priming, below). Eventually, however, familiarity tends to increase for targets and decrease for foils, as shown in the central panel of Figure 4. Recognition decisions are based on the accumulated change in familiarity from the starting time t_0 , denoted $x(t)$:

$$x(t) = \sum_{\tau=t_0}^{t-1} [\phi(\tau+1) - \phi(\tau)] = \phi(t) - \phi(t_0). \quad (9)$$

Fitting the model to data. To allow for efficient quantitative fitting of this model, we developed an approximation described in Appendix A. This approximation allows for the computation of likelihoods for both response-signal and free response data. We verify that model parameters can be correctly recovered via maximum-likelihood fitting in Appendix B.

Response-signal. 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. In these situations, we assume that content features are sampled and join the context features already present in the probe of memory beginning at time t_0 after the onset of the test stimulus. Each model time step takes a constant amount of time ρ . 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 $1/[1 + \exp(\theta)]$, in accord with the assumption that $x(t)$ lies on a log-odds scale. Admittedly, more complex accounts are possible, such as that SAT reflects a mixture of ongoing and already-finished decisions (Meyer, Irwin, Osman, & Kounios, 1988; Ratcliff, 1988), or that additional features can be sampled in the time between the response signal and when the participant actually makes their response (typically around 200 ms after the signal). Nonetheless, our simplified model suffices for present purposes and it is not clear that additional complexity would affect the qualitative forms of any SAT curves we consider.

Free-response. 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. It is possible, however, that the probe becomes saturated with features before the accumulated change reaches a constant boundary. Therefore, we assume that the decision boundaries gradually collapse over time as a function of the expected completeness of the probe, namely,

$$v^*(t) = 1 - \left(1 - \frac{1}{N_C^*}\right)^t, \quad (10)$$

which differs from the equation for the *actual* probe completeness, $v(t)$ in Equation 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 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,

$$\begin{aligned} B_O(t) &= \left(\frac{1}{2} - b\right)A_0 + v^*(t)\frac{A_0}{2} \\ B_N(t) &= \left(\frac{1}{2} - b\right)A_0 - v^*(t)\frac{A_0}{2}. \end{aligned} \quad (11)$$

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 centered around zero, if $b < \frac{1}{2}$ boundaries are “new”-biased, and if $b > \frac{1}{2}$ boundaries are “old”-biased. The process by which familiarity trajectories and response boundaries interact to predict response probabilities and response time distributions is depicted in Figure 4.

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

$$\widehat{RT} = T_R + \rho t \quad (12)$$

where T_R is residual time and ρ is the amount of time per model time-step. This residual time includes not only the time before feature sampling begins (t_0), but also the time needed to detect the test stimulus and execute a motor response. Because of this added variability, we allow T_R to vary from trial to trial according to a Gamma distribution characterized by mean \bar{T}_R and standard

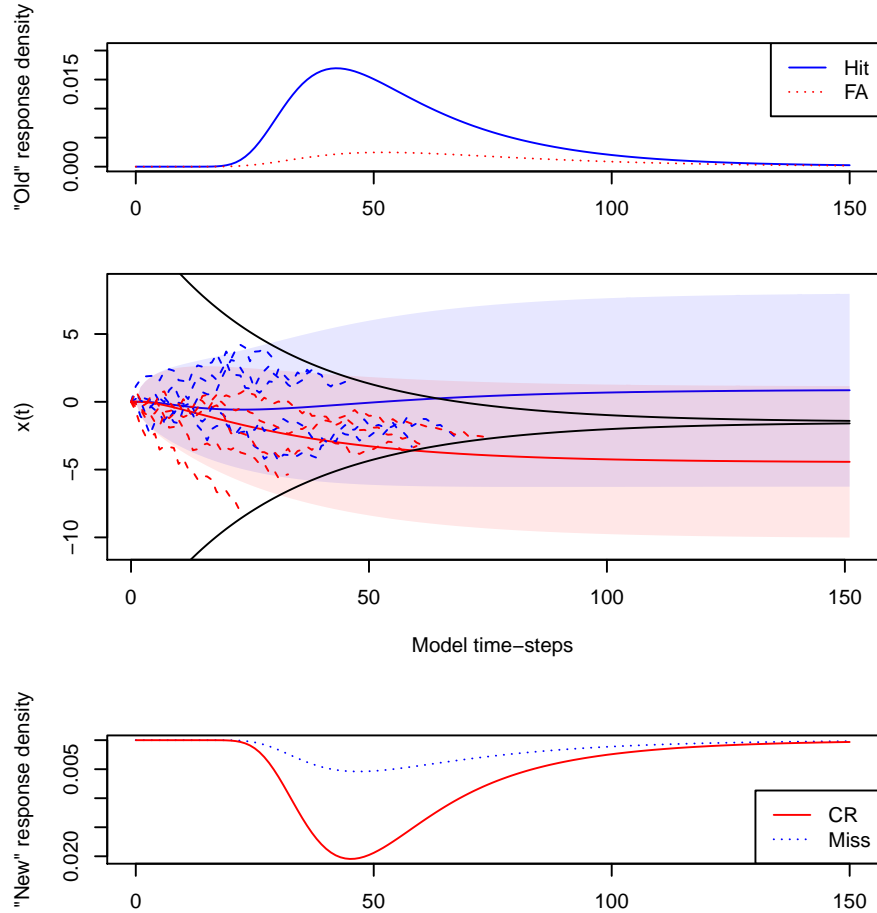


Figure 4. This figure illustrates how recognition decisions are made in free response situations. On any given trial, the accumulated change in familiarity ($x(t)$) takes a noisy trajectory governed by the feature-sampling process as well as the degree of completeness (u) and encoding fidelity (c_S) of the traces stored in memory. Sample $x(t)$ paths for both targets (blue) and foils (red) are shown in dashed lines; the solid blue and red lines show the mean target and foil trajectories, respectively, while the shaded regions depict 95% probability regions around each mean. A response is made when $x(t)$ intersects one of two collapsing decision boundaries; if the upper boundary is hit first, the response is “old”, but if the lower boundary is hit, the response is “new”. The distribution of response times at each boundary, accrued over many trials, is shown in the upper and lower panels. Each model time-step takes a constant amount of time p .

deviation σ_R (yielding shape parameter $s = \frac{\bar{T}_R^2}{\sigma_R^2}$ and rate parameter $r = \frac{\bar{T}_R}{\sigma_R}$). The predicted RT distributions are thus a convolution of this Gamma distribution with the first-passage time densities of the model’s accumulation process (after being multiplied by the time-per-sample ρ).

Core aspects of the dynamic approach

Having now described both the conceptual framework and mathematical model behind our dynamic approach to recognition, we reiterate several core aspects of this approach that serve as starting points for the explorations and insights in the rest of this paper:

In most situations, the current context serves as a baseline against which subsequent evidence is judged. A consequence of this is that, initially, the traces activated will be those formed recently, e.g., during the study list. For a trace from prior life history to become active and contribute to familiarity, its content features must match enough of those in the memory probe to overcome the mismatching context features (see Figure 2). This is particularly relevant for studying word frequency effects in recognition: If such effects are due to interference from prior episodes in which the word was encountered, this interference can only be manifest *after* enough features of the word have entered the memory probe to activate the traces of those prior episodes. In other words, word frequency effects should be diminished if a response must be made quickly, relative to slow responses when enough content features of the word have accumulated to activate its prior event traces. We explore this dynamic prediction in Section 1.

Recognition decisions depend on tracking changes in familiarity over time. What happens when the current context is *not* the target context? One task in which this occurs is called “list discrimination”, in which participants must recognize events that occurred prior to the current list context. In this case, the memory probe still begins with features of the current context and participants must reinstate the target context over the course of the recognition trial, as we show in Section 2. Tracking changes resulting from contextual reinstatement can, just like tracking changes due to content features, be used to support list discrimination judgments.

When a brief masked prime is presented prior to a trial, features of the prime can join those of the current context to define the initial level of familiarity. In a sense, the features that “leak” from the prime into the memory probe become a part of the context for that trial. We explore this situation in Section 3, using our model to formulate and decide between different hypotheses regarding the effect of short-term primes.

Recognition dynamics are governed by the time at which features enter the pool available for sampling. Just as history traces require time to be activated and context takes time to reinstate, not all content features are always available at the beginning of a trial. In these situations, the dynamics of recognition are chiefly governed by the time at which different features enter the pool available for sampling, but different kinds of features may be delayed for different reasons. In Section 4, we show that features that are integral to the memory representation of a word, like its base form and semantic features, are part of the pool from the beginning of the trial but that other features, like the final letter that distinguishes between singular and plural nouns or the presentation modality of a word, may not be available until later. In Section 5, we argue that associative features are also delayed relative to features of the items that are associated, and that this delay results from the fact that the emergence of associative information depends on having enough item features already sampled.

Section 1: Word Frequency and Speed-Accuracy Trade-Off

It is well-documented that 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). This 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 also 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 many of these theories make specific dynamic predictions, such that by taking our dynamic approach it is possible to rule out a variety of theories that would otherwise be indistinguishable.

First, it is important to emphasize that LF and HF words differ in many respects: LF words may attract more attention than HF words encouraging deeper processing or criterion setting (Glanzer & Adams, 1990; Malmberg & Nelson, 2003); LF words are comprised of relatively unusual semantic (Steyvers & Malmberg, 2003; Adelman, Brown, & Quesada, 2006) and orthographic (Malmberg, Steyvers, Stevens, & Shiffrin, 2002) features; and HF words have more traces in memory from prior contexts than LF words (Nelson & Shiffrin, 2013). The word frequency mirror effect could arise if HF words were subject to more interference from these many prior historical traces (Dennis & Humphreys, 2001), but if so, it would not appear until later in a trial, when enough content features of the word had accumulated to activate those history traces. In a similar vein, dual-process theories propose that the hit rate advantage for LF words cannot appear until its target trace has had time to be recalled (Reder et al., 2000). Finally, it may be that an early attentional or criterion-setting process yields the mirror effect (e.g., Gillund & Shiffrin, 1984; Glanzer & Adams, 1990; Hintzman, 1994; Malmberg & Nelson, 2003), in which case it would, once again, not appear until a participant had enough time to process the word during the trial. In contrast to each of these theories, which predict that word frequency effects should not appear until later in the trial, if such effects were due to the inherently more distinctive features of LF words relative to HF words (Shiffrin & Steyvers, 1997), the frequency effect should persist throughout a trial as soon as features enter the available pool for sampling.

Only the feature distinctiveness account predicts that the word frequency mirror effect should be present throughout the recognition process; all other theories propose that it should be delayed in some fashion. Although the distinction between these classes of theory is qualitative, we augment this with a quantitative comparison of the feature distinctiveness account to an exemplar of a delay theory, namely, that frequency effects result from interference from prior history traces. We test these theories using data from two paradigms: a response signal experiment reported by Hintzman, Caulton, and Curran (1994) and two free-response experiments (Starns, Ratcliff, & McKoon, 2012; Rae, Heathcote, Donkin, Averell, & Brown, 2014) in which word frequency interacts with instructions to emphasize speed versus accuracy. Along the way, we demonstrate showing how our dynamic model can account for various common experimental manipulations, including increased study time, response bias, and the effects of speed emphasis. To anticipate, across all of these experiments, we find that the word frequency mirror effect persists regardless of speed vs. accuracy emphasis or response signal delay, arguing against any account of word frequency that predicts a difference in the effect as a function of time. Instead, word frequency effects are best explained by the relative distinctiveness of the features used to encode words of different frequencies.

Table 3

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		
	Distinctiveness	History traces	Both
u	0.371	0.372	0.371
c_S	0.982	0.984	0.982
K_L	0*	89	0
K_H	0*	1134	0
g_L	0.484	0.5*	0.484
θ	0.037	0.054	0.037
t_0	121.809	94.295	121.809
ρ	6.686	7.718	6.686
R^2_{Adj}	0.958	0.920	0.950

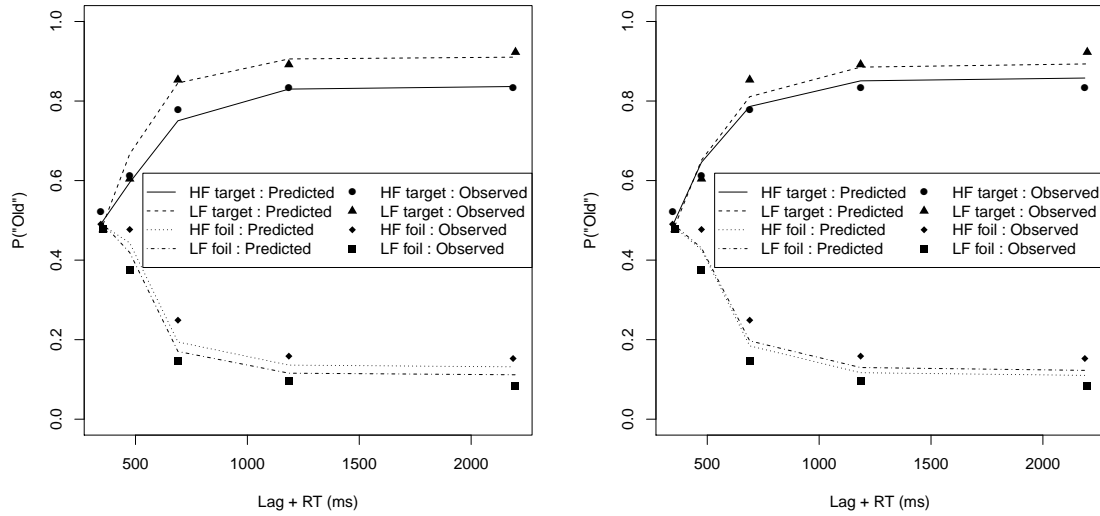
Quantitative Model

Response signal. In their Experiment 6, Hintzman et al. (1994) studied the source of the word frequency effect using a response signal procedure. 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 ,

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

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^2_{Adj} 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 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



(a) Feature distinctiveness only.

(b) History traces only.

Figure 5. Observed probability of responding “old” as a function of signal lag and response time in Hintzman et al. (1994, Exp. 6), with predictions from two different dynamic accounts of word frequency effects.

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.

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 5: 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, one in which we model group performance and a second in which we model individual trial-by-trial recognition.

Group free response. In the experiment reported by Starns et al. (2012), four participants each completed 20 sessions of a recognition memory task¹. 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. In addition, test lists contained different proportions of target words, 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. The resulting data are the response proportions and 0.1, 0.3, 0.5, 0.7, and 0.9 RT quantiles for each response in each condition,

¹We thank these authors for providing us with the data to be analyzed in this section.

averaged across the four individuals (via Vincentizing Ratcliff, 1979).

If word frequency effects were due to factors that varied as a function of time, word frequency effects should be mitigated under speed emphasis relative to when accurate responding is emphasized, however this was not the case, nor did the different levels of bias reduce the word frequency effect (see Starns et al., 2012, tables 1 and 2). Thus, these data provide qualitative evidence that effects of word frequency depend on factors—like feature distinctiveness—that are essentially invariant with respect to both temporal dynamics and instructions. We have two primary aims in this section: First, to provide additional quantitative evidence that feature distinctiveness underlies the difference in recognition between low and high frequency words; and second, to examine the degree to which model parameters can give a deeper interpretation of these results than a more general model of response time and accuracy.

We jointly fit the group accuracy and response time distributions in this experiment via quantile maximum likelihood (Heathcote, Brown, & Mewhort, 2002), enabling quantitative model comparisons using the Akaike Information Criterion (AIC)². We assumed that feature storage probability u increased with the number of study repetitions (each value of u is thus labeled u_1 , u_2 , and u_4). We allowed for both initial boundary separation A_0 and bias b to vary as a function of speed/accuracy emphasis and target proportion, however mean residual time \bar{T}_R and the standard deviation of residual time σ_R were only allowed to vary between speed/accuracy focus. Finally, we assumed single values of encoding fidelity c_S and time-per-sample ρ across all conditions.

Word frequency. As above, we first fit three versions of our model to these data, each representing a different source of word frequency effects. In model 1, prior history traces alone were assumed to underly word frequency effects, represented, once again, by the parameters K_L and K_H for the number of prior history traces in memory for low- and high-frequency words, respectively. In model 2, feature distinctiveness alone was assumed to underly word frequency effects, represented by the single parameter g_L for the probability of a chance feature match for a low-frequency word; the corresponding probability for high-frequency words, was set to $g_H = 1 - g_L$. Finally, model 3 allows for both differing numbers of history traces and increased low-frequency feature distinctiveness. As shown in columns 1–3 of Table 4, the preferred model of word frequency is model 2, in which only feature distinctiveness differs as a function of word frequency. Indeed, when the number of prior history traces is allowed to vary, the best fit is obtained when both $K_L = K_H = 0$.

Speed emphasis. Both Starns et al. (2012) and Rae et al. (2014, see below) argue that an emphasis on speed over accuracy can affect not just response boundaries, but decision evidence as well. One possibility is that a focus on response speed means that participants cannot maintain the features of the memory probe to as high a fidelity as when they are focused on accurate responding. Thus far, we have assumed that features of the test item that are sampled into the probe are sampled with their correct values, but we can model noise in the probe by allowing c_T , the probability of sampling the correct value of a probe feature, to be less than one; this is labeled “model 4” in Table 4. Another possibility is that, under speeded responding, participants are willing to commit to a response not just on the basis of weaker evidence (smaller boundary separation), but also with a less complete probe. We can instantiate this in our model as a reduction in N_C^* , which governs the rate at which response boundaries collapse. If $N_C^* < N_C$, then boundaries will collapse faster than the rate at which features are sampled and participants will tend to respond when the probe has fewer features than under accuracy emphasis; this is labeled “model 5” in Table 4. Finally, both c_T

²In most cases, BIC, which imposes a stronger penalty on number of free parameters than AIC, yields the same conclusions.

Table 4

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 4-6 vary parameters related to speed/accuracy trade-off; the overall preferred model by Akaike Information Criterion (AIC) is indicated in boldface.

Parameter	Emphasis	Target proportion	Model					
			1	2	3	4	5	6
u_1			0.354	0.282	0.282	0.282	0.289	0.289
u_2			0.424	0.393	0.393	0.393	0.395	0.395
u_4			0.505	0.477	0.477	0.477	0.481	0.481
c_S			0.908	0.942	0.942	0.942	0.940	0.940
K_L			0	0*	0	0*	0*	0*
K_H			0	0*	0	0*	0*	0*
g_L			0.5*	0.459	0.459	0.459	0.460	0.460
A_0	Accuracy	0.21	22.7	22.6	22.6	22.6	23.6	23.6
		0.32	23.2	22.7	22.7	22.7	23.8	23.8
		0.50	25.1	24.3	24.3	24.3	25.4	25.4
		0.68	26.6	25.7	25.7	25.7	26.8	26.8
		0.79	26.7	25.7	25.7	25.7	26.9	26.9
	Speed	0.21	13.0	11.4	11.4	11.4	14.8	14.8
		0.32	13.4	11.6	11.6	11.6	15.2	15.2
		0.50	14.2	12.2	12.2	12.2	16.0	16.0
		0.68	13.8	11.8	11.8	11.8	15.3	15.3
		0.79	14.8	12.7	12.7	12.7	16.3	16.3
b	Accuracy	0.21	0.476	0.445	0.445	0.445	0.450	0.450
		0.32	0.517	0.491	0.491	0.491	0.494	0.494
		0.50	0.563	0.543	0.543	0.543	0.544	0.544
		0.68	0.610	0.596	0.596	0.596	0.594	0.594
		0.79	0.649	0.640	0.640	0.640	0.636	0.636
	Speed	0.21	0.499	0.435	0.435	0.435	0.457	0.457
		0.32	0.555	0.507	0.507	0.507	0.514	0.514
		0.50	0.586	0.552	0.552	0.552	0.548	0.548
		0.68	0.622	0.596	0.596	0.596	0.583	0.583
		0.79	0.672	0.661	0.661	0.661	0.635	0.635
\tilde{T}_R	Accuracy		0.376	0.388	0.388	0.388	0.379	0.379
	Speed		0.350	0.374	0.374	0.374	0.362	0.362
σ_R	Accuracy		0.0553	0.0567	0.0567	0.0567	0.0568	0.0568
	Speed		0.0516	0.0577	0.0577	0.0577	0.0609	0.0609
ρ			0.00414	0.00405	0.00405	0.00405	0.00414	0.00414
N_C^*	Speed		30*	30*	30*	30*	25.9	25.9
c_T	Speed		1*	1*	1*	1.00	1*	1.00
AIC			11063	10482	10486	10484	10335	10337

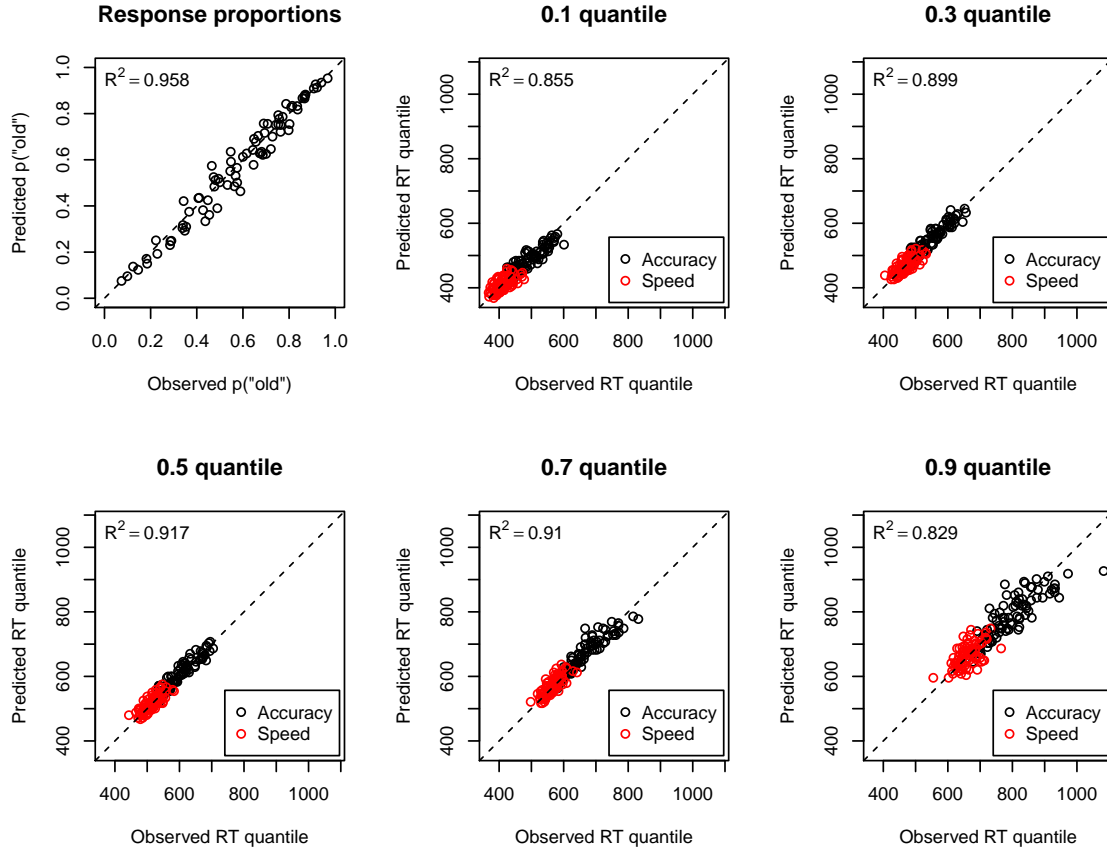


Figure 6. Diagnostic plots for the preferred model fit (#5) to the data from Starns et al. (2012), with proportion of variance of each statistic accounted for indicated in the upper left corner.

and N_C^* might be affected under speed emphasis, which we label “model 6”. As shown in columns 4–6 of Table 4, varying c_T does not increase the fit of the model, but allowing N_C^* to vary does. This provides evidence that speed emphasis does not necessarily impair participants’ ability to maintain the memory probe, but it does encourage participants to respond on the basis of not only a lower *quality* of evidence (smaller initial boundaries), but also a lower *quantity* of evidence (fewer probe features).

Quality of fit. As shown in Figure 6, our preferred model (number 5) does a good job accounting for both observed accuracy and RT quantiles. For comparison, Starns et al. (2012) fit a diffusion model to these data by maximizing χ^2 , a statistic that is very close to quantile maximum likelihood given the large number of trials involved. Their best-fitting diffusion model achieved a χ^2 of 2418, which is superior to the χ^2 of 3597 achieved by our model 5. However, Starns et al. (2012) also report that the RTCON model (Ratcliff & Starns, 2009, 2013), a nonlinear leaky competing accumulator model of RT and confidence, reached a χ^2 of 5533, suggesting that our model still provides a good quantitative account of these data despite using only 31 parameters, relative to the 68 needed by the diffusion model (a comparison of our model’s parameters to those of the diffusion is provided in Appendix C).

Individual free response. In addition to group-level descriptions of accuracy and response time, we investigate the nature of word frequency effects and their interaction with response speed at the level of individual trial-by-trial performance. In so doing, we ensure that the conclusions reached above are not an artifact of the unusual properties of the response signal task nor of averaging across individuals. The relevant data are from an experiment reported by Rae et al. (2014) in which 46 participants engaged in several study-test blocks³. Each study list consisted of 56 words, of which 50 appeared as targets on the subsequent test list, along with 50 foils. Within each study-test block, participants were instructed to emphasize either accuracy or speed in their responses. Although Rae et al. (2014) did not report effects of word frequency, we divided their stimuli into low- and high-frequency words based on a median split in word frequency (using the MRC database; Coltheart, 1981). Each participant contributed between 42 and 48 observations in each combination of speed/accuracy emphasis, word frequency, and study status (target/foil) for a total of 17,617 trials. For further details of the experimental procedure, the reader is directed to their paper.

For each model variant, we found parameters for each participant that maximized the joint likelihood of their responses across all conditions. As there was only one level of study time/repetition, we fit a single value of encoding probability u for each participant, as well as encoding fidelity c_S and time-per-sample p . We fit separate values of initial boundary separation A_0 , bias b , mean residual time \bar{T}_R , and residual time standard deviation σ_R for speed emphasis and accuracy emphasis.

Word frequency. As above, we fit three different model variants representing different theories of the source of word frequency effects: Prior history traces only (model 1), feature distinctiveness only (model 2), and both prior history and feature distinctiveness (model 3). As shown in columns 1–3 of Table 5, the overall preferred model in terms of AIC, is model 2, in which word frequency effects are due entirely to differences in feature distinctiveness between low- and high-frequency words. As above, when the number of history traces is allowed to vary, it tends to be estimated to be quite low, with median estimates across participants hovering around 1 for LF words and 8–10 for HF words, hardly enough to make a significant impact on recognition performance. The only individual who is better fit by model 1 ($AIC = -300.41$, with $K_L = 1$ and $K_H = 155$) is fit nearly as well by model 2 ($AIC = -300.36$, with $g_L = 0.491$), offering little support for the history trace account. Across all individuals, we again find that word frequency effects are best accounted for by the relative distinctiveness of the features of which words are comprised.

Speed emphasis. As above, we fit three additional model variants to each individual to assess the effect of speed emphasis on their decision process. One model allows for features in the probe to be represented with error ($c_T < 1$, model 4), one allows response boundaries to collapse faster in speed conditions ($N_C^* < N_C$, model 5), and the final model allows for both effects (model 6). Summed AIC across all individual fits favors model 5, in which speed instructions increase the rate of boundary collapse, enabling participants to respond on the basis of fewer sampled features than they would when accuracy is emphasized. At the individual level, the majority of participants (25/46) are best fit by model 5, although there is considerably more ambivalence here than with regard to word frequency effects, suggesting that speed emphasis may, for some individuals, also affect the quality with which they encode probe features.

Quality of fit. Figure 7 shows the fit of model 5—the model preferred by summed AIC—to each individual in each condition, illustrating that the model does a good job accounting for both

³We are grateful to the authors for providing us with the raw data from this experiment.

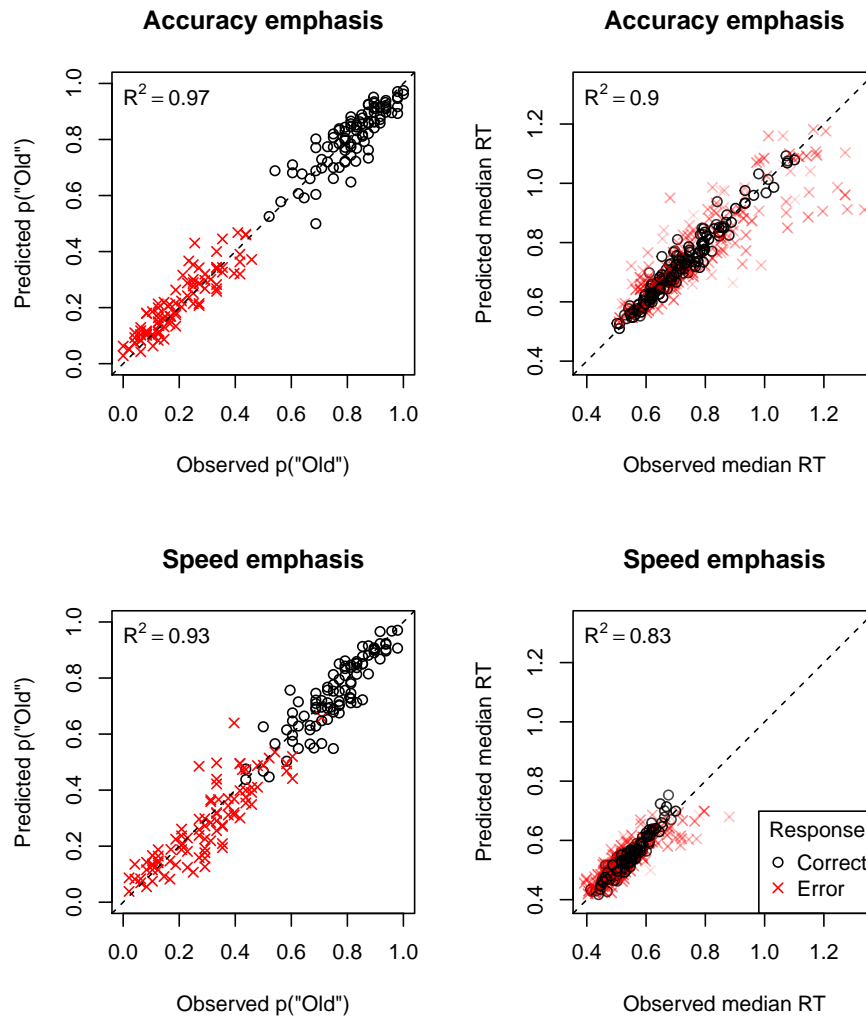


Figure 7. Observed response probabilities and median response times for each participant in each condition of the experiment reported by Rae et al. (2014), compared with predictions from the fit of model 5 to each participant. The upper left corner of each panel gives the proportion of variance R^2 in each measure accounted for by the model, weighted by the number of observations in each condition. The number of observations at each point is also used to determine the weight of each plotted point, hence the lighter colors for error response times, which are less common than correct responses.

Table 5

Medians of the best-fitting parameter values for each participant from the recognition experiment reported by Rae et al. (2014), as well as Akaike Information Criterion (AIC) summed over individual model fits to each participant. Also listed are the number of participants best fit by each model variant according to AIC for two relevant comparisons: the effect of word frequency and the effect of speed emphasis. Values fixed for each model are marked with an asterisk and preferred models are shown in bold.

Parameter	Emphasis	Model					
		1	2	3	4	5	6
u		0.506	0.542	0.544	0.511	0.514	0.500
c_S		0.935	0.935	0.931	0.936	0.936	0.942
K_L		1	0*	1	0*	0*	0*
K_H		10	0*	8	0*	0*	0*
g_L		0.5*	0.467	0.466	0.466	0.465	0.465
A_0	Accuracy	44.5	43.2	42.2	42.5	40.2	41.0
	Speed	16.2	14.9	14.1	14.5	65.2	29.3
b	Accuracy	0.525	0.524	0.524	0.523	0.525	0.526
	Speed	0.566	0.571	0.574	0.570	0.513	0.521
\bar{T}_R	Accuracy	0.352	0.372	0.369	0.371	0.371	0.378
	Speed	0.398	0.404	0.405	0.401	0.371	0.385
σ_R	Accuracy	0.0586	0.0534	0.0543	0.0541	0.0480	0.0543
	Speed	0.0547	0.0603	0.0597	0.0597	0.0640	0.0623
ρ		0.00455	0.00471	0.00470	0.00470	0.00538	0.00531
N_C^*	Speed	30*	30*	30*	30*	17.3	26.3
c_T	Speed	1*	1*	1*	0.999	1*	0.999
AIC		-6541	-6893	-6736	-6842	-7126	-7055
Number of participants best word frequency fit		1	43	2			
Number of participants best speed-accuracy trade-off fit			11		4	25	6

response proportions and response times. Although the model appears to occasionally underestimate median error RTs, these are estimated from fewer observations (an average of 10.9 responses, ranging from 0 to 34) than the median correct RTs (an average of 37.0 responses, ranging from 14 to 48). Overall, the strong correspondence between model predictions and observations ensures that meaningful theoretical inferences can be drawn on the basis of estimated model parameters and comparisons between different model variants.

Parameter distributions. By fitting to individual rather than group data, we can explore how model parameters covary with each other and across individuals. We estimated the joint distributions of parameters using 1,000,000 bootstrap samples from the set of best-fitting parameters for the preferred model (5). The means of median group parameters are shown in Table 6, showing that g_L is reliably less than 0.5, reflecting the higher distinctiveness of the features of LF than HF features. The residual time distributions, reflecting processes like the time needed for knowledge access and response execution, do not substantially differ between speed and accuracy emphasis. Instead, initial boundary separation (A_0) is sometimes higher under speed emphasis, but the faster rate of collapse results in faster responses; in essence, under speed instructions, participants place less emphasis on responding based on the quality of evidence (boundary separation) than on the quantity of evidence (number of desired features).

Mean Kendall's τ ordinal correlations among parameters are shown in Table 7. The negative

Bootstrapped mean of median parameter values (with 99% credible intervals in parentheses) for model 5, as applied to individuals from the recognition experiment reported by Rae et al. (2014).

Parameter	Emphasis	Median (99% CI)
u		0.520 (0.432, 0.616)
c_S		0.934 (0.917, 0.949)
g_L		0.466 (0.457, 0.476)
A_0	Accuracy	40.318 (28.413, 55.274)
	Speed	62.653 (31.272, 100.245)
b	Accuracy	0.527 (0.516, 0.545)
	Speed	0.515 (0.502, 0.535)
\bar{T}_R	Accuracy	0.372 (0.331, 0.414)
	Speed	0.370 (0.320, 0.404)
σ_R	Accuracy	0.049 (0.025, 0.070)
	Speed	0.063 (0.042, 0.082)
ρ		0.005 (0.004, 0.006)
N_C^*		16.930 (12.741, 22.404)

Bootstrapped mean of Kendall's τ correlations (with 99% credible intervals in parentheses) for model 5, as applied to individuals from the recognition experiment reported by Rae et al. (2014). Correlations with 99% credible intervals that exclude zero are shown in bold.

	c_S	g_L	A_0	Accuracy	Speed	b	Accuracy	Speed	\tilde{T}_R	Accuracy	Speed	σ_R	Accuracy	Speed	ρ	N_c^+	Speed
u	-0.64 (-0.8, -0.42)	0.29 <i>(-0.02, 0.55)</i>	0.19 <i>(-0.13, 0.5)</i>	-0.1 <i>(-0.41, 0.22)</i>	0.33 (0.01, 0.61)	0.48 (0.21, 0.71)	-0.14 <i>(-0.39, 0.13)</i>	0 <i>(-0.27, 0.28)</i>	-0.08 <i>(-0.34, 0.2)</i>	-0.24 <i>(-0.5, 0.03)</i>	0.1 <i>(-0.21, 0.38)</i>	0.16 <i>(-0.14, 0.45)</i>					
c_S	—	-0.25 <i>(-0.5, 0.04)</i>	-0.04 <i>(-0.35, 0.26)</i>	-0.01 <i>(-0.32, 0.28)</i>	-0.43 (-0.67, -0.14)	-0.45 (-0.67, -0.16)	0.07 <i>(-0.21, 0.34)</i>	0.04 <i>(-0.25, 0.33)</i>	0.04 <i>(-0.25, 0.31)</i>	0.08 <i>(-0.18, 0.35)</i>	-0.1 <i>(-0.39, 0.2)</i>	0.03 <i>(-0.3, 0.28)</i>					
g_L	—	—	0.1 <i>(-0.22, 0.4)</i>	0.13 <i>(-0.15, 0.39)</i>	0.27 (-0.04, 0.55)	0.23 <i>(-0.1, 0.5)</i>	-0.1 <i>(-0.39, 0.19)</i>	0 <i>(-0.28, 0.27)</i>	-0.02 <i>(-0.3, 0.25)</i>	-0.13 <i>(-0.4, 0.16)</i>	-0.01 <i>(-0.26, 0.23)</i>	0.03 <i>(-0.28, 0.33)</i>					
A_0	Accuracy	—	—	0.03 <i>(-0.25, 0.32)</i>	-0.35 (-0.65, -0.04)	0.12 <i>(-0.17, 0.4)</i>	-0.31 (-0.53, -0.04)	-0.02 <i>(-0.3, 0.26)</i>	-0.06 <i>(-0.38, 0.24)</i>	-0.21 <i>(-0.47, 0.07)</i>	-0.15 <i>(-0.38, 0.1)</i>	0.11 <i>(-0.2, 0.4)</i>					
	Speed	—	—	—	0.09 <i>(-0.21, 0.37)</i>	-0.33 <i>(-0.67, 0.03)</i>	-0.11 <i>(-0.39, 0.16)</i>	-0.22 <i>(-0.5, 0.06)</i>	-0.15 <i>(-0.41, 0.12)</i>	0.02 <i>(-0.29, 0.29)</i>	0.12 <i>(-0.2, 0.42)</i>	-0.6 (-0.78, -0.37)					
b	Accuracy	—	—	—	—	0.29 <i>(-0.03, 0.56)</i>	0.1 <i>(-0.17, 0.36)</i>	0.04 <i>(-0.25, 0.34)</i>	-0.12 <i>(-0.41, 0.21)</i>	-0.02 <i>(-0.32, 0.29)</i>	0.15 <i>(-0.11, 0.39)</i>	-0.03 <i>(-0.35, 0.31)</i>					
	Speed	—	—	—	—	—	-0.11 <i>(-0.35, 0.15)</i>	-0.05 <i>(-0.3, 0.21)</i>	-0.06 <i>(-0.33, 0.2)</i>	-0.32 (-0.5, -0.12)	-0.06 <i>(-0.38, 0.26)</i>	0.37 (0.06, 0.65)					
\tilde{T}_R	Accuracy	—	—	—	—	—	—	0.39 (0.12, 0.61)	0.31 (0.03, 0.58)	0.31 (0.03, 0.57)	-0.32 (-0.56, -0.06)	-0.01 <i>(-0.29, 0.29)</i>					
	Speed	—	—	—	—	—	—	—	0.3 (0.07, 0.52)	0.43 (0.19, 0.62)	-0.33 (-0.58, -0.07)	0.07 <i>(-0.25, 0.4)</i>					
σ_R	Accuracy	—	—	—	—	—	—	—	—	0.3 (0, 0.56)	-0.29 (-0.5, -0.07)	0.09 <i>(-0.2, 0.38)</i>					
	Speed	—	—	—	—	—	—	—	—	—	-0.07 <i>(-0.34, 0.18)</i>	-0.3 <i>(-0.57, 0.01)</i>					
ρ	—	—	—	—	—	—	—	—	—	—	—	-0.23 <i>(-0.48, 0.03)</i>					

correlation between A_0 and N_C^* in the speed condition is consistent with participants either reducing the quality of evidence needed to respond (lower boundary separation) or the quantity of evidence (fewer probe features), but not both. The negative correlation between u and c_S *must* exist because if both of these parameters were low, accuracy would drop to chance regardless of whether speed or accuracy were emphasized. Their correlations with bias are not entailed by any aspect of the model (see Appendix B), but reflect a property of this group of participants. Positive correlations between the mean and standard deviation of residual times likely reflect the fact that residual time cannot be negative, although it is sensible that these parameters are also correlated between conditions, since they reflect many of the same processes of knowledge activation and response execution. Importantly, these correlations do not suggest any strong trade-offs between the explanatory power of model parameters.

Discussion

By taking a dynamic approach, we were able to apply a qualitative test to one of the most important phenomena in recognition memory: why are low-frequency words better recognized than high-frequency words? Word frequency mirror effects exist from the moment information regarding the test word begins to be retrieved, ruling out theories of the word frequency effect that predict that it should change within a trial. Such accounts include those that depend on criterion setting (Glanzer & Adams, 1990; Hintzman, 1994), a redeployment of attentional resources (Malmberg & Nelson, 2003), or a slow-acting recall process (Reder et al., 2000). Using our mathematical model, we additionally made a quantitative test between two theories: that word frequency effects depend on interference from prior episodes in which the word occurs (Dennis & Humphreys, 2001); and/or that such effects depend on the relative distinctiveness of the features used to represent low frequency words (Shiffrin & Steyvers, 1997). Across all model fits, the feature distinctiveness account was preferred at both the group and individual levels, reinforcing the fact that word frequency effects persist throughout a trial, since historical traces would only begin to interfere when enough features of the word had joined the memory probe for those traces to become active. The core principle behind our conclusions in this section is that the current context determines which traces in memory are initially active; any model that embraces this role for context should yield similar conclusions. Although one could imagine a model in which historical traces were activated immediately, such a model would need additional mechanisms to account for effects of recency and context-specificity and would be inconsistent with functional neural data (McKenzie et al., 2014), all of which all point to a special role for the current context in event retrieval.

Although we have shown that interference from *specific* prior episodes is not the primary source of word frequency effects in recognition, we do not claim that such interference cannot exist (cf. Maddox & Estes, 1997; Nosofsky, Cao, Cox, & Shiffrin, 2014), and we certainly do not argue that prior experience is irrelevant to episodic memory. Rather, prior experience can have two distinct effects: On the one hand, episodes stored in other contexts can interfere with retrieval in the present context (Dennis & Humphreys, 2001), while on the other hand, experience in prior contexts can lead to expertise that allows items to be encoded more distinctively (Czerwinski, Lightfoot, & Shiffrin, 1992; Shiffrin & Lightfoot, 1997; Gauthier & Tarr, 1997). In situations where experience with test items is limited to experimental contexts that are highly similar to the recognition context, such as in the training studies reported by Nelson and Shiffrin (2013), interference is likely to play a more important role than expertise. In contrast, the diverse linguistic and situational contexts in which words are encountered—which tends to be highly dissimilar to the recognition context—appears to

result in expertise outweighing any effect of interference.

We posit that the expertise learned from experience with words relates primarily to words' semantic features, which in turn enter the pool of active features used to encode a word event. A wealth of research on semantic memory has supported theories in which semantic features are dependent on the contexts in which a word occurs, whether that context is other nearby words (e.g., Lund & Burgess, 1996; 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)⁴. As outlined in our model, when a word is encountered in a study episode, the content features of the resulting trace are partially derived 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 (Adelman et al., 2006), the semantic features by which they are encoded are correspondingly more distinctive, just as we have inferred from our model fits (Jones, Johns, & Recchia, 2012; Nelson & Shiffrin, 2013). In light of these results, we believe a productive route for future research to travel is the interplay between episodic encoding and semantic learning across the lifespan (Johns & Jones, 2010; Kumaran & McClelland, 2012; Nelson & Shiffrin, 2013).

Section 2: Contextual Reinstatement

The reason historical traces take time to be activated—such that they can only interfere late in the recognition process—is that prior to the sampling of content features, the memory probe contains features of the current context. Because historical traces were formed in different contexts, they can only be activated when enough content features have joined the memory probe. What a participant is asked to recognize events from *outside* the current context? Such a situation occurs in list discrimination paradigms, in which participants are told to endorse an item only if it appeared on a particular study list, which may not be the most recent one.

As an exemplary set of results, Hintzman et al. (1998, Exp. 1) conducted 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 8. 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.

One account of the nonmonotonic false alarm curve is that participants use the test item to cue recall of the study context. If this recall process took longer than the process by which content features of the item are sampled and added to an evolving memory probe, List 2 foils would initially and erroneously be called “old” before recall can be used to correctly reject such items. An alternative account is a form of contextual reinstatement (Polyn, Norman, & Kahana, 2009; Howard, Shankar, Aue, & Criss, 2015). Memory is initially and automatically probed with the

⁴Perceptual and linguistic context appear to carry a great deal of redundant information about word meaning (Riordan & Jones, 2010), although the two forms of experience can interact to support additional learning and inference (Johns & Jones, 2012).

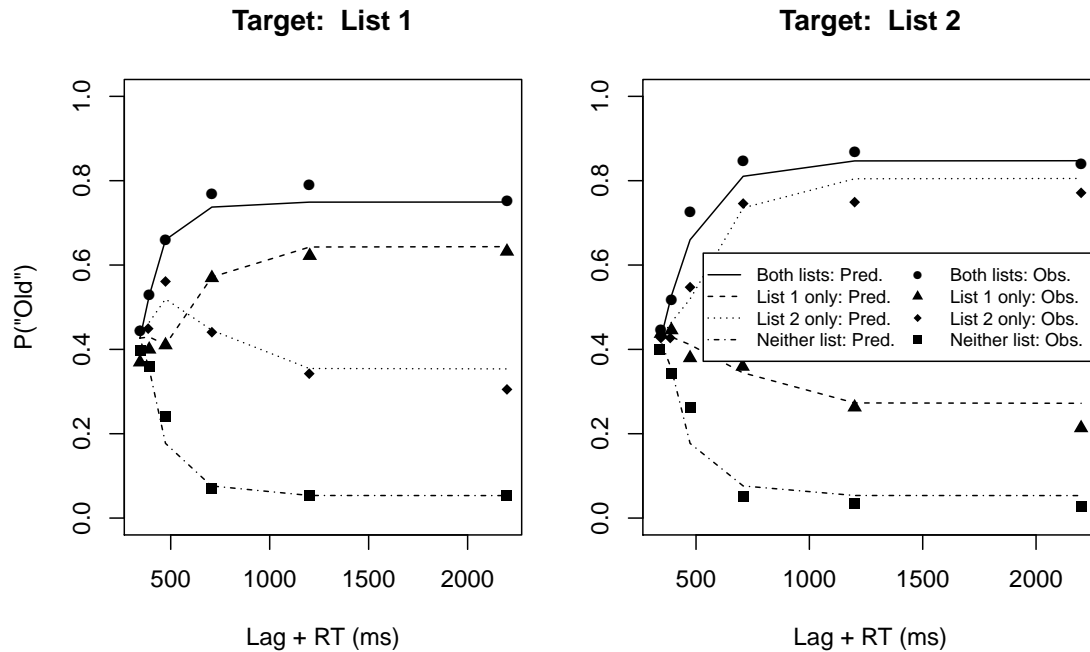


Figure 8. Data and model predictions for Hintzman, Caulton, and Levitin (1998, Exp. 1). Parameters used to generate model predictions are given in Table 8.

current context, leading to a high false alarm rate to recent List 2 items. The features of the context that are available automatically tend not to discriminate very well between the two lists, however, since they occur nearby in time and in the same location. Thus, participants reinstate contextual features that are specific to the target context, enriching context representation in the probe to make it better “tuned” to the specific target context and causing false alarms to diminish with additional processing time. 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 difference between repeated and

Table 8

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

Parameter	Experiment 1	Experiment 2
u_1	0.488	0.422
u_2	—	0.549
c_S	0.963	0.940
s_X	0.969	0.975
p_X	0.198	0.255
t_X	397.322	487.308
$\theta_{\text{Recognition}}$	0.295	-1.924
$\theta_{\text{List discrim.}}$	—	-0.279
t_0	94.459	0.957
ρ	4.152	8.528

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 now demonstrate how our dynamic model can implement a form of contextual reinstatement to illustrate what happens when the target recognition context differs from the current context. We also show how one of the key features of our model—that recognition decisions depend on tracking changes in familiarity—can be used to support list discrimination decisions by tracking changes in familiarity as context is reinstated.

Quantitative Model

We model both experiments from Hintzman et al. (1998) using a form of contextual reinstatement. Reinstating context features is accomplished using the same mechanisms as sampling of content features. As a result, a proportion p_X of the probe’s capacity for content features is taken up by these reinstated context features, consistent with the idea that contextual reinstatement requires cognitive resources and interferes with the ability to form a complete memory probe. At time t_0 , the probe of memory consists of the N_X features of the current context, as usual. We presume that these “ambient” context features tend to overlap considerably between successive lists, given that they are experienced close in time and in the same location; these original N_X context features perfectly match the context of the immediately preceding list (as in other applications of our model) and match those in the previous list with probability s_X . Until time t_X , only $(1 - p_X) \times N_C$ content features of the item are available for sampling, but at time t_X the $p_X \times N_C$ reinstated context features join the pool. These reinstated context features match *only* the target list context, “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.

The fit of this model to the data from Hintzman et al. (1998, Exp. 1) is shown in Figure 8, with model parameters given in Table 8. 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 enables list discrimination judgments by tracking changes in familiarity, in the same way that participants perform a typical recognition task. Instead of tracking changes from time t_0 , changes are tracked from time t_X , the time at which contextual reinstatement begins. Just as in normal recognition, the change in familiarity over time as features of the target context enter the probe is diagnostic of the degree to which that word had been experienced in that context, such that the evidence for list discrimination $x_{LD}(t)$ is given by

$$x_{LD}(t) = \sum_{\tau=t_X+1}^t [\phi(\tau) - \phi(\tau-1)] = \phi(t) - \phi(t_X). \quad (13)$$

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. Therefore, to maximize efficiency and minimize interference from non-target traces, the starting time for contextual reinstatement t_X can be chosen by the participant such that enough content features have been sampled by that time to identify the relevant memory trace. Note, again, that this mechanism does not predict a difference in retrieval dynamics as a function of item strength: Retrieval dynamics in this model are entirely a function of (content and context) feature sampling during the test trial, rather than recovery of context features from the memory trace. We predict that item strength only affects asymptotic performance, given that additional storage time allows for more features to be stored.

We fit this model simultaneously to both the recognition and list discrimination tasks from Hintzman et al. (1998, Exp. 2), as shown in Figure 9, with parameters given in Table 8. 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. 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. As in Experiment 1, these features perfectly match the storage context of List 2 (the most recent list), they only match the context features of List 1 with probability s_X . The time t_X at which context features are reinstated corresponds to roughly 46 model “time-steps”, at which point there is an 79% chance that all content features have been sampled (see Equation 1), consistent with the idea that list discrimination must wait until the target trace is effectively isolated.

Discussion

In this section, we have shown how we can model contextual reinstatement using the same feature-sampling principles that underly item recognition to explain the dynamics of list discrimination judgments. However, we have left open a reasonable question: why, if participants know the target context at the beginning of the trial, do they not begin with the features of that context in their initial probe of memory? Why must reinstatement occur *during* the trial, rather than before? Doshier (1984b, Exp. 3) found that, when participants could use a pre-experimental semantic relationship

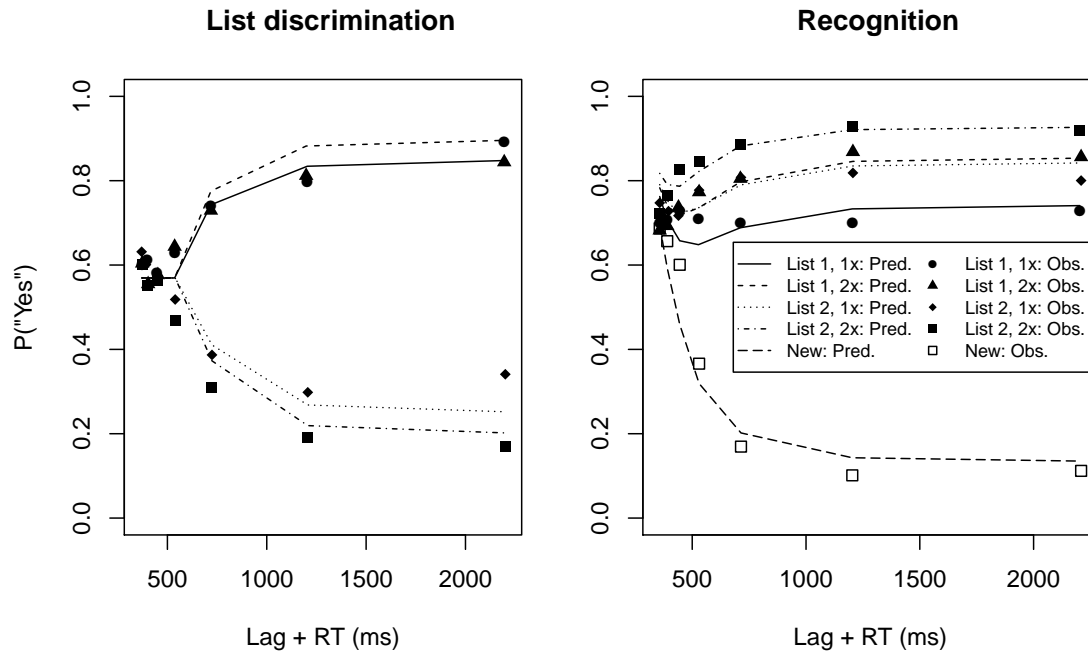


Figure 9. Data and model predictions for Hintzman et al. (1998, Exp. 2). In recognition, positive “yes” responses should be given to any studied item, regardless of list. In list discrimination, “yes” responses should be given only to words studied in List 1. Parameters used to generate model predictions are given in Table 8.

to immediately reject a word pair in an associative recognition task, they failed to do so until later in retrieval. Features of the ambient context are also used automatically at storage (Malmberg & Shiffrin, 2005). It would thus appear that, at least under the conditions we are considering, the initial context features present in the memory probe are not under any kind of strategic control. Although the term “context” is an overloaded one in the study of memory (Klein et al., 2007), these results suggest a useful operative definition: Context *is* whatever features are probing memory (or would be used to probe memory) in the absence of a specific test item. Any features that join the probe as a result of processing a test item or engaging in cognitive effort constitute “content” features—even if those new features were originally part of a prior context. We admit that, given each researcher’s preferred sense of the term “context”, this operational definition may not gain traction, but it will also help to understand subsequent results.

Finally, we note that our approach assumes that contextual reinstatement relies on the same mechanisms by which content features are sampled into the memory probe, thus representing a capacity trade-off: In most recognition situations, the ambient context features that are initially present in the probe are sufficient to pick out the target context, but resources must be expended in order to reinstate features that are specific to a particular context. The amount of resources needed to reinstate a context presumably grows as the target context gets more distant from the current context. Consistent with this account, 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), only more spread out, as would occur if feature sampling were extended to allow for the reinstatement of additional target context features.

Section 3: Masked Priming

Using data from list discrimination experiments, we showed that memory probes at the start of each trial contain features of the current context. When the target context differs from the current context, the target context must be reinstated on each trial using the same feature sampling processes as for content features. As a result, participants can use the change in familiarity driven by contextual reinstatement to make list discrimination judgments. We now consider a situation in which the initial memory probe contains *content* features from a prime as well as context features. In an operation sense, these item features become context features for the current trial; participants continue to track changes in familiarity, but it is from a baseline dictated not just by context but by the content features initially present in the probe as well. As a consequence, participants display a form of short-term priming that has been traditionally attributed to a “fluency” heuristic.

Experiment

According to the fluency account (Jacoby & Dallas, 1981), studying an item enhances its ability to be perceived later at test, and this enhanced “fluency” with regard to processing the item is taken as evidence that the item had been studied. This account predicts that any manipulation that enhances perception of the test item will lead to it being more likely to be called “old”. Although the construct of “fluency” has been implicated in a variety of cognitive domains (Oppenheimer, 2008), it has also been used to explain why masked presentation of the test item immediately prior to recognition leads participants to call the item “old”, whether or not it had actually been studied (Jacoby & Whitehouse, 1989). We present a replication of this result and show how it can be directly explained by our dynamic recognition model without recourse to a separate fluency heuristic⁵.

Participants. 79 undergraduate students from Indiana University participated in the experiment in exchange for course credit. Data from two additional participants was excluded due to below-chance performance.

Stimuli. Stimuli were drawn from a pool of nouns of moderate length (4–7 letters, mean 5.2 letters; 1–3 syllables, mean 1.4 syllables) and frequency (mean log-frequency 8.4, range 3.85–12.66; Burgess & Livesay, 1998).

Procedure. Words were presented in lowercase letters in the center of a computer monitor. 90 words were presented to participants at study, with participants instructed to read the words aloud as they appeared on the screen and to remember them for a later test of memory. The study phase was divided into two blocks, one with words presented for one second, another with words presented for three seconds; the order of these blocks was randomized.

Immediately after study, participants were given a recognition test list comprised of the 90 studied target words and 90 unstudied foils with the instruction to give an “old” response if the test word came from the study list, and otherwise to respond “new”. Old and new responses were randomly mapped to the “A” or “L” key for each participant. Each recognition test word was preceded by a non-diagnostic subliminal prime: on 1/3 of trials, the prime was identical to the test word, on another 1/3 of trials, the prime was a different word that had not been previously seen during the

⁵We gratefully acknowledge the work of Nicholas J. Lewis in designing the experiment and collecting data in the replication study reported here.

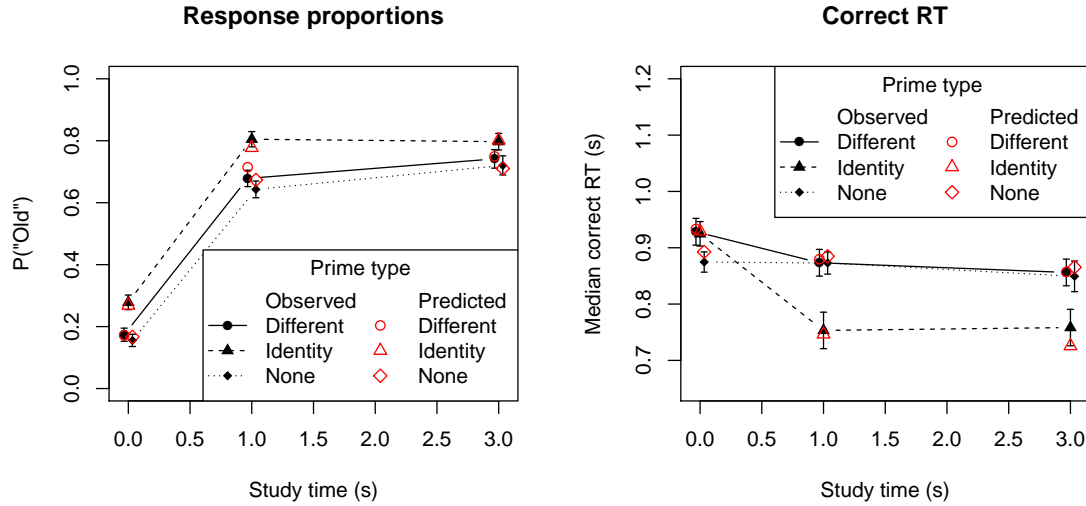


Figure 10. Observed and predicted probability of responding “old” as a function of prime type and study time (left panel) and median correct RT (right panel) in our replication of the “fluency” effect due to masked priming. Error bars on observations depict 95% confidence intervals around the mean. Predictions are from the overall preferred model (“L2”), in which priming effects are a result of features leaking from the prime into the memory probe; see text for model details.

experiment, and on another 1/3 of trials, the prime was a neutral string of characters (“XOXOXO”). On each trial, a pre-mask (“&&&&&&”) was presented for 500 ms followed by a prime (same, different, or neutral) for 50 ms and a post-mask for an additional 500 ms. The screen was blank for 300 ms before the test word was presented. After the participant made a response, the screen was cleared for 1865 ms until the next test trial. Participants were not informed that the primes would be present.

Results. Prior to analysis, trials with RT that were too fast (less than 200 ms) or too slow (longer than 3 s) were excluded. All subsequent analyses and modeling is based on the remaining 14,017 trials from 79 participants.

The observed mean probability of responding “old” in each condition is shown in the left panel of Figure 10. For purposes of analysis, we consider the probability of responding “old” to targets and foils separately. There is a significant effect of prime type on probability of false alarm ($F(2, 78) = 56.8, p \approx 0$); the probability of committing a false alarm was significantly higher for identity primes than for non-word primes ($t(78) = 9.65, p \approx 0$), but only marginally higher for different primes than for non-word primes ($t(78) = 1.68, p = 0.10$). For targets, there was also a significant main effect of prime type ($F(2, 78) = 47.7, p \approx 0$), as well as a significant main effect of study time ($F(1, 78) = 18.3, p \approx 0$) and interaction between prime type and study time ($F(2, 78) = 6.85, p = 0.001$). The probability of a correct “old” response increases with study time ($t(78) = 4.28, p \approx 0$) and is greater following an identity prime than a non-word prime ($t(78) = 9.90, p \approx 0$). A non-identical prime also leads to an increased probability of giving a correct “old” response ($t(78) = 2.23, p = 0.03$).

Observed median correct RT is shown in the right panel of Figure 10. There is a significant main effect of prime type on median correct rejection RT ($F(2, 78) = 8.47, p = 0.0003$), with correct

rejections taking longer following both identity primes ($t(78) = 3.83$, $p = 0.0003$) and different primes ($t(78) = 3.87$, $p = 0.0002$) relative to non-word primes. There was also a significant main effect of prime type on hit RT ($F(2, 78) = 56.7$, $p \approx 0$), but no significant main effect of study time ($F(1, 78) = 1.86$, $p = 0.18$) nor any interaction between prime type and study time ($F(2, 78) = 0.91$, $p = 0.40$). Median correct hit RT is significantly faster following an identity prime ($t(78) = 7.70$, $p \approx 0$) relative to a non-word prime, but there is no significant difference in hit RT between different primes and non-word primes ($t(78) = 0.52$, $p = 0.61$).

In sum, we replicate the key result of Jacoby and Whitehouse (1989) in that subliminal identical primes lead to an increase in the probability of calling the subsequent test item “old”, regardless of whether the test item had actually been studied. In addition, we found that there is a small but detectable increase in the probability of responding “old” when the test item is preceded by a different-word prime, an effect also found but unexplained in the original results of Jacoby and Whitehouse (1989). Finally, we found that correct recognition of a studied word is speeded by a subliminal identical prime while the time to correctly reject an unstudied item increases following both identical and non-identical word primes.

Model comparison

We now use our dynamic recognition model to explore the mechanisms that underly this suite of findings.

Priming mechanisms. A variety of mechanisms might give rise to these priming effects, none of which are mutually exclusive.

Enhanced knowledge access. It is likely that the prime activates the trace in lexical memory corresponding to the prime word, meaning that those features are in the available pool earlier than they would be otherwise. For an identical prime, this means that features could be sampled earlier, but for a different prime, there would be no benefit since the prime would have activated the “wrong” lexical trace. If this were the only mechanism operating, it would predict a response time benefit for identical primes, but no differences in accuracy as a function of prime type. In model terms, enhanced lexical access is reflected in the residual time distribution governed by its mean (\bar{T}_R) and standard deviation (σ_R).

Fluency. If the prime leads to enhanced knowledge access, this greater fluency might lead participants to think that the test item was old. In model terms, this would mean there was a negative correlation between residual time (\bar{T}_R) and bias (b). This account does not specify how non-identical primes might lead to a response bias since these would not be expected to enhance knowledge access for the test item.

Prime storage. The prime might itself be stored as a memory trace—although perhaps only weakly. The result would be to introduce an extra matching trace following an identical prime, thereby increasing the familiarity of the test item. This mechanism would lead to a greater probability of endorsing an item following an identical prime along with faster hit and false alarm RT’s for such items, but would only account for half of the effect of different primes: Introducing a new memory trace would effectively increase the length of the list, leading to the observed higher false alarm rates, but a lower hit rate, which is the opposite of the observed direction.

Feature leakage. Finally, features of the prime may “leak” into the representation of the test item (Huber, Shiffrin, Lyle, & Ruys, 2001). The effect of this is that the memory probe begins not only with context features, but with some content features of the prime as well. The initial familiarity $\phi(t_0)$ thus reflects the match to memory of context features as well as some prime features

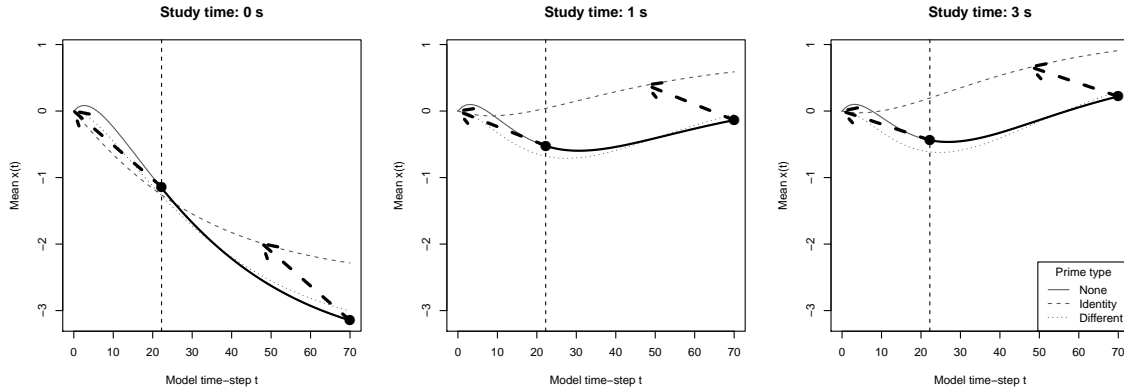


Figure 11. Mean change in familiarity ($x(t)$) from the start of a trial as a function of study time and prime type in our masked priming experiment. Identical primes result in features of the test item being present at the start of a trial (time t_0), effectively “skipping” the initial feature samples. The number of “skipped” time-steps as a result of an identical prime is depicted by the dashed vertical line, while the dashed arrows indicate how the change in starting point affects the familiarity trajectory on a trial following an identical prime.

and as participants accumulate subsequent changes in familiarity, they do so from a baseline that changes with the type of prime. If the prime is identical to the subsequent test item, it effectively removes the first part of the familiarity trajectory, as illustrated in Figure 11. The effect of a non-identical prime is to add some initial noise to the probe by introducing features that are eventually replaced as features are sampled from the subsequent test item.

An identical prime leads to a bias to respond “old” to foils by virtue of eliminating the first few negative steps provided by the first few features. By starting from a lower level of familiarity, the additional features sampled from the consciously-perceived test item cannot drive down familiarity as far, making it harder for a foil’s familiarity trajectory to reach the “new” boundary. This also predicts the higher RT for correct rejections following an identical prime. An identical prime also leads to a bias to respond “old” to targets for slightly more complex reasons. Although familiarity for a target typically increases, as shown in Figure 11, targets often lead to a dip in familiarity early on caused by the fact that, although the features of the probe may match those in a single target trace, they typically do not match those in the other traces formed at study. These non-matching traces drive down familiarity until enough mismatching features have accumulated to cause them to deactivate, at which point familiarity typically increases as the matching target trace becomes dominant.

Features of non-identical primes could also leak into the memory probe. However, because of the time needed to access the lexical trace of the test word, which is different from that of the prime, some of the leaked features could deactivate before sampling begins. The remaining leaked features will not tend to match any of the traces formed at study, so the effect of a non-identical prime on a foil is essentially the same as for an identical prime. For a target, the non-identical prime features still effectively remove some of the initial dip in familiarity that would occur without a prime, meaning that as features of the test item get sampled and gradually replace those of the prime, familiarity has nowhere to go but up, yielding a positive response bias.

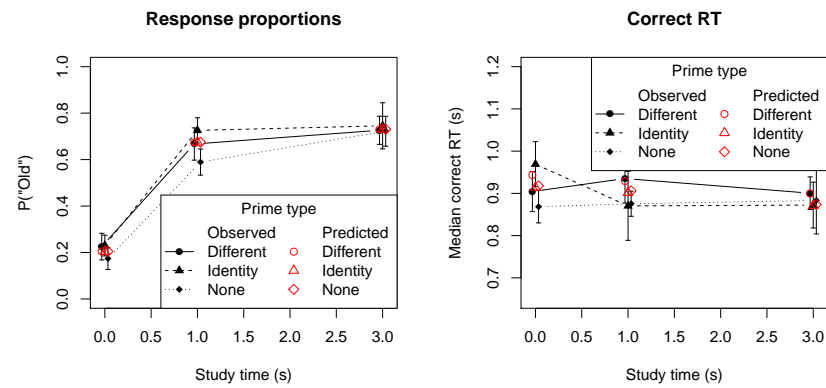
Quantitative model. In qualitative terms, only the feature leakage account is capable on its own of accounting for the full set of masked priming effects. However, it is possible that a combination of mechanisms is actually at work and/or that different participants are affected by the primes in different ways. To determine this, we implemented the three models above—along with all possible combinations—within our dynamic recognition framework in order to fit them quantitatively to the joint accuracy and RT data from each participant.

In all models, we assumed, as usual, that each studied word results in a single trace in memory (for simplicity—and in light of our word frequency results, above—we did not model storage of prior history traces). The probability of storing a content feature in a trace can take one of two values, u_1 or u_3 , depending on study time, while encoding fidelity c_S , boundary separation A_0 , and time-per-sample ρ were assumed equal across conditions. We allowed the mean and standard deviation of residual time (\bar{T}_R , σ_R) to vary as a function of prime type in all models, allowing the prime to affect knowledge access without necessarily affecting episodic recognition.

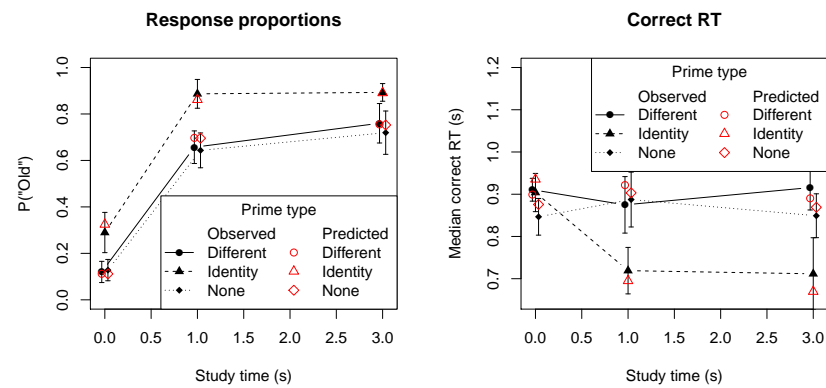
We implemented a fluency heuristic (“FI”) by allowing response bias b to vary as a function of prime type under the constraint that b and \bar{T}_R must be inversely monotonically related. This constraint was purely ordinal—we did not assume any particular functional form for the relationship between b and \bar{T}_R . Prime storage (“St”) was implemented by storing an additional memory trace corresponding to the prime. This trace is presumed to have all N_X context features stored and a proportion u_P of its content features stored (with the same encoding fidelity c_S as all other traces). Feature leakage was implemented by assuming that a proportion of the content features in the memory probe are filled with prime features at time t_0 . In the less restrictive version of this model (“L2”), the proportion of content features “leaked” into the initial probe differs between identical (p_I) and different (p_D) primes, while the restricted version (“L1”) requires that these proportions are equal ($p_I = p_D$).

Median best-fitting parameters for all models are given in Table 9. Although each mechanism in isolation generally improves the fit of the model, the overall preferred model is “L2”, in which features leak from the prime into the initial memory probe, but at different rates for identical versus different primes. When this mechanism is combined with prime storage, the estimated degree of prime storage has a median of zero; similarly, when L2 is combined with a fluency-like bias variation, bias ends up not varying much between prime types. Thus, feature leakage evidently plays a more important explanatory role than the other two accounts.

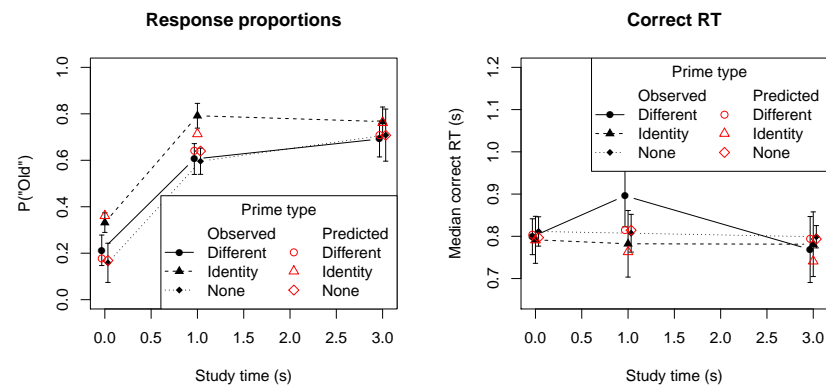
This is not necessarily the case for all participants, however. Although the majority (45/79) are best fit by a model that includes a feature leakage mechanism (either L1 or L2), 10 are best fit by a fluency-like bias shift, 8 by storage of a prime trace, and 13 by no effect of any prime. The data and model predictions for these three groups of participants are shown in Figure 12. When primes can only affect the residual time distribution (model “N”), they cannot alter response proportions, as shown in Figure 12a; it is clear that the priming effects for these participants are sufficiently small relative to the other variability in the data that no mechanism needs to be invoked to explain them for these participants. A fluency-like bias shift as a result of shorter residual time (model “FI”) can predict priming for identical primes but not, in general, for different primes since these are usually associated with either no change or an increase in residual time, as shown in Figure 12b. Storage of the prime as a memory trace (model “St”) also predicts priming for identical primes but not for different primes, but importantly does *not* entail a noticeable speed advantage for identical primes, as shown in Figure 12c. The small effect of different primes on these latter two groups of participants means that bias shifts and/or prime storage can adequately explain their results, even



(a) No prime effect ("N").



(b) Fluency-like bias shift ("Fl").



(c) Prime storage ("St").

Figure 12. Observed and predicted response proportions and median correct response time (RT) for three groups of participants that are best fit by the corresponding model variant according to AIC (Akaike Information Criterion). Error bars denote 95% confidence intervals about the mean.

Table 9

Median best-fitting parameters of each model variant to each participant from our masked priming experiment. “N” denotes no effect of prime other than on residual time, “Fl” denotes a fluency mechanism (b varies inversely with \bar{T}_R), “St” denotes a prime storage mechanism ($u_P > 0$), “L1” denotes restricted feature leakage ($p_D = p_I$), and “L2” denotes free feature leakage ($p_D \neq p_I$). “AIC” is summed Akaike Information Criterion over participants while “num. best fit” is the number of participants for which each model had the lowest AIC.

Parameter	Prime type	Model											
		N	Fl	St	L1	L2	St+L1	St+L2	Fl+St	Fl+L1	Fl+L2	Fl+St+L1	Fl+St+L2
u_1		0.562	0.579	0.587	0.549	0.545	0.559	0.551	0.598	0.553	0.54	0.575	0.546
u_3		0.577	0.601	0.628	0.577	0.587	0.591	0.575	0.627	0.572	0.583	0.579	0.583
c_S		0.895	0.886	0.891	0.893	0.902	0.901	0.914	0.883	0.893	0.904	0.907	0.916
A_0		29.477	29.485	28.861	26.783	25.892	26.175	25.892	28.861	26.651	25.49	25.051	25.171
ρ		0.01	0.01	0.01	0.009	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
\bar{T}_R	Neutral	0.32	0.324	0.337	0.353	0.367	0.364	0.373	0.337	0.366	0.369	0.375	0.383
	Identity	0.269	0.289	0.269	0.275	0.26	0.274	0.241	0.275	0.293	0.26	0.295	0.25
	Different	0.345	0.356	0.356	0.348	0.379	0.363	0.383	0.357	0.364	0.381	0.375	0.384
σ_R	Neutral	0.007	0.007	0.007	0.009	0.007	0.012	0.007	0.008	0.01	0.008	0.014	0.008
	Identity	0.16	0.153	0.16	0.159	0.15	0.153	0.15	0.152	0.155	0.144	0.152	0.143
	Different	0.011	0.01	0.009	0.01	0.008	0.013	0.012	0.009	0.011	0.009	0.011	0.011
b	Neutral	0.543	0.535	0.536	0.535	0.531	0.528	0.529	0.534	0.53	0.53	0.527	0.526
	Identity	—	0.571	—	—	—	—	—	0.563	0.556	0.539	0.542	0.535
	Different	—	0.54	—	—	—	—	—	0.535	0.527	0.525	0.527	0.523
u_P		—	—	0.151	—	—	0.151	0	0.021	—	—	0	0
p_I		—	—	—	0.3	0.521	0.292	0.489	—	0.271	0.48	0.243	0.464
p_D		—	—	—	—	0.218	—	0.224	—	—	0.23	—	0.236
AIC		16192	15972	15940	16014	15708	15842	15797	15959	15893	15935	15941	16046
Num. best fit		13	10	8	8	26	8	1	3	1	0	1	0

though, as in the complete group data, there remains evidence that different primes have a similar but smaller effect to identical primes.

Finally, we note that the differences in residual time between identical and different primes is consistent with our account of why different numbers of features can leak into the memory probe between these two different prime types. In model L2, median residual time (\bar{T}_R) following an identical prime is $0.379 - 0.260 = 0.119$ s shorter than following a different prime while the median probability of feature leakage is 0.521 following an identical prime and 0.218 following a different prime. If the difference in feature leakage is due to features “leaking out” of the probe during the time needed to access knowledge, as we suggest, and if features leak out of the probe at the same rate they are sampled, we can calculate the average time for the probability of feature activation to drop from 0.521 to 0.218: $\rho \times \frac{\log[1-(p_I-p_D)]}{\log(1-1/N_C)} \approx 0.010 \times \frac{\log[1-(0.521-0.218)]}{\log(1-1/30)} \approx 0.106$ s, nearly equal to the additional residual time needed to access knowledge following a different prime.

Discussion

By embodying various hypotheses within the context of our dynamic recognition model, we have shown that masked priming leads to surreptitious “leakage” of features from the prime into the initial memory probe, explaining the effect of both identical and different primes on recognition accuracy and response time. This is a natural outgrowth of our core assumption that recognition decisions depend on tracking *changes* in familiarity over the course of a trial, since the content features of the prime result in a different initial level of familiarity than would be expected from context alone.

The residual time component of our model encompasses the processes involved in accessing lexical traces stored in knowledge as a source of content features. Residual time was roughly equivalent between different and neutral primes, but was much faster and more variable following identical primes (see Table 9), in support of our hypothesis that identical primes pre-activate the relevant lexical trace for the test word and allow feature sampling to begin almost immediately. This result is also consistent with the serial position effects reported by McElree and Doshier (1989), in which retrieval speed was greater for the final list item in a short-term recognition task but did not differ between other serial positions; the final test item in essence serves as a prime for itself. That an identical prime provides a “head-start” to both lexical access and feature sampling is also supported by ERP data (Woollams, Taylor, Karayanidis, & Henson, 2008), in which masked identity priming is associated with an early positive component (P200) for both targets and foils—during which lexical access typically occurs (Serenio & Rayner, 2003)—as well as an earlier onset of parietal components related to episodic recognition. The increased variability in residual time following an identical prime then reflects variability in the degree to which the masked prime is able to activate a lexical trace.

Although the leakage of prime features into the probe is sufficient to account for subliminal priming effects in recognition, we suspect more complicated mechanisms will be necessary to explain longer-duration priming (e.g., Doshier, McElree, Hood, & Rosedale, 1989). Although positive priming effects are observed at the short durations we used, long prime durations—allowing for a clear percept of the prime to form—yield either smaller effects or even a negative priming effect (Jacoby & Whitehouse, 1989). This crossover has been attributed to either active discounting of prime features when the prime is known to be uninformative (Huber et al., 2001) or habituation to the features of the prime (Huber & O'Reilly, 2003). The effect of leaked prime features on the starting level of familiarity can be viewed as a form of habituation, in that those features get treated as part of the context rather than as part of the test item itself (Huber, Curran, Clark, & Winkielman, 2008). Clearly-presented primes, on the other hand, may be more likely than masked primes to be stored as traces in episodic memory and/or form a compound probe with the test item (Ratcliff & McKoon, 1988). The compounding together of prime features and test item features would result in a probe that represented a mixture of the two items, leading to enhanced performance if the prime and test item were similar (i.e., shared features) but interference if they were not (Gillund & Shiffrin, 1984; Ratcliff & McKoon, 1988).

Section 4: Delayed Content Features

Given that recognition decisions depend on tracking changes in familiarity as features join an evolving probe of memory, it is clear that the dynamics of recognition depend on the time at which features enter the available pool and subsequently join the probe. In Section 2, we explored how the dynamics of list discrimination could be explained in terms of when context features became available to help reinstate the target context. In this section, we examine the dynamics with which *content* features become available, contrasting features that are available from the onset of retrieval with those that are delayed (Brockdorff & Lamberts, 2000). This distinction between early and late information may be contrasted with a distinction between separate fast and slow recognition processes (e.g., Mandler, 1980); although some information may be delayed, it still ends up contributing to a single ongoing recognition process. Our dynamic approach thus disentangles processes, which unfold in time, from the static information that is acted upon by those processes (analogous to the contrast between cognitive algorithm—process—and representational content—

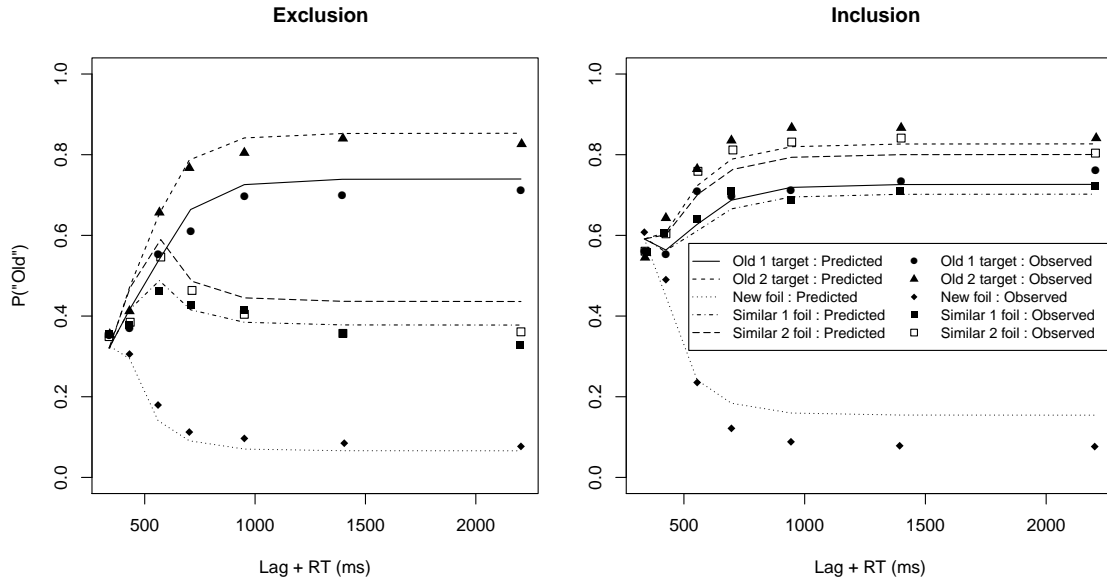


Figure 13. Observed and predicted probability of responding “old” as a function of response signal lag plus response time from Hintzman and Curran (1994, Exp. 3). Parameter values used to generate predictions were $u_1 = 0.413$, $u_2 = 0.492$, $c_S = 0.978$, $\theta_{\text{Inclusion}} = 0.565$, $\theta_{\text{Exclusion}} = 1.440$, $p_P = 0.150$, $t_P = 326$, $p_I = 0.0613$, $t_0 = 84.7$, and $p = 5.13$. See text for details.

information—made by Cowell, 2012). Specifically, we contrast the usual case in which studied words are recognized by their semantic content and base word-form with those in which the word’s plurality or modality are also relevant to the decision. The features that carry information about plurality and modality do not enter the pool of available features until after the features carrying the word’s semantic and base (i.e., uninflected) form information, which are available as soon as lexical knowledge of the word is accessed.

Delay of plurality and modality features

Hintzman and Curran (1994) report a set of response signal studies in which participants studied a list of nouns. In their second and third experiments, test items comprised old words, 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 13, “exclusion”). Similar results were found in a set of response signal experiments by McElree et al. (1999) in a study of source memory, based on the exclusion procedure of Jacoby (1999). In their experiments, participants were asked to recognize only those items that had been studied in the auditory modality, and to reject novel words or words that had been studied visually. As shown in Figure 14, the dynamics of these judgments closely resemble those found by Hintzman and Curran (1994) for reversed-plurality foils in their exclusion

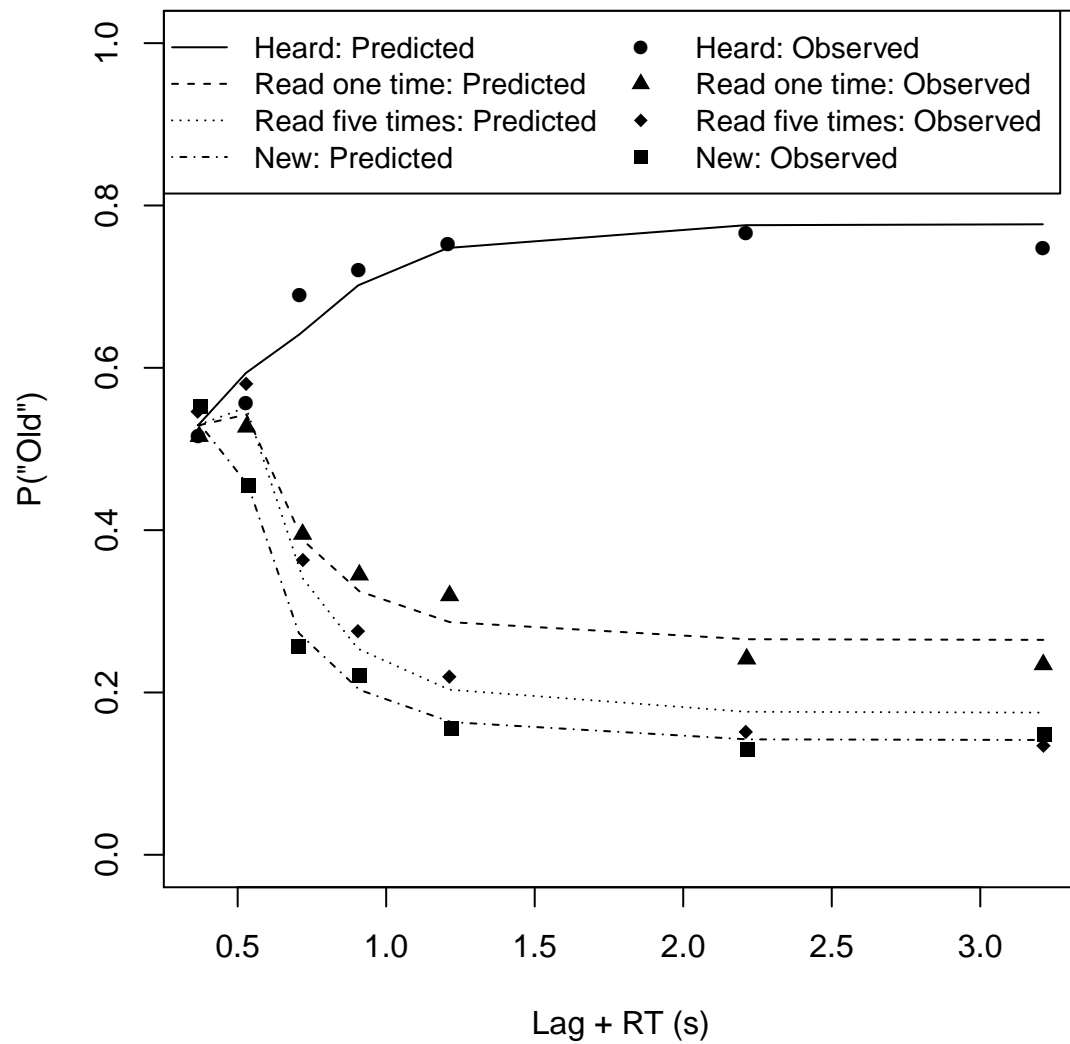


Figure 14. Probability of endorsing a word as a function of response signal lag and response time in McElree, Dolan, and Jacoby (1999, Exp. 2). Parameter values used to generate model predictions are were $u_A = 0.483$, $u_{V1} = 0.604$, $u_{V5} = 0.987$, $c_S = 0.972$, $p_M = 0.466$, $t_M = 183$, $\theta = -0.117$, $m = 1.801$, $t_0 = 87$, $\rho = 10.4$. See text for details.

condition, with a reversal in false alarms to words from the non-target modality⁶.

One possibility is that participants use the test word as a cue to retrieve a memory trace of the corresponding word. If the plurality information stored in the trace fails to match that of the test word (Hintzman & Curran, 1994), or if the modality information stored in the trace is visual rather than auditory (McElree et al., 1999), the participant rejects the item. This retrieval process, often termed recollection, is assumed to operate more slowly than the process by which other word information is retrieved, thereby resulting in a reversal of the false-alarm curve. From a dynamic perspective, however, a stronger memory trace should be retrieved more quickly than a weaker memory trace, such that the false-alarm curve should reverse earlier the more an item is repeated at study.

Instead, the inflection point for repeated items in both experiments is the same regardless of the number of repetitions. This is more consistent with recognition dynamics being governed by the time at which the relevant plurality or modality information becomes available to enter the memory probe. Because the memory probe evolves over time independently of the memory traces formed at study, plurality or modality features should become available at the same time regardless of the strength of the target item. This time, though, is clearly later than the time at which the other features of the word—its semantic features and base word-form—are available, thus yielding a reversal in false alarms to similar foils (Brockdorff & Lamberts, 2000). In phenomenological terms, this means that participants first obtain some information about the word itself and then “imagine” experiencing it in the appropriate plurality or modality, thereby adding additional features to the probe and allowing words to be rejected if these features fail to match anything in memory.

Quantitative Model. Under the assumption that plurality or modality information may not be available until later in a trial, we fit a version of our dynamic recognition model to Experiment 3 from Hintzman and Curran (1994) and to Experiment 2 from McElree et al. (1999).

Plurality. A certain proportion of content features, p_P , represent the plurality of a word and become available for sampling at time t_P ⁷. 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* 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

⁶We note that reversals in false-alarm curves could result from different response biases at different signal lags (Rotello & Heit, 1999), but there is no reason to expect this to occur nor is there evidence for such shifts except in extreme cases (Hintzman et al., 1994).

⁷The proportion of features dedicated to representing plurality may be less in experiments where plurality is not critical for distinguishing targets from foils.

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. As shown in Figure 13, the model captures the trends in the data both qualitatively and quantitatively.

Modality. We assumed 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 ⁸. Modality-related features may be stored at a different rate than other content features (like semantic features), as represented by the parameter m : The probability of storing a modality-related feature is u^m , such that modality features could be stored at a greater rate than other features if $m < 1$, and at a lower rate if $m > 1$.

Semantic and word-form features are available early

The close quantitative match between our model and the data reported by Hintzman and Curran (1994) and McElree et al. (1999) should not be surprising, given that it was deliberately tailored to fit those data. However, it serves as a useful contrast to situations in which word events are distinguished by features that are integral to their representation, namely, their semantic features and base word-form. If, as we argue, such features are present in the available pool as soon as knowledge has been accessed, there should be no evidence for any reversals of false alarms to similar foils. Consistent with this prediction, when foils must be rejected on the basis of their base word-form (Rotello & Heit, 1999, see Figure 15) or semantic features (Heit et al., 2004, see Figure 16), there is no evidence that the relevant information is delayed.

Quantitative Model.

Word-form. 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 the base word-form of the stimuli. The “unpaired” condition of Rotello and Heit (1999, Exp. 2) replicates the general design of the exclusion condition of Hintzman and Curran (1994, Exp. 3), with two exceptions: strong items were studied three times instead of twice; and similar foils were generated not by adding or subtracting final letters, but by swapping internal letters (e.g., PRUMIR to PRAMAD). As a result, even similar foils differed in their overall orthographic form, even if some letters overlapped. As shown in Figure 15, 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 the orthographic features that distinguish old items from similar foils. We note that our model still predicts a slight nonmonotonicity in the false alarm curves to similar foils as a consequence 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.

Semantics. 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 16), three sets of instructions were given to different groups of participants with regard to the recognition test: 1) a

⁸These modality-related features might not be used solely to encode modality *per se*, but also other features related to modality.

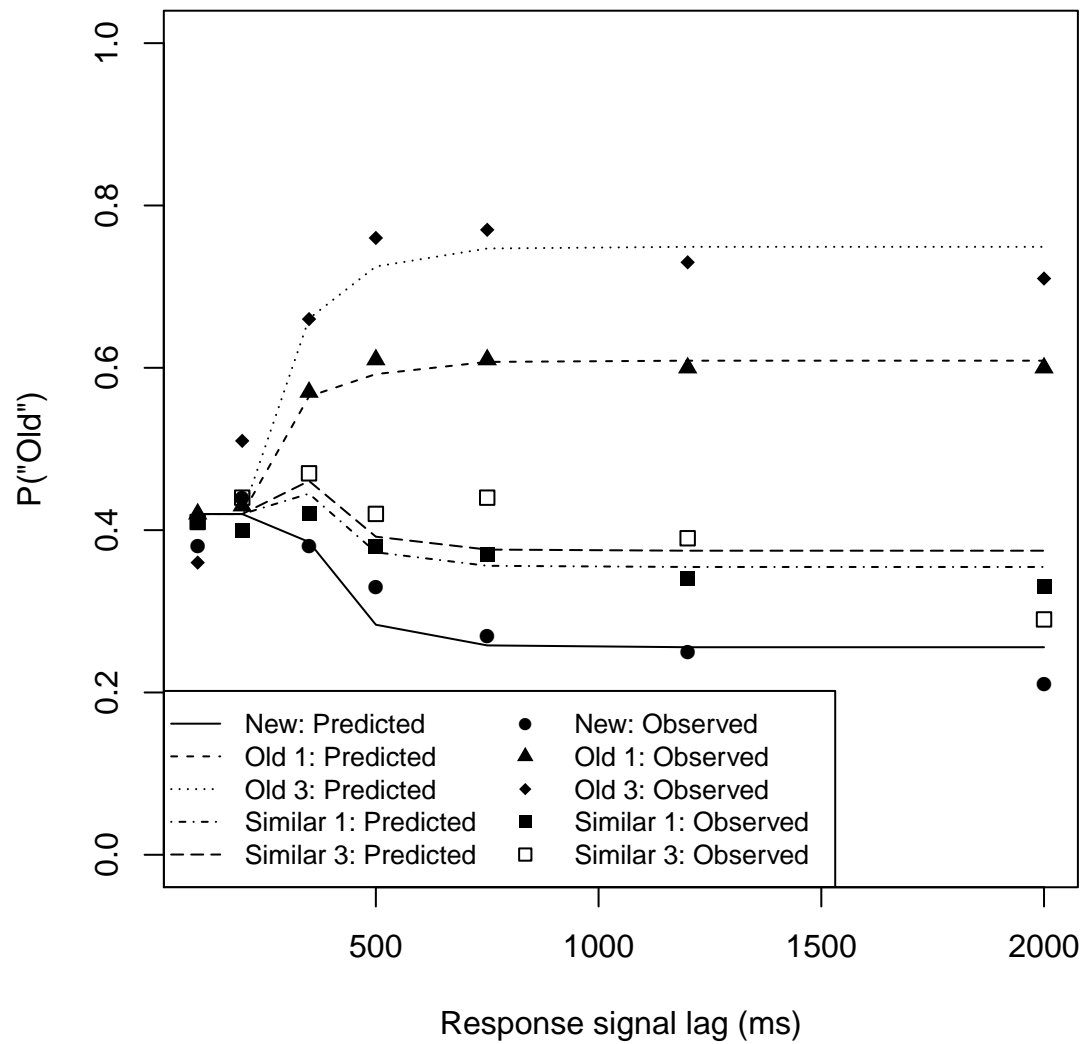


Figure 15. Observed and predicted probability of responding "old" as a function of response signal lag for the three "unpaired" condition in Rotello and Heit (1999, Exp. 2). Parameter values used to generate predictions were $u_1 = 0.298$, $u_3 = 0.388$, $c_S = 0.974$, $s = 0.711$, $\theta = 0.323$, $t_0 = 196$, and $\rho = 3.449$. See text for details.

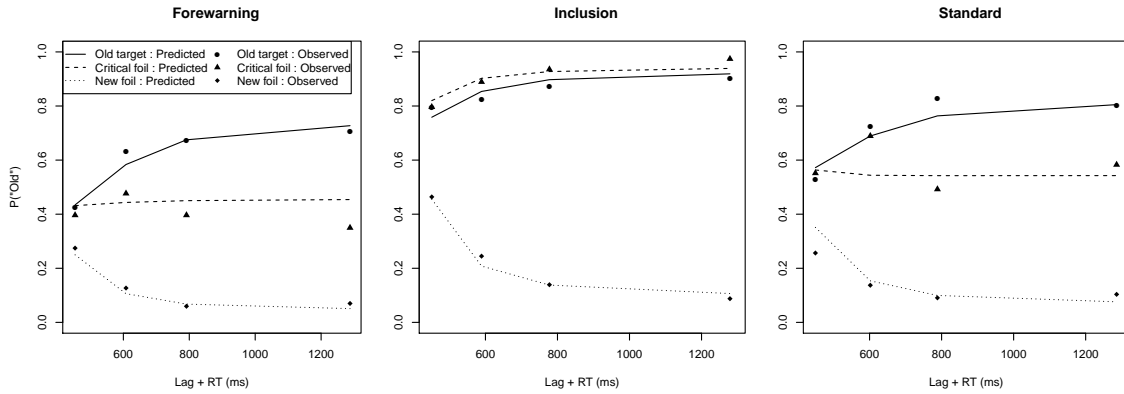


Figure 16. Observed and predicted probability of responding “old” as a function of response signal lag plus response time for the three instructional conditions in Heit, Brockdorff, and Lamberts (2004, Exp. 2). Parameters values used to generate predictions were $u = 0.444$, $c_S = 0.971$, $p_S = 0.736$, $p_{S,Inc} = 1.000$, $s_C = 0.769$, $\theta_{Forewarning} = 1.709$, $\theta_{Inclusion} = 0.155$, $\theta_{Standard} = 0.901$, $t_0 = 61.5$, and $\rho = 7.31$. See text for details.

“standard” set of instructions in which participants were told to respond “old” only to words that were on the study list; 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 of each probe/trace encode the semantic features pertaining to each critical word, and that this proportion is higher in the Inclusion condition (denoted $p_{S,Inc}$) owing to the increased attention to similarity in response to instructions (which were given prior to study and hence could influence both encoding and retrieval; for further support of this conclusion, see Gallo, Roediger, & McDermott, 2001). We modeled similarity within each group of associates by assuming that each associate has probability s_C of sharing one of its p_S content feature values with that of the critical word, otherwise its value is random (and could, therefore, match by chance). 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.

As shown in Figure 16, the model captures the critical finding that, in the Standard and Forewarning conditions, participants are initially likely to false alarm to the critical word, but this tendency does not increase as time passes. This is a direct consequence of the 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 ($p_{S,Inc} \approx 1$) to maintain a high rate of endorsement of the critical word. If shared semantic features were available earlier than item-specific features, the critical word would be more likely to be recognized than actual old words at early lags even in the Standard and Forewarning conditions—that this is not the case provides additional qualitative evidence that the semantic features of a word are treated in an integral fashion and are all available at the same time. We also note that 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 most conservative in the Forewarning condition, and that forewarning participants of

similar foils only appears to affect their willingness to respond “old” rather than any other aspects of encoding or retrieval (in accord with Gallo et al., 2001). Other studies have, however, found effects of forewarning on retrieval beyond having a more stringent criterion (Starns, Lane, Alonzo, & Roussel, 2007), which could, in contrast to inclusion instructions, reflect allocation of attention away from shared category features such that fewer of them ended up in the pool available for sampling.

Discussion

In this section, we have taken a dynamic approach to understand when and how certain kinds of information become available for retrieval, distinguishing between semantic and word-form features which are available early from plurality and modality information which are available later. Our approach differs from traditional dual-process accounts in that we separately consider the kinds of information that are relevant for particular memory tasks from the processes that make use of this information. Although plurality and modality information are delayed relative to other word features, they are used in the same way to drive changes in familiarity over time. The accounts of false recognition provided by fuzzy trace theory (Brainerd & Reyna, 2002) and conjoint recognition theory (Brainerd, Reyna, & Mojardin, 1999) also assume a difference in the content of the information used for recognition, specifically between “gist” and “verbatim” information. In our model, the gist/verbatim distinction manifests as the degree to which features are shared between a target and a related foil. Likewise, although it would be possible to label early semantic and word-form features as “familiarity-based” and later features as “recollection-based”, doing so conflates processes and the information those processes act upon (Cowell, Bussey, & Saksida, 2010; Cowell, 2012). Because processes unfold in time, we believe they are best identified by their dynamic properties: For example, in the General Discussion, we consider a case in which retrieval dynamics were found to be affected by the strength of the target memory trace (Hintzman & Caulton, 1997), unlike any of the situations considered here; this is a signal that something other than the features in the memory probe are driving and we present an extension of our model that employs a recollection-like feature recovery mechanism.

It is worth considering *why* plurality and source do not enter the pool of available features until relatively late in a trial. We suggested above that the delay could result from participants manufacturing these features themselves and adding them to the probe, since they are not necessarily present in the test item itself. It is unlikely that the delay for plurality-related features is result of a strategy adopted by participants specifically for that task, since such information is not easily ignored or suppressed (this is embodied in the imperfect suppression of plurality features in our model; Jou & Cortes, 2012). There is more support for the idea that the delay for modality-related features could 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). At the risk of conflating two already overloaded terms, particularly when presentation modality is blocked at study (as in McElree et al., 1999), participants may attempt to reinstate the context associated with the modality, as described in Section 2, rather than retrieve anything about the modality *per se*. In either case, when the relevant modality or context information is not already present in the environment or in the presentation of the test item, participants must generate (“imagine”) that information themselves and add it to the memory probe; the need for similar processes in both context reinstatement and source memory also highlights the need for an operational definition of “context”

that can distinguish it from other constructs⁹. On the other hand, by analogy to the Stroop effect, it may not be that modality or plurality are “delayed”, per se, but rather the semantic and base form features of a word are available exceptionally *quickly*. Indeed, the semantic features that constitute a word’s meaning are accessed very rapidly (Gibson, 1971; LaBerge & Samuels, 1974; Feldman, O’Connor, & Del Prado Martín, 2009), even faster than single letters can be identified within the word (Johnson, 1975), suggesting that such properties may be a fundamental characteristic of access to lexical knowledge.

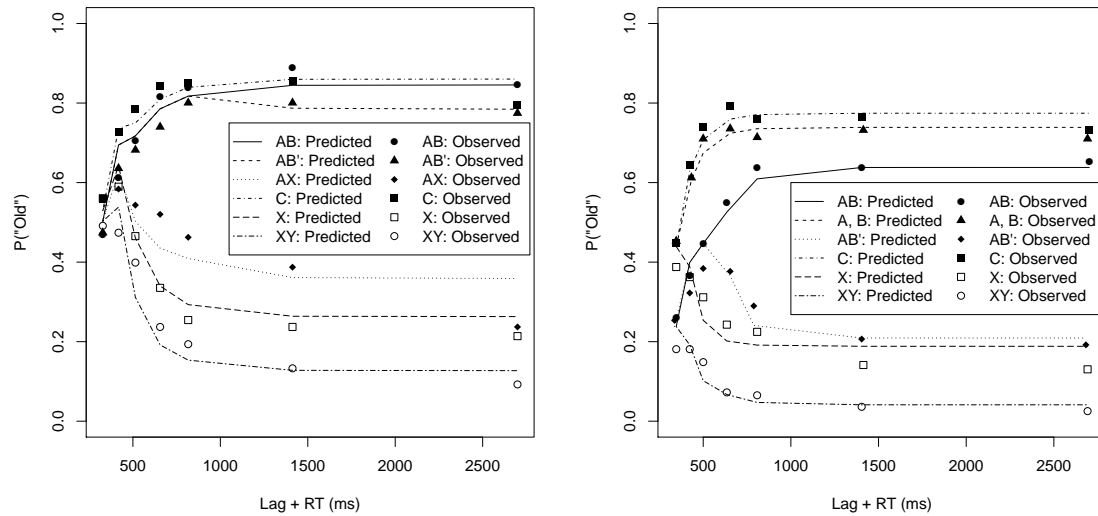
Section 5: Associative Recognition

Another domain of memory that involves distinctions between early and late retrieval is the recognition of multiple items and their associations. As in Section 4, we argue that the dynamics of associative retrieval are governed by the time at which different features become available for sampling. Although features of individual items are available at the beginning of the trial, participants use these item features to generate novel features that represent the association between the items. The delay with which associative features are available can be attributed to the need to generate these features each time a combination of otherwise-unrelated items is presented.

Associative recognition requires that participants discriminate between pairs of items that had been studied together—called “intact” pairs—from pairs of items that had been studied, but at different times—“rearranged” pairs. Recognizing the items individually is insufficient to make this distinction—some additional information is required. Dual-process theories propose that, in associative recognition, participants engage in a recall or recollection process that is qualitatively different from single-item recognition (e.g., Rotello & Heit, 2000; Malmberg & Shiffrin, 2005; Buchler, Light, & Reder, 2008; Malmberg, 2008). Specifically, participants are hypothesized to use one member of the pair as a recall cue; if the item recalled from memory in response to this cue matches the other member of the test pair, the pair is judged to be “intact”; if not, it is judged to be “rearranged”. Alternatively, studying a pair might result in storage not just of the individual elements of the pair, but of relational information representing the compound of the two items. If so, this relational information—what we refer to as associative features—could form a compound cue which is then used similarly to the item cues in typical recognition, comparing this additional information to the associative features stored in memory (Murdock, 1982; Doshier & Rosedale, 1989, 1997).

A wealth of research has attempted to distinguish between these two views of associative recognition on the basis of accuracy data. The shape of receiver-operating characteristic (ROC) curves has been argued to support the involvement of recall in associative recognition (Yonelinas, 1997), but these data are equally consistent with a signal detection process that combines two sources of evidence, as in a compound cue model (Wixted, 2007). Process-dissociation procedures (Jacoby, 1991) have also been used to argue in favor of a recall-like process in associative recogni-

⁹An interesting alternative in the case of source information 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. Supplying these features is analogous to contextual reinstatement in that only the features of the current context are immediately available, while the participant must restate 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.



(a) Experiment 1 (inclusion instructions).

(b) Experiment 2 (exclusion instructions).

Figure 17. 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 10.

tion. However, process-dissociation will 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 et al., 1995). Process-dissociation also relies on the strong assumption that familiarity and recall are stochastically independent of one another; when the assumption is violated, estimates obtained from this procedure are uninterpretable (Curran & Hintzman, 1995; Hillstrom & Logan, 1997). The assumption of independence is also inconsistent with positive correlations between the probability of correct recall and that of correct recognition (Tulving & Wiseman, 1975). Finally, even if the assumptions of the process dissociation procedure are satisfied, it is insufficient for identifying the relevant processing components (Humphreys et al., 2000).

Given the inability of accuracy data alone to shed much light on the nature of the processes involved in associative recognition, we argue that a dynamic approach is more informative. Gronlund and Ratcliff (1989) report a set of response signal experiments that are particularly illustrative, the results of which are reproduced in Figure 17. In their first experiment, 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 addition, participants were tested on single items and had to distinguish between studied (C) and unstudied (X) words, where associative information is once again irrelevant. 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 (although participants were, again, also tested on single-item recognition). Under the Experiment 2 exclusion instructions, participants initially endorsed rearranged pairs (AB') before

this tendency reversed at later signal lags. The nonmonotonic false alarm curve to AB' pairs under exclusion instructions indicates the involvement of a process or source of information that is not available early in associative recognition.

As mentioned above, this process could be a form of recall, or as in Section 4, it could be a signal that new features—in this case, associative features—become available later than the features that encode the individual items. Reiterating the logic from Section 4, to the extent that the dynamics of associative recognition are invariant with respect to the strength of the target information in memory, this is evidence that dynamics are driven by the time at which features are able to join the memory probe. If recall is involved, as the target memory trace gets stronger—e.g., as a function of increased study time or repetition—the stored associate should be retrieved more quickly (e.g., Raaijmakers & Shiffrin, 1981; Anderson, 1983; Sederberg, Howard, & Kahana, 2008). This would manifest as an earlier inflection point and/or faster approach to asymptote in the false alarm curve to AB' pairs. If, instead, associative recognition depends on the delayed availability of associative features, there should be no difference in the dynamics of associative recognition as a function of memory strength, only a difference in asymptotic performance.

These predictions were directly tested by Doshier (1984a), who presented pairs for varying amounts of time at study and tested participants on both intact pairs and rearranged pairs formed by mixing the left and right members from pairs that had been studied for equal amounts of time. Although her initial analyses suggested that repetition—but not study time—led to faster processing of the pair, a modern Bayesian analysis of those results, presented in Appendix D, finds no evidence that either the inflection point or rate of approach to asymptote changes as a function of study time or repetition. Instead, only asymptotic accuracy differs with strength. This mirrors the findings of Nobel and Shiffrin (2001) and Wickelgren and Corbett (1977), in which memory strength was manipulated both by study time and by list length, yet had no effect on the dynamics of associative recognition, only on asymptotic accuracy (see Appendix D). An additional test of a recall-based account of associative recognition was conducted by Gronlund and Ratcliff (1989, Exp. 4), in which the first word of each test pair was presented 200 ms before the second; if participants were using a recall-like 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, although participants could reject completely novel foils (XY) more quickly when given a 200 ms head start, the extra time conferred no advantage in rejecting AB' foils (while a longer head start could have more of an effect, too long and it would force participants to engage in cued recall, e.g., Wickelgren & Corbett, 1977).

In sum, the dynamics of associative recognition do not resemble those that would be expected if recall were operating. Instead, associative recognition seems to depend on a set of associative features that is delayed relative to the features used to encode the items. We argue that this delay is due to the fact that associative features, at least for novel associations, are formed in the moment and must therefore be built from item features; only when enough item features are available can the association begin to emerge. We present a formal model of this process as an outgrowth of our model of single-item recognition and show that it provides an excellent qualitative and quantitative account of associative recognition.

Quantitative model

Representing associations. Although there are many ways associations might be represented (see, e.g., Criss & Shiffrin, 2005; Murdock, 1982), we adopt the representation depicted in

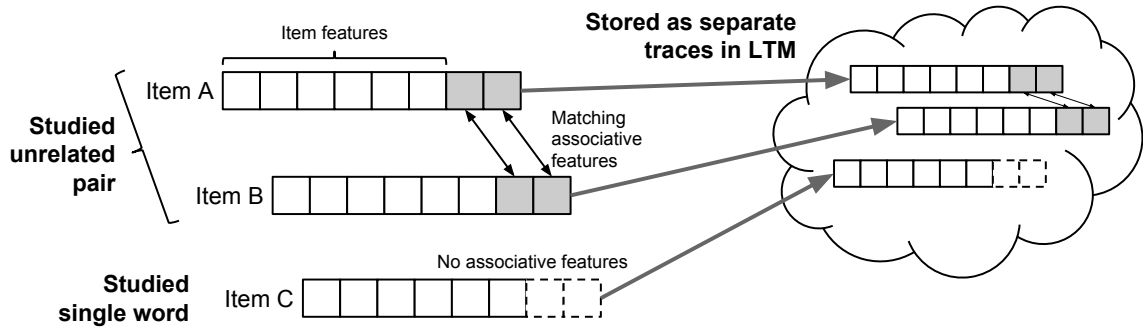


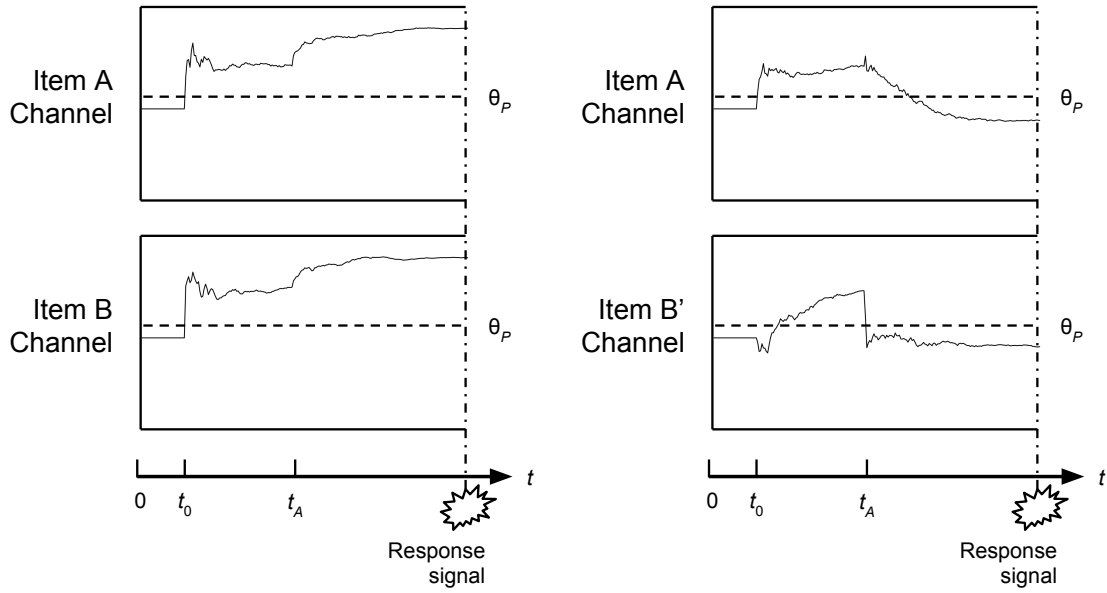
Figure 18. Schematic depiction of how studied items and pairs are stored as memory traces. See text for detail.

Figure 18. When a pair 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 emergent associative features. These associative features are “configural” in that they go beyond the features of either word alone and are “emergent” in that they arise from the relations between the features of each item. These 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). Studying a single item does not lead to storage of any of these emergent associative features. 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.

Pair and associative recognition. In pair recognition, participants should call a pair “old” if both of its members were studied, regardless of whether they were studied together (e.g., AB or AB’). 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, the 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 associative features become available later. 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 19. The model allows for both single- and multiple-item recognition 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).

Model fits.



(a) Intact (AB) pair.

(b) Rearranged (AB') pair.

Figure 19. Schematic depiction of our model for associative recognition. Each member of the pair is processed in parallel channels that each operate as single-item recognition processes; sample paths for memory evidence $x(t)$ are shown for each channel.

Table 10

Best-fitting parameters of the dynamic model for associative recognition to various datasets.

	Gronlund and Ratcliff (1989)		Rotello and Heit (2000)
	Exp. 1	Exp. 2	Exp. 1
u_S	0.316	0.414	—
u_P	0.339	0.388	0.439
c_S	0.988	0.980	0.984
p_A	0.017	0.170	0.162
t_A	0.752	0.412	0.437
p_S	0.937	0.869	—
θ_S	-0.118	0.245	—
θ_P	-0.891	0.048	1.098
t_0	0.089	0.094	0.083
ρ	0.006	0.004	0.008

Table 11

Best-fitting parameters of the dynamic model to individual participant and aggregate data from Doshier (1984a, Exp. 2).

Participant	u_1	u_2	u_4	u_6	c_S	p_A	t_A	θ	t_0	ρ
B.G.	0.529	0.586	0.567	0.602	0.932	0.082	0.190	0.257	0.100	0.005
D.B.	0.428	0.461	0.475	0.497	0.963	0.143	0.411	0.708	0.241	0.007
E.G.	0.346	0.489	0.522	0.533	0.974	0.313	0.430	-0.692	0.006	0.014
J.J.	0.407	0.419	0.448	0.478	0.971	0.162	0.634	-0.331	0.002	0.010
W.A.	0.328	0.394	0.423	0.437	0.968	0.167	0.435	0.243	0.189	0.007
Participant median	0.407	0.461	0.475	0.497	0.968	0.162	0.430	0.243	0.100	0.007
Aggregate data	0.356	0.420	0.437	0.463	0.969	0.190	0.447	-0.228	0.187	0.010

Gronlund and Ratcliff (1989). 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 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 17, showing that the model closely matches the data, with the fitted parameter values in Table 10. 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.

Rotello and Heit (2000). In Experiment 1 reported by Rotello and Heit (2000), participants only encountered pairs during study and test, never single items. Accordingly, we assumed that $p_S = 1$ in this experiment since, without tests of both single items and pairs, this parameter becomes unidentifiable, although it remains likely that participants are subject to capacity limitations on the number of item features that can be sampled into the two probes. 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 20 while best-fitting parameters are given in Table 10. Once again, these data demonstrate a non-monotonic false alarm curve to AB' pairs and not to XY pairs.

Doshier (1984a). Doshier (1984a) reported two experiments in which participants only had to distinguish between intact (AB) and rearranged (AB') pairs, where the strength of the pairs was varied either by study time (Experiment 2) or study time and repetition (Experiment 3). In Experiment 2, pairs were studied for either 1, 2, 4, or 6 seconds each while in Experiment 3, pairs were studied for single presentations of 1, 2, or 5 seconds or two presentations of one second each. Foils were formed by rearranging items within each level of study/time repetition. Because novel

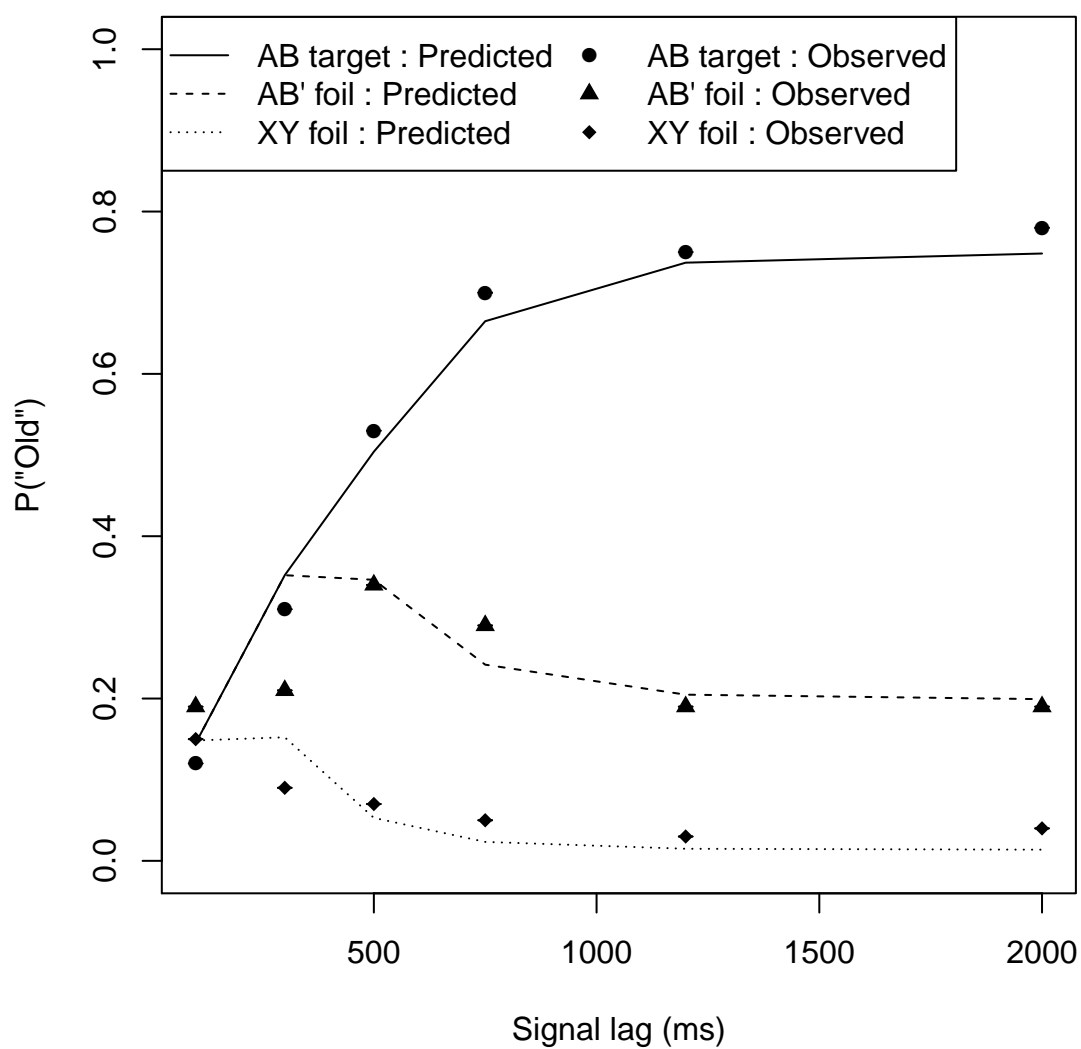


Figure 20. Probability of endorsing a test item as a function of response signal lag from Rotello and Heit (2000, Exp. 1), along with predictions from our dynamic recognition model. Parameters used to generate these predictions are given in Table 10.

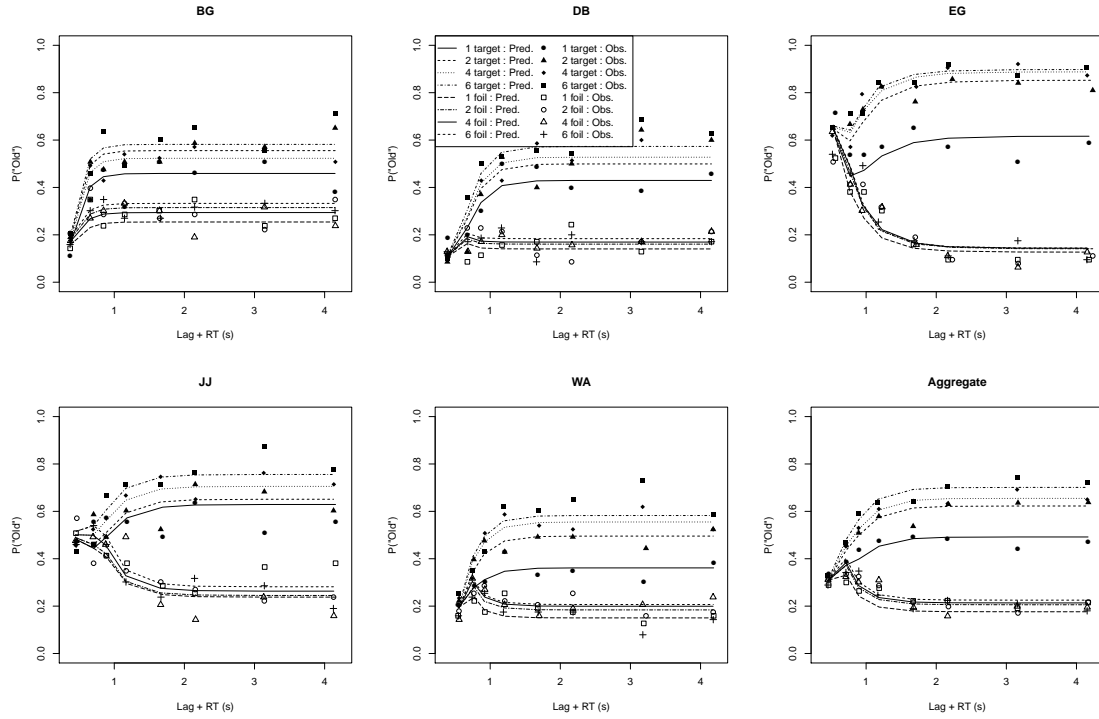


Figure 21. Probability of endorsing a test item as a function of response signal lag and response time for each participant from Doshier (1984a, Exp. 2), along with predictions from our dynamic recognition model. Parameters used to generate each participant's predictions are given in Table 11.

Table 12

Best-fitting parameters of the dynamic model to individual participant and aggregate data from Doshier (1984a, Exp. 3).

Participant	u_1	u_{1+1}	u_2	u_5	c_S	p_A	t_A	θ	t_0	ρ
A.H.	0.560	0.671	0.634	0.717	0.971	0.309	0.059	0.367	0.002	0.011
M.H.	0.381	0.511	0.502	0.684	0.982	0.496	0.117	-0.181	0.010	0.014
M.W.	0.515	0.659	0.604	0.647	0.955	0.235	0.079	0.260	0.000	0.010
R.S.	0.552	0.653	0.623	0.661	0.937	0.179	0.111	1.466	0.006	0.012
S.N.	0.488	0.624	0.589	0.617	0.969	0.302	0.131	0.200	0.005	0.013
Participant median	0.515	0.653	0.604	0.661	0.969	0.302	0.111	0.260	0.005	0.012
Aggregate data	0.483	0.596	0.562	0.614	0.966	0.280	0.111	0.236	0.000	0.011

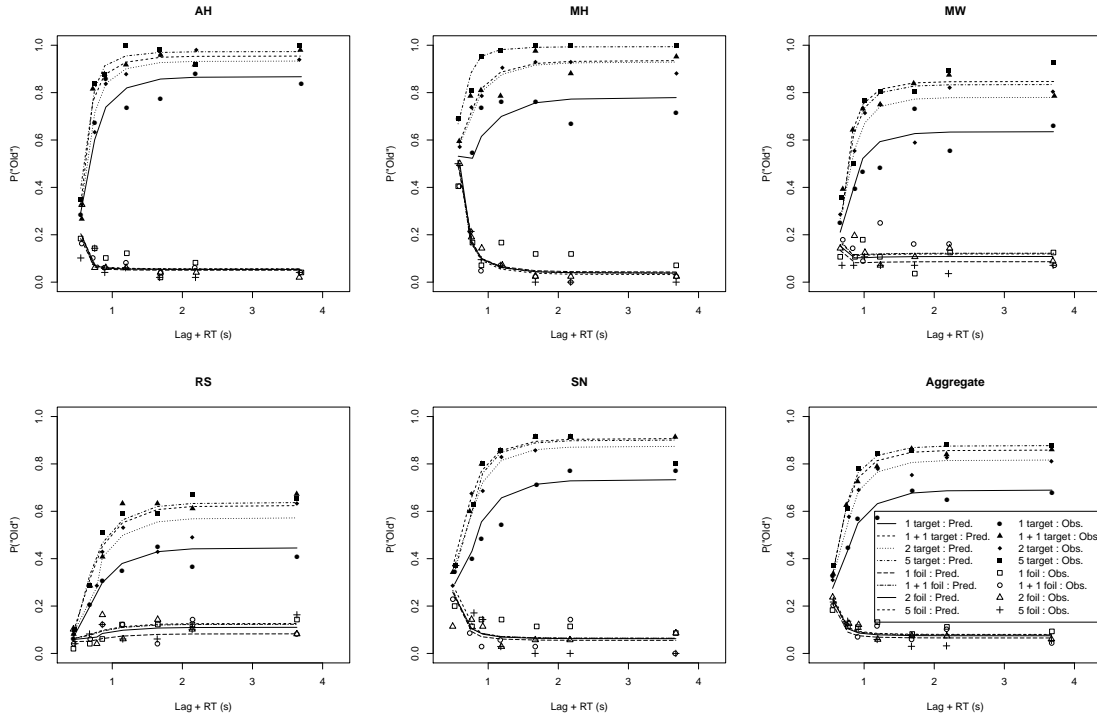


Figure 22. Probability of endorsing a test item as a function of response signal lag and response time for each participant from Doshier (1984a, Exp. 3), along with predictions from our dynamic recognition model. Parameters used to generate each participant's predictions are given in Table 12.

items never appeared in these experiments, distinguishing targets and foils relies exclusively on associative information. We assumed that each level of study time/repetition was associated with a particular storage probability u and that p_A proportion of features are associative features that become available at time t_A after stimulus onset. The best-fitting parameters for the participants and aggregate data from Experiment 2 are given in Table 11 with the associated predictions shown in Figure 21. For Experiment 3, the best-fitting parameters are given in Table 12 with predictions shown in Figure 22. Participants had overall better asymptotic performance in Experiment 3, which is modeled by increasing feature storage probabilities as well as a higher proportion of critical associative features (p_A).

Estimates for t_0 and t_A are on average lower for Experiment 3 than Experiment 2. This could be an artifact of our simplifying assumption that feature sampling occurs only during the interval between stimulus onset and the response signal. Response latencies were on average longer in Experiment 3 than Experiment 2, and it is possible that participants in fact processed stimuli for slightly longer after the response signal but before they began to initiate a response. Because our simple model has no mechanism for this, it must account for it by allowing feature sampling to begin earlier.

Wickelgren and Corbett (1977). Wickelgren and Corbett (1977) studied the dynamics of associative recognition under different levels of associative interference. Participants studied a mixture of pairs and triplets of words and, at test, were asked to distinguish between pairs of words that had occurred together (intact pairs) or on different study trials (rearranged pairs). Words that had

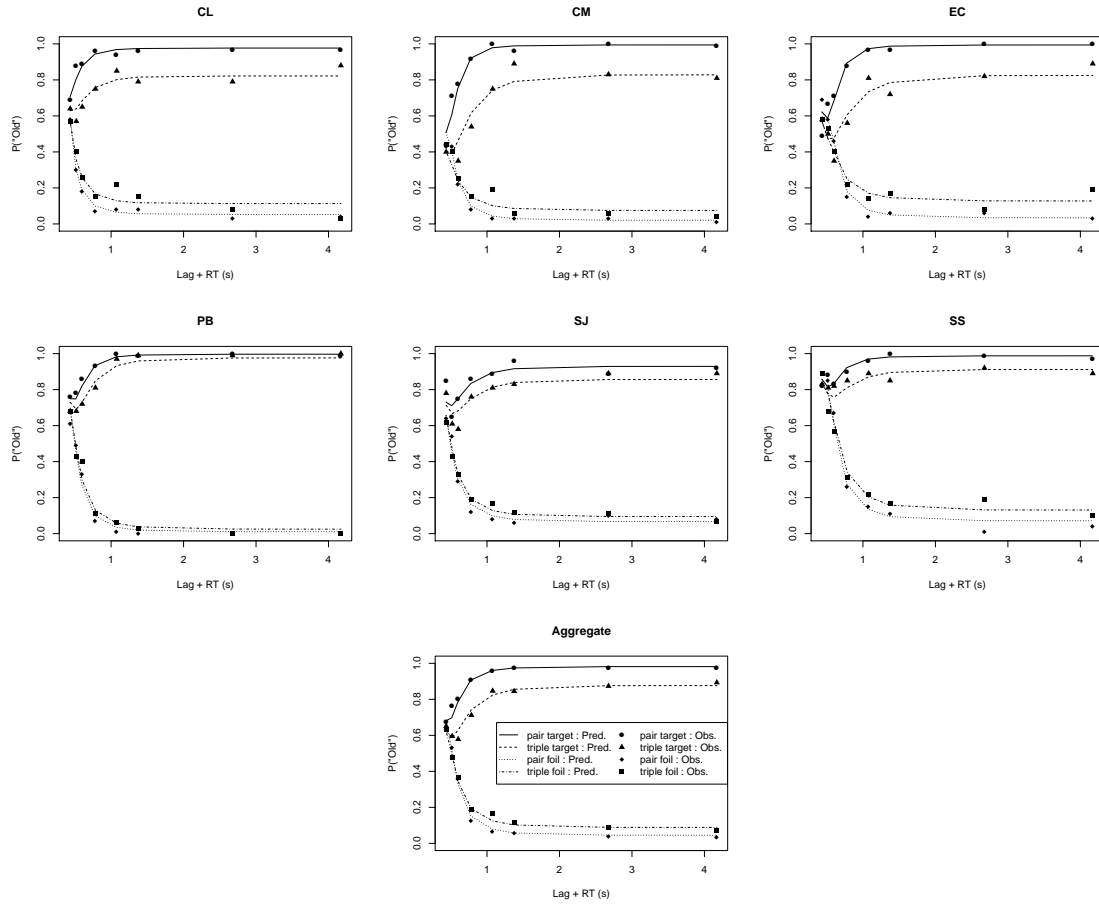


Figure 23. Probability of endorsing a test item as a function of response signal lag and response time for each participant from Wickelgren and Corbett (1977), along with predictions from our dynamic recognition model. Parameters used to generate each participant's predictions are given in Table 13.

Table 13

Best-fitting parameters of the dynamic model to individual participant and aggregate data from Wickelgren and Corbett (1977).

Participant	u_P	u_T	u_{TA}	c_S	p_A	t_A	θ	t_0	ρ
C.L.	0.739	0.610	0.355	0.955	0.520	0.179	-2.081	0.002	0.009
C.M.	0.870	0.637	0.393	0.963	0.430	0.278	-0.666	0.010	0.013
E.C.	0.832	0.623	0.326	0.965	0.428	0.350	-1.002	0.067	0.013
P.B.	0.745	0.685	0.520	0.973	0.554	0.194	-1.640	0.050	0.014
S.J.	0.600	0.562	0.436	0.959	0.474	0.148	-1.971	0.011	0.011
S.S.	0.779	0.646	0.438	0.947	0.528	0.334	-3.233	0.004	0.012
Participant median	0.762	0.630	0.415	0.961	0.497	0.236	-1.805	0.011	0.012
Aggregate data	0.763	0.613	0.452	0.956	0.444	0.265	-1.877	0.004	0.012

Table 14

Bootstrapped median and 99% confidence intervals for best-fitting associative recognition parameters from Doshier (1984a) and Wickelgren and Corbett (1977). Estimates were made from 1,000,000 bootstrap samples.

Parameter	Median (99% CI)
c_S	0.964 (0.955–0.971)
θ	-0.269 (-1.640–0.260)
p_A	0.334 (0.167–0.496)
t_A	0.219 (0.117–0.411)
t_0	0.011 (0.002–0.067)
ρ	0.011 (0.009–0.013)

been studied as part of a triple were, therefore, subject to greater interference at both the item level (less time can be distributed across the items) and the associative level (each word in the triple is associated with not just one, but two other words). Our proposed model instead matches performance in this experiment quite well, as shown in Figure 23. We assumed that feature storage probability could vary between words that had been studied as part of a pair (u_P) or a triple (u_T). Although we assumed that the probability of storing both item and associative features were equal for pairs, we allowed the probability of storing an associative feature to differ for triples (u_{TA}). All other model parameters would assumed to be equal between pairs and triples, with the best-fitting parameters reported in Table 13. For all participants, feature storage was lower for triples than pairs, as would be expected if the same amount of study time had to be distributed over three rather than two items ($u_T < u_P$). Storage of associative features is even more strongly impaired for triples than pairs ($u_{TA} < u_T < u_P$), suggesting that storing such features requires additional effort or time beyond that required for item features.

Parameter distributions. The three experiments above for which we fit individual participant data demonstrate a variety of both experimental manipulations and model parameters. We first note that the median of individual participant parameters is close to the best-fitting parameters to aggregate data from each experiment, particularly c_S , p_A , t_A , t_0 , and ρ (bottom two rows of Tables 11, 12, and 13). Feature storage probabilities (u parameters) tend to vary more between participants, but the aggregate fits are qualitatively in line with the median participant. In sum, we believe this suggests that group SAT functions are generally representative of individual performance, implying that our fits to group data in Gronlund and Ratcliff (1989) and Rotello and Heit (2000) are unlikely to reflect artifacts of averaging.

However, we wished to better understand how model parameters might be distributed in the general population and how they might relate to one another. We began by collecting the six parameters shared across experiments (c_S , θ , p_A , t_A , t_0 , and ρ) from each of the sixteen participants. While it may be that differences between participants are a function of experimental manipulations, we hoped that by pooling across experiments, we could build a better picture of how these parameters may be distributed in the general population.

We estimated group parameter median and associated confidence intervals via bootstrapping, with the results shown in Table 14. The most important parameters are those that pertain to the storage of associative features— p_A and t_A —since these are the only source of evidence that can discriminate intact from rearranged pairs in these experiments. On average, roughly one third of probe

Table 15

Bootstrapped mean and 99% confidence intervals for Kendall's τ correlations between best-fitting associative recognition parameters from Doshier (1984a) and Wickelgren and Corbett (1977). Estimates were made from 1,000,000 bootstrap samples.

	θ	p_A	t_A	t_0	ρ
c_S	-0.03 (-0.58–0.60)	0.15 (-0.53–0.69)	0.13 (-0.39–0.64)	0.00 (-0.46–0.46)	0.37 (-0.16–0.72)
θ		-0.58 (-0.87– -0.19)	-0.20 (-0.65–0.35)	0.10 (-0.45–0.62)	-0.27 (-0.70–0.21)
p_A			-0.12 (-0.58–0.39)	-0.12 (-0.62–0.52)	0.52 (-0.04–0.91)
t_A				0.30 (-0.27–0.77)	-0.03 (-0.46–0.46)
t_0					0.07 (-0.50–0.77)

features are given over to the encoding of associative features, although there remains considerable inter-subject variability. Likewise, while there is still uncertainty regarding the time t_A at which associative features become available, it is on average 219 ms after stimulus onset, but might be as brief as 117 ms or as long as 411 ms.

Do any of these parameters trade-off with one another or represent other correlations? We estimated Kendall's τ ordinal correlations between these 6 parameters via bootstrapping, along with associated 99% confidence intervals, all of which are given in Table 15. The only parameters for which there is good evidence of a correlation are θ and p_A , specifically, as the proportion of associative features in the probe grows, response threshold θ becomes more liberal. Examining Figures 21, 22, and 23, we see that participants that show poor asymptotic discriminability—associated with lower p_A —tend to be biased toward giving negative responses—associated with a higher θ . Parameters related to sampling dynamics, particularly t_0 and t_A , are positively correlated as one might expect (since $t_A > t_0$), but this correlation is sufficiently weak to be indiscriminable from zero. There is no evidence for any tradeoff between encoding fidelity c_S and the proportion of associative features p_A , suggesting that these account for separable aspects of memory performance.

Discussion

In this section, we illustrated how our dynamic model for recognition of single items can be extended to the recognition of pairs. Based on the dynamic properties of associative recognition, we showed that its dynamics can be best explained in terms of when associative features become available to join the developing probe of memory, in contrast to theories that assume associative information can only be retrieved by a separate recall-like process. We propose that associative features represent a type of compound cue (cf. Doshier & Rosedale, 1989, 1997) that represents the interrelations between the features of the items being associated. Thus, the delay in the availability of associative features is due to the fact that participants must first accumulate some information about the items before these relational associative features can be formed. In other words, participants must first know something about what is being associated before they can form the association. This implies that the representation of an association is not independent of the items that comprise the association, consistent with an absence of interference between pairs of different item types or pairs of items encountered in different contexts (Aue, Criss, & Fischetti, 2012; Criss & Shiffrin, 2004b, 2005).

In support of the notion that associative features depend on item information, the presence of a pre-existing semantic relationship between items both enhances correct recognition of an intact pair and aids rejection of rearranged pairs that “break” the studied semantic relationship (Doshier,

1984b; Doshier & Rosedale, 1991). Such relationships in rearranged pairs have also been found to lead to a bias to erroneously call such pairs “intact”, at least early in processing (Doshier, 1984b; Doshier & Rosedale, 1991). Consistent with associative features being primarily relational, situations where relationships are easier to detect leads to greater encoding of associative features. This kind of “feedback loop” between the presence of a semantic relationship and the encoding of an episodic pair association implies that associative features represent an intermediate step between completely separate entities/events and those that have become “unitized”, i.e., viewed as a single whole (Czerwinski et al., 1992; Gauthier & Tarr, 1997; Shiffrin & Lightfoot, 1997): For previously unrelated items (e.g., foot-carrot), only a few associative features can be stored; for related items (foot-hand), more such features can be stored; and for unitized items (foot-ball), storage may consist solely of “associative” features. According to the theory we have outlined here, this transition should be reflected in the dynamics of retrieval: To the extent that associative information is episodic, it should be delayed relative to the onset of item features; to the extent that associative information is semantic, it should be encoded as an item feature that is available as soon as item features begin to be sampled. Functional neuroimaging supports this continuum, with perirhinal cortex being engaged when encoding novel or especially difficult associative information (foot-carrot) and earlier cortical regions encoding well-learned object representations (Cowell, Bussey, & Saksida, 2006; Haskins, Yonelinas, Quamme, & Ranganath, 2008; Staresina & Davachi, 2010; Cowell et al., 2010).

Across experiments, we estimated that associative features are likely to be available between 117 and 411 ms after stimulus onset (Table 14). ERP studies of associative recognition have found signatures discriminating intact from rearranged pairs in both late positive components—sometimes identified as occurring after 600 ms (D. I. Donaldson & Rugg, 1998) but sometimes as “early” as 400 ms (Jäger, Mecklinger, & Kipp, 2001)—and earlier negative components (after 250–300 ms; Hannula, Federmeier, & Cohen, 2006; Speer & Curran, 2007; Rhodes & Donaldson, 2008), consistent with the range of times at which we estimate associative features join the pool. Typically, such results have been interpreted to mean that associative recognition is partially dependent on both early “familiarity” processes and later “recollection” processes (e.g., Rugg & Curran, 2007), but our model suggests a more fruitful interpretation of this mixed literature in terms of differences between individuals (some of whom show exceptionally fast associative encoding) and random variation in the dynamics of feature sampling. Even if associative features always became available at the same time, they might enter the probe quickly on some trials or slowly on others, owing to noise in the feature sampling process. Our dynamic approach—especially its applicability to individual participants—offers a distinct opportunity to clear up the apparently muddy picture of associative retrieval dynamics painted by EEG studies.

General Discussion

We have introduced a dynamic approach to recognition memory based on several core features: that context serves as a baseline against which subsequent memory evidence is judged; that recognition decisions depend on tracking changes in familiarity over time; and that these dynamics are governed by the time at which features become available to sample into an evolving probe of memory. A quantitative model derived from this approach closely matches the behavior of both groups and individuals across a wide variety of paradigms and manipulations, enabling the model to be used to express and test different theories of retrieval. Our dynamic approach emphasizes the importance of how prior experience affects the episodic encoding of low frequency words (Section

1), that recognition uses the current context as a baseline to judge novelty and/or familiarity (Section 2), that recognition depends on tracking changes from an initial level of familiarity which can be influenced by primes (Section 3), and that different kinds of information join the memory probe at different times by virtue of how word events are represented (Section 4) and how novel associations are formed (Section 5). Our model represents a significant advance over most current accounts 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. In this General Discussion, we consider several ways in which a dynamic approach to recognition could be extended to spur further empirical and theoretical development.

Model extensions

Although our theoretical framework is intended to apply generally to any retrieval scenario, our quantitative modeling has focused primarily on two-choice recognition paradigms. Here, we consider how the model might be extended to other tasks or reformulated without altering the underlying framework of the theory.

Feature recovery and source judgments. In Section 4, we explained how source exclusion judgments arise from the sampling of modality-related features of an event that are separable from the semantic features that initially encode a word. Thus, the dynamics of such decisions are governed by the time at which these different types of features enter the *probe*. If, instead, dynamics depend on the time at which features can be recovered from memory, these dynamics should depend on the strength of the memory trace from which they are recovered. Although none of the studies in Sections 4 and 5 reported this result, Hintzman and Caulton (1997) found that dynamics were affected by trace strength when participants had to make source *judgments*. Unlike these other studies, participants know that the test words in source judgment tasks had been studied and have to decide whether they had been presented visually or auditorily. When an item was strengthened by repetition during study, not only did this improve asymptotic modality judgment performance, it also caused correct modality judgments to rise above chance more quickly (Figure 25).

We can model the recovery of source features from memory traces using the same dynamic mechanisms invoked in other recognition tasks. Features of the test item gradually join an evolving probe which is compared in parallel to all the traces stored in memory, as usual. At any given time, rather than tracking the average activation across memory traces, a single memory trace is sampled in proportion to its activation relative to that of other traces (e.g., Raaijmakers & Shiffrin, 1981; Diller et al., 2001; Malmberg, 2008):

$$\Pr(\text{Sample trace } i \text{ at time } t) = \frac{\left[\frac{\lambda_i(t)}{\sum_{j=1}^{N_{Act}(t)} \lambda_j(t)} \right]^\kappa}{\left[\frac{\lambda_i(t)}{\sum_{j=1}^{N_{Act}(t)} \lambda_j(t)} \right]^\kappa + \beta^\kappa},$$

where $N_{Act}(t)$ is the number of active traces (those with $\lambda_j(t) > 1$) at time t , β is a criterion level of relative activation that allows for no trace to be sampled at time t , and $\kappa > 0$ is a scaling factor. If a trace is sampled, its features are recovered. For simplicity, we assume that if the wrong trace has been sampled, participants detect the mismatch and say the test word was auditory with probability p_A (in a more complete model, participants would have a mechanism for restarting their sampling efforts); they do the same if no trace has been sampled by the time of the response signal. If, instead,

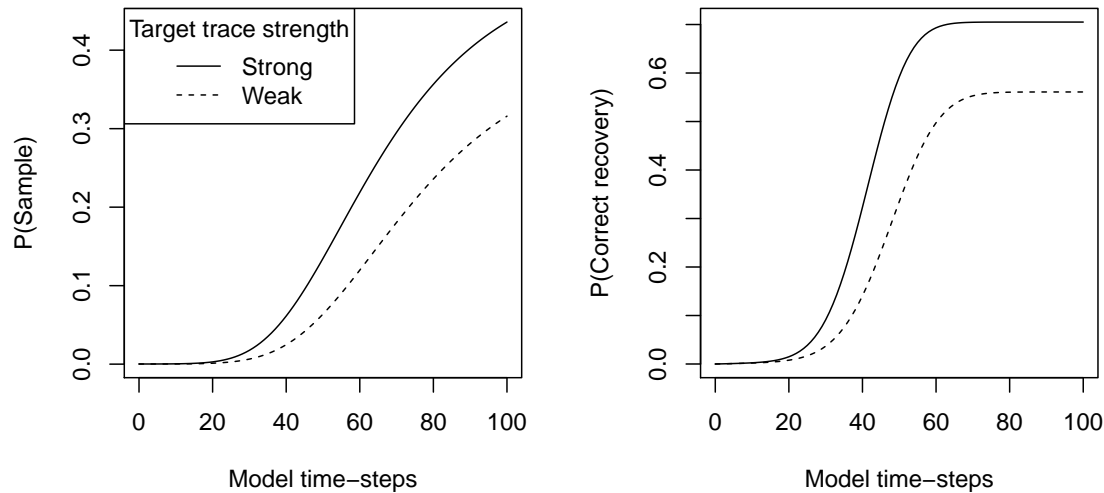


Figure 24. The dynamics of target trace recovery as a function of strength. The left panel shows the relative activation of the target trace as features join the memory probe over time, while the right panel shows the probability of having successfully recovered the target trace by a certain time.

the correct trace has been sampled by the time of the response signal, participants correctly respond with the modality in which the word had been presented. Thus, the dynamics of source judgments will depend on whether the correct trace has been sampled by the time of the response signal, which in turn depends on both the features in the probe (which are used to activate the memory traces) *and* the strength of the correct trace (which governs its relative degree of activation which grows faster for stronger traces; see Figure 24).

We fit this extension to the data from the third experiment reported by Hintzman and Caulton (1997); as shown in Figure 25, the dynamics of feature recovery provide an excellent quantitative account of these data. This is an example of how our dynamic approach might be extended to more elaborate recall-based tasks, including both cued and free recall. Obviously, much more work would be required to flesh out any such theory, but doing so would allow for a more complete unification of recognition and recall-like processes. Such an effort seems especially worthwhile given the prevalence of a familiarity-recollection distinction in modern neuroscience (e.g., D. I. Donaldson & Rugg, 1998; Curran, 2000; Ranganath et al., 2003; Rugg & Curran, 2007). As we suggest below, dynamic models like ours allow the possibility for tightening the link between behavioral and neural measurements.

Collapsing boundaries and competing accumulators. 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. 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, and certainly it is reasonable to expect that participants will not search memory indefinitely. Collapsing boundaries have been implicated in a variety of decision-making models (Balakrishnan & Macdonald, 2011; Busemeyer & Rapoport, 1988; Ditterich, 2006;

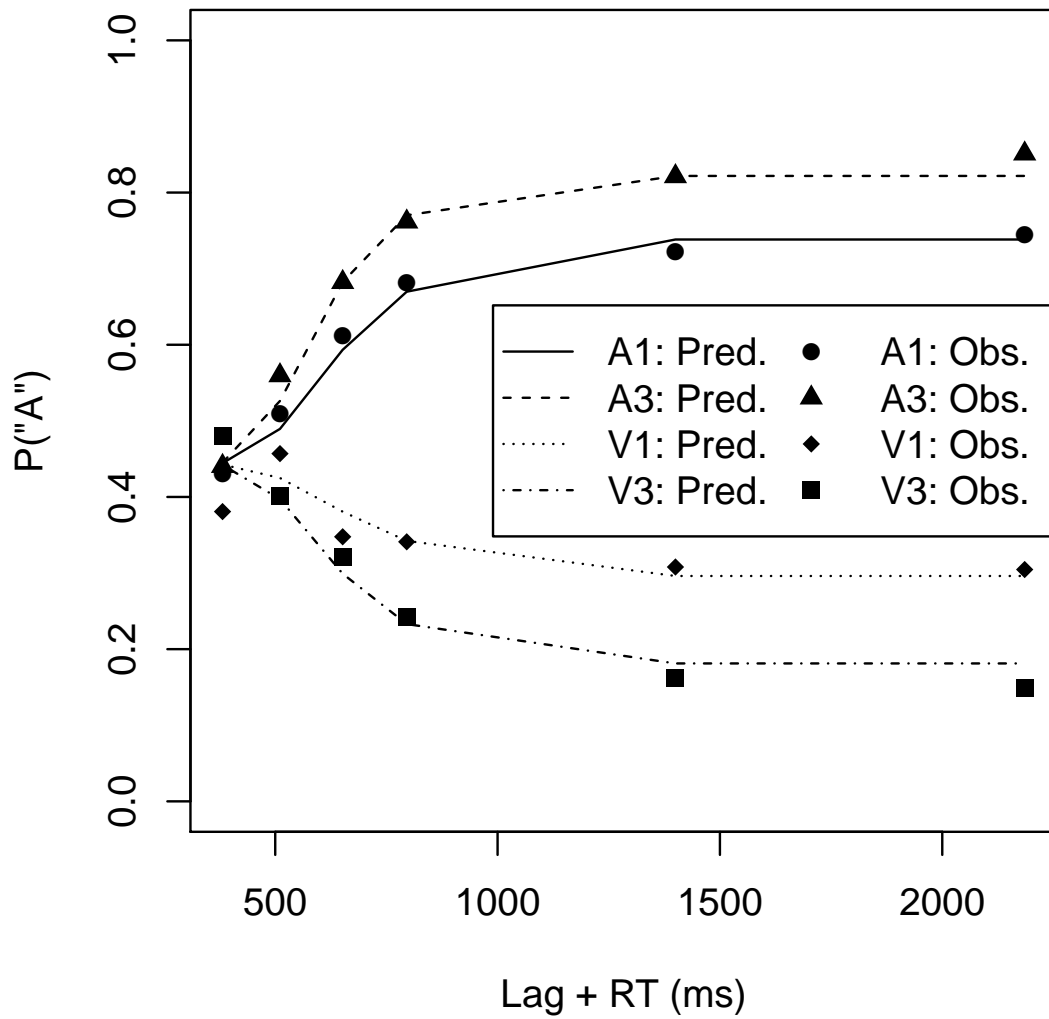


Figure 25. Probability of judging a word as having been presented auditorily in Hintzman and Caulton (1997, Exp. 3). Conditions are labeled by modality (A for auditory, V for visual) and by the number of repetitions at study (1 or 3 times). We fit separate encoding probabilities for each study condition, with the complete set of parameters given by $u_{A1} = 0.284$, $u_{A3} = 0.354$, $u_{V1} = 0.210$, $u_{V3} = 0.309$, $c_S = 0.985$, $p_A = 0.443$, $\beta = 0.005$, $\kappa = 8.197$, $t_0 = 44.019$, and $\rho = 3.094$.

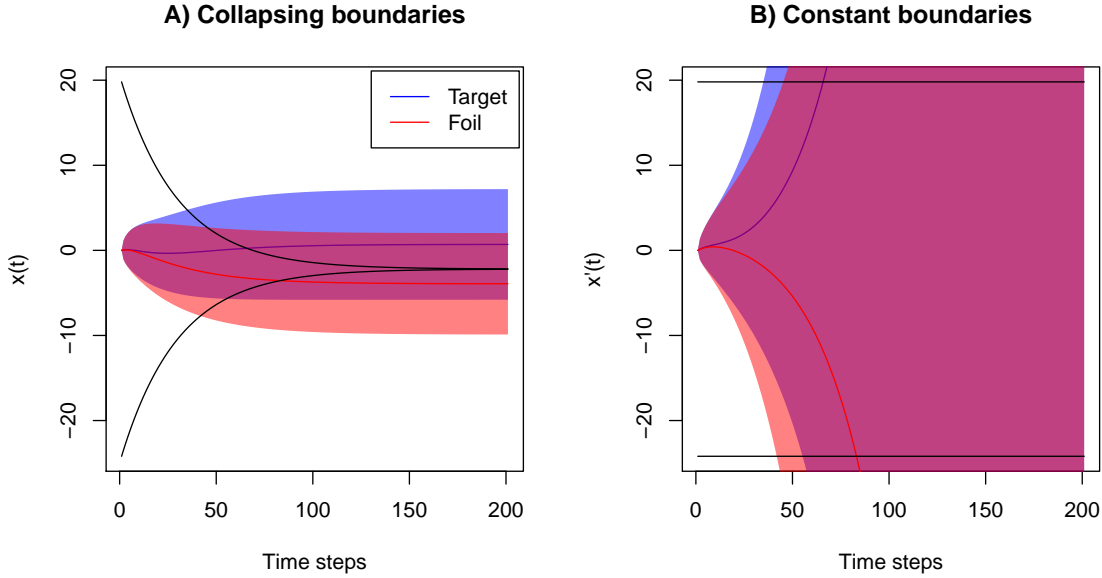


Figure 26. The model’s predicted distributions of familiarity over time for target and foil items, in relation to an upper (“old”) and lower (“new”) response boundary. Solid colored lines depict mean familiarity while colored regions show 95% confidence regions about the mean. A) In the standard formulation 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.

Frazier & Yu, 2008; Ratcliff & Frank, 2012; Zhang, Lee, Vandekerckhove, Maris, & Wagenmakers, 2014), including the recognition decision model of Hockley and Murdock (1987). We also note the analogy between collapsing bounds and a time-dependent “urgency signal” (Cisek, Puskas, & El-Murr, 2009).

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 26. 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 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. 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 (Busmeyer & Townsend, 1993; Usher & McClelland, 2001). While we believe that collapsing boundaries are conceptually transparent, easily interpreted, and well-motivated by the structure of our model, an excitatory-inhibitory accumulator formulation of our model need not alter its predictions and may be better suited for some purposes.

Confidence and remember-know judgments. In this paper, we have been concerned only with two-choice tasks, yet it is common to collect multiple responses per trial in the form of confidence ratings. Our approach toward modeling binary choices is equivalent to having two parallel accumulators representing the two choices that are perfectly anti-correlated, such that a step up in one accumulator (e.g., toward an “old” response) is matched by an equal-sized step down in the other (e.g., away from a “new” response). Because we assume boundaries collapse over time, early responses tend to be based on stronger evidence than later responses, so participants might assign confidence to their response on the basis of how long it took to reach a boundary. This hypothesis, however, implies a stronger correlation between response time and confidence than is typically observed (Ratcliff & Starns, 2009).

Confidence ratings could, instead, be driven by separate partially-correlated accumulators corresponding to each possible response, akin to the RTCON approach (Van Zandt, 2000; Usher & McClelland, 2001; Ratcliff & Starns, 2009, 2013). This method seems less applicable when giving continuous confidence ratings (Province & Rouder, 2012), although in such cases it is possible that participants assign accumulators to a small set of “anchor” values along the confidence scale. Alternatively, confidence judgments might arise from a two-stage process, where the first stage is a binary recognition decision of the kind we’ve addressed here, followed by a second stage in which evidence is further assessed for the purpose of giving a confidence rating (Pleskac & Busemeyer, 2010), potentially making use of collapsing boundaries as in the initial binary decision (Moran, Teodorescu, & Usher, 2015). This second stage might involve the sampling of a different set of features beyond those used to make the initial recognition judgment, thereby reducing the strong correlation between response time and evidence strength predicted by our binary choice model.

A different second-stage process might help align confidence ratings with another kind of multi-alternative recognition response: Remember-know judgments (Tulving, 1985; Reder et al., 2000). In this case, the second stage would be akin to our account of source memory judgments in that, during this stage, participants would directly assess the relative strengths of the activated traces in memory and attempt to recover features from studied traces. As described above, features can be recovered more quickly from a stronger trace, potentially helping to explain the puzzle of how “remember” judgments—supposedly based on a slow, effortful memory search—are often made *faster* than “know” judgments (Dewhurst & Conway, 1994; Dewhurst, Holmes, Brandt, & Dean, 2006; Rotello & Zeng, 2008). In addition, if feature recovery were also the basis for confidence ratings, this would help explain the strong correlation between remember-know and confidence judgments (W. Donaldson, 1996; Hirshman & Master, 1997; Dunn, 2004, 2008; Wixted & Mickes, 2010; Pratte & Rouder, 2012), since both kinds of responses would depend on the same evolving probe of memory activating the same memory traces and using those activations as the basis for feature recovery (cf. Gillund & Shiffrin, 1984). Although further exploration of this hypothesis is needed, we note that feature recovery dynamics can be dissociated from the dynamics of feature sampling, emphasizing the utility of dynamic measures to further understand the mechanisms behind confidence and remember-know judgments.

Alternative decision mechanisms. As noted in Appendix C, many of the parameters of our model can be mapped onto those of more general response time models, such as diffusion models (e.g., Ratcliff, 1978), with our model providing a more concise, theoretically motivated account of those parameters. A major drawback of these general response time models, as noted in the Introduction, is that they cannot easily allow for evidence to vary as a function of time, which is a core element of our dynamic approach. On the other hand, these general methods have been

extensively explored and there are efficient computational means for deriving their predictions and fitting them to data. Thus, in limited situations, it may be possible to use one of these general models and “work backwards” to determine what kinds of memory processes from our approach could yield the behavior of these general models. This effort could have practical benefits, in the same way that removing within-trial variability yields “ballistic” response time models that, while less realistic, still yield important insights (S. Brown & Heathcote, 2005, 2008). It could also have theoretical merit in that much work has gone toward investigating the neural implementation of diffusion-like models (Smith, 2010), but see our discussion below regarding prospects for a dynamic approach to the neuroscience of retrieval.

Alternative memory representations. Our model assumes that different events are stored as separate traces in memory. An alternative view is that events are stored in a single composite trace (e.g., Murdock, 1982; Metcalfe & Eich, 1982). Composite trace theories, as typically expressed, have difficulty accounting for the fact that strengthening memory for one item does not harm memory for other items in the same context (Ratcliff, Clark, & Shiffrin, 1990; Shiffrin, Ratcliff, & Clark, 1990). The separate trace view that we adopt allows for stronger memories to become more distinctive such that they do not interfere with weaker traces (Shiffrin & Steyvers, 1997; McClelland & Chappell, 1998; Criss, 2006). It is, however, possible to achieve the same effect using a composite representation if individual events are encoded in a sufficiently sparse manner (Kanerva, 1988), meaning our theory need not be restricted to the assumption of separate traces.

Likewise, although we assumed binary independent features for simplicity in our mathematical model, this is not entailed by the theory and we envision some situations in which this assumption should be revised. Different kinds of features may have distributions of feature values (e.g., Shiffrin & Steyvers, 1997; McClelland & Chappell, 1998) which could be derived from the environmental properties of the features used to encode events (Johns & Jones, 2010). When stimuli are well controlled, with features/dimensions that have been explicitly designed and/or discovered by procedures like multidimensional scaling, it seems reasonable to assume that features—and their associated sampling dynamics—correspond to these well-defined aspects of the stimuli (e.g., Nosofsky, 1992; Lamberts, Brockdorff, & Heit, 2002). Finally, it may be necessary to assume correlations between features. For example, if “bank” is encountered in a financial context, the semantic features that are encoded in the corresponding event trace will tend to be those pertaining to its meaning as a financial institution and *not* those related to its meaning as the side of a river. In this case, one might adopt a feature representation that allows for correlations (Mueller & Shiffrin, 2006) or represent correlations as their own kind of higher-order feature (e.g., A, B, and AB).

Other aspects of memory

We have demonstrated how our model of recognition dynamics can be extended to the different memory task of list discrimination, 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. The temporal context model (Howard & Kahana, 2002), for example, has 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 et al., 2008; Polyn et al., 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. Response times have also been used to motivate a more elaborate model of temporal context

in which recency judgments involve a kind of reverse mental time-travel in which participants search backwards through time from the present (Howard et al., 2015), similar to how we model contextual reinstatement in list discrimination. The dynamics of 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” intentionally stored by the participant, or might be reflect the different temporal context in which the various repetitions occurred (e.g., Howard et al., 2015). Given the similar dynamics between frequency judgments and recognition, whatever new features are introduced might act like content features that are available to be sampled throughout a trial. 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 may not do so during the test phase. 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 judgments are not qualitatively different, although more empirical work would be required to put any such theory on firm footing.

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 accounts of this paradigm as well (Ratcliff, 1978; Hockley & Murdock, 1987; Nosofsky, Little, Donkin, & Fific, 2011; Donkin & Nosofsky, 2012b; Nosofsky, Cox, Cao, & Shiffrin, 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 (e.g., Donkin & Nosofsky, 2012a; Nosofsky, Cox, et al., 2014). Decay could result from the loss of content features in the trace (decreasing u), most likely due to subsequent items interfering with the transfer of features from short-term to long-term storage (Atkinson & Shiffrin, 1968). 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” (Howard et al., 2015), but would need to be reconciled with the fact that context shifts within a list tend to be quite minor (Klein et al., 2007).

Knowledge access. The first stage in our model involves using the physical features of the test item in conjunction with context to access knowledge about the item, which then con-

tributes to the pool of features that are available to be sampled into the memory probe. At present, we do not explicitly model this process, however, we have found that this stage is critical to explaining recognition performance for opposite-plurality foils, in which plurality is available later than other semantic/orthographic features, and following masked primes, where an identical prime speeds knowledge access. Existing accounts of knowledge access make use of feature-sampling mechanisms that are closely related to those we propose for event recognition (Wagenmakers et al., 2004). It is, therefore, possible that essentially the same recognition process operates in a continuous flow, albeit on two different kinds of memory traces: a first process that activates and accesses knowledge traces followed by a second similar process—the one we describe in this paper—that activates event traces. We expect future theoretical and experimental work to further explore the interface between knowledge access and event recognition (for progress along these lines, see Nelson & Shiffrin, 2013).

Encoding. We have focused here on modeling the processes involved in memory retrieval, but we believe that a dynamic approach to memory storage will also prove fruitful. By developing explicit models of the dynamics of the study event—both within the study trial and in the ongoing study context—similarities between encoding and retrieval might be revealed. 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, which presumably involves retrieval of the length of each path and its associated reward, resembles a temporally-compressed version of the hippocampal activity that occurs when the mice are learning the maze (Jadhav, Kemere, German, & Frank, 2012).

Extended events. Our model describes how familiarity evolves over relatively short spans of time, on the order of a few seconds at most. This timescale is entailed by the kinds of stimuli used in the experiments we have considered, namely, visually presented words or images. Such stimuli, once they appear, are static and are not extended in time. Most events take place over longer timescales and have their own internal 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 expect that the processes we have described for recognition over short timescales should continue to apply to events of longer duration. 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 form a memory probe in exactly the way our model describes. The familiarity signal resulting from such a probe would evolve over time, just as we have described, and could then serve as the basis for recognition or other memory decisions.

Categorization. In Section 4, we distinguished between features that were integral to the representation of a word, its semantics and base form, with those that were more peripheral, like modality and plurality. This is analogous to the distinction often made in categorization between “separable” features, which can be processed independently, and “integral” features, which are treated holistically (see, e.g., Lockhead, 1966; Garner & Morton, 1969; Garner, 1970; Treisman & Gelade, 1980; Ashby & Townsend, 1986; Nosofsky, 1986). Although it is possible for separable features to be processed at different times (Fific, Little, & Nosofsky, 2010; Little, Nosofsky, & Denton, 2011), integral features are processed together in parallel (Little, Nosofsky, Donkin, &

Denton, 2013), consistent with our account that a word's integral features are available together whereas separable features like plurality can be delayed. Thus, our dynamic approach to recognition is conceptually linked to categorization via a focus on the nature of the information in memory—whether features of a memory trace are separable or integral—and on the dynamics of the processes that retrieve this information.

As an even tighter link between recognition memory and 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). Conversely, it is possible to view categorization as a kind of recognition, by analogy with our account of list discrimination in Section 2. Consider binary categorization, where participants study items from either category A or category B. Each studied exemplar results in a memory trace where the content features include both item 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 reflect 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 particular account of “categorization as recognition” only applies in the binary case but highlights one of the many ways memory could be employed in an ostensibly different task (Brockdorff & Lamberts, 2000; Nosofsky et al., 2011).

Prospects for neuroscience

Above, particularly in Sections 4 and 5, we noted some ways in which our model relates to measures of the internal neural state of participants while performing recognition tasks. While fMRI allows researchers to track dynamics to some extent (e.g. Criss, Wheeler, & McClelland, 2012), only measures like EEG and MEG do so at a time-scale similar to that of our model, i.e., small fractions of a second. In accord with our model, such measures have found support for the notions that source/modality information (Staresina, Fell, Lam, Axmacher, & Henson, 2012; Staresina, Fell, Dunn, Axmacher, & Henson, 2013), plurality information (Curran, 2000), and associative information (Staresina & Davachi, 2008; Kumaran & Maguire, 2009) are all separable from other (e.g., semantic) item information. Much of this data has been used to support a distinction between a familiarity process and a recall-like recollection process (D. I. Donaldson & Rugg, 1998; Rugg & Curran, 2007), although the early and late ERP signals thought to index these processes do not consistently map onto retrieval of specific kinds of information (e.g., Greve, van Rossum, & Donaldson, 2007; Speer & Curran, 2007; Addante, Ranganath, & Yonelinas, 2012). By focusing on the information conveyed by specific dynamic processes, our approach suggests a different distinction: because the dynamics of recognition are governed by how the memory *probe* evolves rather than by how the activations of memory *traces* evolve, these neural signals reflect the time at which informa-

tion enters the probe, either because of how knowledge is accessed, task-specific strategies, or the dependence of associative information on item information. Although future research will have to determine the relative utility of these distinctions, we emphasize that such determinations are only possible if memory theories specify the dynamics of retrieval at a sufficiently fine timescale.

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. 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, such as diffusion or ballistic accumulator models which define a participant's state in terms of a whole-trial "drift rate" (Bogacz, Wagenmakers, Forstmann, & Nieuwenhuis, 2010; Turner et al., 2013). Dynamic models like ours, on the other hand, make 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. On the one hand, one could compare the correlation matrices of whole-brain activity at different time points within a trial to that predicted by our (or another) dynamic model, thereby assessing the degree to which the model accounts for internal state changes during recognition. On the other hand, one could identify particular brain regions that demonstrate the pattern of dynamic activity predicted by the model and, to the degree that those regions were sensible (for example, one would expect heavy involvement of the hippocampus in event recognition), this would lend support to the model. We believe a dynamic approach to memory theory is an advance not only at the behavioral level, but represents a distinct opportunity to link retrieval processes with underlying neural activity.

References

- Addante, R. J., Ranganath, C., & Yonelinas, A. P. (2012). Examining ERP correlates of recognition memory: Evidence of accurate source recognition without recollection. *NeuroImage*, 62, 439–450.
- 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). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, 22(3), 261–295.
- Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, 2(6), 396–408.
- Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, 93(2), 154–179.
- 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.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89–195). New York: Academic Press.
- Aue, W. R., Criss, A. H., & Fischetti, N. W. (2012). Associative information in memory: Evidence from cued recall. *Journal of Memory and Language*, 66, 109–122.
- 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.

- 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.
- Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current Directions in Psychological Science*, 11(5), 164–169.
- Brainerd, C. J., Reyna, V. F., & Mojardin, A. H. (1999). Conjoint recognition. *Psychological Review*, 106(1), 160–179.
- 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, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, 107(1), 127–181.
- 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.
- 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.
- Burgess, C., & Livesay, K. (1998). The effect of corpus size in predicting reaction time in a basic word recognition task: Moving on from Kucera and Francis. *Behavior Research Methods, Instruments, & Computers*, 30(2), 272–277.
- 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.
- 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.
- 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.
- Cowell, R. A. (2012). Computational models of perirhinal cortex function. *Hippocampus*, 22, 1952–1964.
- Cowell, R. A., Bussey, T. J., & Saksida, L. M. (2006). Why does brain damage impair memory? a connectionist model of object recognition memory in perirhinal cortex. *Journal of Neuroscience*, 26(47), 12186–12197.
- Cowell, R. A., Bussey, T. J., & Saksida, L. M. (2010). Components of recognition memory: Dissociable cognitive processes or just differences in representational complexity? *Hippocampus*, 20, 1245–1262.
- 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. (2006). The consequences of differentiation in episodic memory: Similarity and the strength based mirror effect. *Journal of Memory and Language*, 55, 461–478.
- 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. (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.
- Czerwinski, M., Lightfoot, N., & Shiffrin, R. M. (1992). Automatization and training in visual search. *American Journal of Psychology*, 105(2), 271–315.
- Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of experimental psychology*, 58(1), 17.
- 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.
- Dewhurst, S. A., Holmes, S. J., Brandt, K. R., & Dean, G. M. (2006). Measuring the speed of the conscious components of recognition memory: Remembering is faster than knowing. *Consciousness and Cognition*, 15(1), 147–162.
- 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.
- 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., & 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.
- Doshier, 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.
- Doshier, B. A. (1984b). Discriminating preexperimental (semantic) from learned (episodic) associations: A speed-accuracy study. *Cognitive Psychology*, 16, 519–555.
- Doshier, B. A., McElree, B., Hood, R. M., & Rosedale, G. (1989). Retrieval dynamics of priming in recognition memory: Bias and discrimination analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(5), 868–886.
- Doshier, 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.
- Doshier, 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.
- Doshier, B. A., & Rosedale, G. (1997). Configural processing in memory retrieval: Multiple cues and ensemble representations. *Cognitive Psychology*, 33, 209–265.
- 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.

- 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.
- Fific, M., Little, D. R., & Nosofsky, R. M. (2010). Logical-rule models of classification response times: A synthesis of mental-architecture, random-walk, and decision-bound approaches. *Psychological Review*, 117(2), 309–348.
- 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.
- Gallo, D. A., Roediger, H. L., & McDermott, K. B. (2001). Associative false recognition occurs without strategic criterion shifts. *Psychonomic Bulletin & Review*, 8(3), 579–586.
- Garner, W. R. (1970). The stimulus in information processing. *American Psychologist*, 25(4), 350–358.
- Garner, W. R., & Morton, J. (1969). Perceptual independence: Definitions, models, and experimental paradigms. *Psychological Bulletin*, 72(4), 233–259.
- Gauthier, I., & Tarr, M. J. (1997). Becoming a “greeble” expert: Exploring mechanisms for face perception. *Vision Research*, 37(12), 1673–1682.
- Genz, A. (2004). Numerical computation of rectangular bivariate and trivariate normal and t probabilities. *Statistics and Computing*, 14, 251–260.
- Gibson, E. J. (1971). Perceptual learning and the theory of word perception. *Cognitive Psychology*, 2, 351–368.
- 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.
- 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.
- Hannula, D. E., Federmeier, K. D., & Cohen, N. J. (2006). Event-related potential signatures of relational memory. *Journal of Cognitive Neuroscience*, 18(11), 1863–1878.
- Haskins, A. L., Yonelinas, A. P., Quamme, J. R., & Ranganath, C. (2008). Perirhinal cortex supports encoding and familiarity-based recognition of novel associations. *Neuron*, 59, 554–560.
- 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. (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. (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.
- 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.
- 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., Aue, W. R., & Criss, A. H. (2015). A distributed representation of internal time. *Psychological Review*, 122(1), 24–53.
- 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.
- Huber, D. E., Curran, T., Clark, T. F., & Winkielman, P. (2008). Effects of repetition priming on recognition memory: Testing a perceptual fluency-disfluency model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(6), 1305–1324.
- Huber, D. E., & O'Reilly, R. C. (2003). Persistence and accommodation in short-term priming and other perceptual paradigms: Temporal segregation through synaptic depression. *Cognitive Science*, 27, 403–430.
- Huber, D. E., Shiffrin, R. M., Lyle, K. B., & Ruys, K. I. (2001). Perception and preference in short-term word priming. *Psychological Review*, 108(1), 149–182.
- 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.
- 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.
- Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, 110(3), 306–340.
- Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General*, 118(2), 126–135.
- Jadhav, S. P., Kemere, C., German, P. W., & Frank, L. M. (2012). Awake hippocampal sharp-wave ripples support spatial memory. *Science*, 336, 1454–1458.
- Jäger, T., Mecklinger, A., & Kipp, K. H. (2001). Intra- and inter-item associations doubly dissociate the electrophysiological correlates of familiarity and recollection. *Neuron*, 52, 535–545.

- 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.
- Johnson, N. F. (1975). On the function of letters in word identification: Some data and a preliminary model. *Journal of Verbal Learning and Verbal Behavior*, 14, 17–29.
- 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.
- Jou, J., & Cortes, H. M. (2012). Can people strategically control the encoding and retrieval of some morphologic and typographic details of words? *Consciousness and Cognition*, 21, 1280–1297.
- Kahana, M. J. (2012). *Foundations of human memory*. New York, NY: Oxford University Press.
- Kanerva, P. (1988). *Sparse distributed memory*. Cambridge, MA: MIT Press.
- Kent, C., & Lamberts, K. (2008). The encoding–retrieval relationship: Retrieval as mental simulation. *Trends in Cognitive Sciences*, 12(3), 92–98.
- 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.
- Kruschke, J. K. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd ed.). London: Academic Press.
- Kumaran, D., & Maguire, E. A. (2009). Novelty signals: a window into hippocampal information processing. *Trends in Cognitive Sciences*, 13(2), 47–54.
- 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.
- LaBerge, D. (1962). A recruitment theory of simple behavior. *Psychometrika*, 27(4), 375–396.
- LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information in reading. *Cognitive Psychology*, 6, 293–323.
- 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.
- 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.
- Little, D. R., Nosofsky, R. M., & Denton, S. E. (2011). Response-time tests of logical-rule models of categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(1), 1–27.
- Little, D. R., Nosofsky, R. M., Donkin, C., & Denton, S. E. (2013). Logical rules and the classification of integral-dimension stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(3), 801–820.
- Liu, C. C., & Smith, P. L. (2009). Comparing time–accuracy curves: Beyond goodness-of-fit measures. *Psychonomic Bulletin & Review*, 16(1), 190–203.
- Lockhead, G. R. (1966). Effects of dimensional redundancy on visual discrimination. *Journal of Experimental Psychology*, 72(1), 95–104.
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, 95(4), 492–527.
- 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.
- Maddox, W. T., & Estes, W. K. (1997). Direct and indirect stimulus-frequency effects in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(3), 539–559.
- 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. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, 87(3), 252–271.
- 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., & 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 complementary 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.
- McKenzie, S., Frank, A. J., Kinsky, N. R., Porter, B., Rivière, P. D., & Eichenbaum, H. (2014). Hippocampal representation of related and opposing memories develop within distinct, hierarchically organized neural schemas. *Neuron*, 83, 202–215.
- Metcalfe, 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.
- Moran, R., Teodorescu, A. R., & Usher, M. (2015). Post choice information integration as a causal determinant of confidence: Novel data and a computational account. *Cognitive Psychology*, 78, 99–147.
- 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.
- Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, 89(3), 609–626.
- Murnane, K., & Shiffrin, R. M. (1991). Interference and the representation of events in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(5), 855–874.
- 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. (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.
- Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Science*, 12(6), 237–241.
- Pike, R. (1973). Response latency models for signal detection. *Psychological Review*, 80(1), 53–68.
- Pleskac, T. J., & Busmeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review*, 117(3), 864–901.
- Plummer, M. (2013). *JAGS: Just another Gibbs sampler*. Retrieved from <http://mcmc-jags.sourceforge.net/>
- 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.
- Pratte, M. S., & Rouder, J. N. (2012). Assessing the dissociability of recollection and familiarity in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(6), 1591–1607.
- 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 accumulation 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., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review*, 95(3), 385–408.
- 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., 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.
- Rhodes, S. M., & Donaldson, D. I. (2008). Electrophysiological evidence for the effect of interactive imagery on episodic memory: Encouraging familiarity for non-unitized stimuli during associative recognition. *NeuroImage*, 39, 873–884.
- 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.
- 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.
- Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences*, 11(6), 251–257.
- 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.
- Sereno, S. C., & Rayner, K. (2003). Measuring word recognition in reading: Eye movements and event-related potentials. *Trends in Cognitive Sciences*, 7(11), 489–493.
- Shiffrin, R. M., Huber, D. E., & Marinelli, K. (1995). Effects of category length and strength on familiarity in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(2), 267–287.
- Shiffrin, R. M., & Lightfoot, N. (1997). Perceptual learning of alphanumeric-like characters. In R. L. Goldstone, P. G. Schyns, & D. L. Medin (Eds.), *The psychology of learning and motivation* (Vol. 36, pp. 83–126). San Diego, CA: Academic Press.
- 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.

- 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., & 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.
- Staresina, B. P., & Davachi, L. (2008). Selective and shared contributions of the hippocampus and perirhinal cortex to episodic item and associative encoding. *Journal of Cognitive Neuroscience*, 20(8), 1478–1489.
- Staresina, B. P., & Davachi, L. (2010). Object unitization and associative memory formation are supported by distinct brain regions. *Journal of Neuroscience*, 30(29), 9890–9897.
- Staresina, B. P., Fell, J., Dunn, J. C., Axmacher, N., & Henson, R. N. (2013). Using state-trace analysis to dissociate the functions of the human hippocampus and perirhinal cortex in recognition memory. *Proceedings of the National Academy of Sciences*, 110(8), 3119–3124.
- Staresina, B. P., Fell, J., Lam, A. T. A. D., Axmacher, N., & Henson, R. N. (2012). Memory signals are temporally dissociated in and across human hippocampus and perirhinal cortex. *Nature Neuroscience*, 15(8), 1167–1173.
- Starns, J. J., Lane, S. M., Alonzo, J. D., & Roussel, C. C. (2007). Metamnemonic control over the discriminability of memory evidence: A signal detection analysis of warning effects in the associative list paradigm. *Journal of Memory and Language*, 56, 592–607.
- 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.
- Sternberg, S. (1966). High speed scanning in human memory. *Science*, 153, 652–654.
- 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.
- Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. New York: Cambridge University Press.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12, 97–136.
- 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.
- Underwood, B. J. (1969). Attributes of memory. *Psychological Review*, 76(6), 559–573.
- 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.

- 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.
- Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, 114(1), 152–176.
- Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review*, 117(4), 1025–1054.
- Woollams, A. M., Taylor, J. R., Karayanidis, F., & Henson, R. N. (2008). Event-related potentials associated with masked priming of test cues reveal multiple potential contributions to recognition memory. *Journal of Cognitive Neuroscience*, 20(6), 1114–1129.
- 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. (1997). Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Memory & Cognition*, 25(6), 747–763.
- 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.
- Zacks, J. M., & Tversky, B. (2001). Event structure in perception and cognition. *Psychological Bulletin*, 127(1), 3–21.
- 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.

Appendix A

Continuous approximation

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.

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):

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

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

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:

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

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 normal approximation has mean and covariance matrix

$$\mu(t) = N_C \begin{bmatrix} p_M(t) \\ p_N(t) \\ p_P(t) \\ p_T(t) \\ p_O(t) \end{bmatrix}$$

$$\Sigma(t) = N_C \begin{bmatrix} p_M(t)(1-p_M(t)) & -p_N(t)p_M(t) & -p_P(t)p_M(t) & -p_T(t)p_M(t) & -p_O(t)p_M(t) \\ -p_M(t)p_N(t) & p_N(t)(1-p_N(t)) & -p_P(t)p_N(t) & -p_T(t)p_N(t) & -p_O(t)p_N(t) \\ -p_M(t)p_P(t) & -p_N(t)p_P(t) & p_P(t)(1-p_P(t)) & -p_T(t)p_P(t) & -p_O(t)p_P(t) \\ -p_M(t)p_T(t) & -p_N(t)p_T(t) & -p_P(t)p_T(t) & p_T(t)(1-p_T(t)) & -p_O(t)p_T(t) \\ -p_M(t)p_O(t) & -p_N(t)p_O(t) & -p_P(t)p_O(t) & -p_T(t)p_O(t) & p_O(t)(1-p_O(t)) \end{bmatrix}$$

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

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

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

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.

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:

$$F_i(x) = \Phi\left(\frac{\log x - \mu_i}{\sigma_i}\right)$$

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

$$F(x) \lesssim \prod_{i=1}^N \Phi\left(\frac{\log x - \mu_i}{\sigma_i}\right) \quad (14)$$

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 14: 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.

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:

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

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:

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

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:

$$\begin{aligned} \mu_\phi(t) &= \mu_Z - \log S \\ \sigma_\phi(t) &= \frac{1}{2} (x_u - x_l) \end{aligned}$$

where x_u and x_l are defined in the previous section.

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

$$\begin{aligned} \mu_d &= D\mu_\phi \\ \sigma_d^2 &= \text{diag}(D^T \Omega D) \end{aligned}$$

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 an normal, independent of all other steps.

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 $x(t-1) < B_O(t-1)$:

$$p_O(t) = \Pr[x(t) > B_O(t) \wedge x(t-1) < B_O(t-1)]. \quad (15)$$

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:

$$\begin{aligned} p_O(t) &= f_O(t) + r_{OO}(t) + r_{NO}(t) \\ f_O(t) &= p_O(t) - r_{OO}(t) - r_{NO}(t). \end{aligned} \quad (16)$$

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

$$r_{OO}(t) = \sum_{\tau=0}^{t-1} f_O(\tau) \Pr[x(t) > B_O(t) | x(\tau) > B_O(\tau)]. \quad (17)$$

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

$$r_{NO}(t) = \sum_{\tau=0}^{t-1} f_N(\tau) \Pr[x(t) > B_O(t) | x(\tau) < B_N(\tau)]. \quad (18)$$

Given the time-varying boundaries $B_O(t)$ and $B_N(t)$ and the mean μ_x and variance Σ_x , the results of Equations 15, 17, and 18 can all be computed and used to solve for the first-passage time density (Equation 16). 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 x reaching either boundary, allowing a decision to be made.

Finally, we convert model “time-steps” to real time via a linear transformation via Equation 12, 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.

Appendix B

Table B1

Ranges from which model parameters were sampled to generate simulated data for parameter recovery.

Parameter	Lower bound	Upper bound
u	0	1
c_S	0.5	1
A_0	5	400
b	0	1
ρ	0.001	0.030
\bar{T}_R	0	0.5
σ_R	0	0.5

Parameter recovery

In many experimental situations, each participant contributes a relatively small number of observations to each condition, which potentially contributes to considerable variability in estimated model parameters. In this section, we investigate whether the parameters of our proposed model are sufficiently constrained by data such that, with increasing number of observations, the true generating parameters can be recovered (and the variability observed in the previous section might be mitigated). To that end, we simulate a limited number of trials from various randomly sampled sets of model parameter values and then find the model parameters that assign the highest likelihood to the simulated data. By comparing the true parameter values that generated the data to those found by maximum likelihood fitting, we can assess the degree to which the data constrain model parameters.

Each simulated dataset consists of an equal number of target and foil trials following study of a list of 40 items. Each parameter used to generate simulated data was sampled independently from a uniform distribution with ranges given in Table B1, under the constraint that the predicted difference in hit rate and false alarm rates was greater than 0.2 and that the predicted hit rate was greater than 0.4. These constraints ensure that the simulated data are within the range of performance typically observed in recognition memory studies. The parameters we allowed to vary were the seven parameters used to fit individual performance in the previous section (u , c_S , A_0 , b , \bar{T}_R , σ_R , and ρ).

We simulated 2000 datasets, 1000 of which contained 30 target trials and 30 foil trials and 1000 of which contained 300 target trials and 300 foil trials. Thus, the 30-trial datasets matched the design of the experiment in the previous section. As shown in the upper left region of Table B2, the accuracy constraints placed on the randomly-sampled true parameters do not induce many strong correlations between parameters with two exceptions: There are negative correlations between storage probability u and encoding fidelity c_S and between boundary separation A_0 and bias b . The negative correlation between u and c_S is a result of excluding parameter sets in which *both* u and c_S are too low to produce sufficient accuracy; similarly, the negative correlation between A_0 and b is a result of excluding parameter sets where there is either a strong “new” bias (low b , lower hit rate) or small boundary separation (low A_0 , lower discriminability). The upper right region of Table B2 indicates that parameter estimates are moderately to strongly correlated with the true generating parameters, with bias (b) and time-per-sample (ρ) most strongly correlated and mean residual time (\bar{T}_R) and feature storage probability (u) less strongly correlated.

The correlations between true and estimated parameters are noticeably stronger when 300

Table B2

Kendall's τ correlations between true model parameters and estimated model parameters across 1000 simulated datasets comprised of 30 target and 30 foil trials.

		True value						Estimated value						
		c_S	A_0	b	ρ	\bar{T}_R	σ_R	u	c_S	A_0	b	ρ	\bar{T}_R	σ_R
True value	u	-0.09	0.03	0.01	-0.01	-0.01	-0.01	0.42	-0.02	0.01	0.02	0.01	-0.07	-0.02
	c_S		0.08	-0.03	-0.03	-0.02	0	-0.08	0.61	0.07	-0.03	-0.01	-0.06	0.02
	A_0			-0.2	-0.01	-0.01	-0.02	0.01	0.06	0.64	-0.13	-0.02	0.01	0.03
	b				0	0.02	0.01	-0.04	-0.01	-0.16	0.78	0.03	-0.02	-0.01
	ρ					0.03	-0.02	0	-0.04	-0.03	0.02	0.72	0.11	0
	\bar{T}_R						0.01	-0.02	0	-0.01	0	0.05	0.45	-0.05
	σ_R							-0.02	0	0.01	0	-0.03	-0.01	0.55
Estimated	u								-0.24	-0.02	0	0	0.02	0
	c_S									0.1	-0.04	-0.02	-0.02	0.01
	A_0										-0.15	-0.15	0.01	0.12
	b											0.06	-0.04	-0.03
	ρ												0.01	-0.09
	\bar{T}_R													0.1

Table B3

Kendall's τ correlations between true model parameters and estimated model parameters across 1000 simulated datasets comprised of 300 target and 300 foil trials.

		True value						Estimated value						
		c_S	A_0	b	ρ	\bar{T}_R	σ_R	u	c_S	A_0	b	ρ	\bar{T}_R	σ_R
True value	u	-0.12	0.03	0.02	0.02	0	0.02	0.69	-0.1	0.04	0.03	0.02	-0.01	0.03
	c_S		0.09	-0.05	-0.01	0.02	-0.01	-0.1	0.81	0.1	-0.05	-0.02	0.01	-0.01
	A_0			-0.24	-0.04	-0.01	-0.02	0.04	0.08	0.86	-0.23	-0.03	0	-0.02
	b				0	0.01	0.01	0.02	-0.05	-0.23	0.92	0	0.01	0.02
	ρ					0.02	-0.02	0.01	-0.02	-0.02	0	0.89	0.04	-0.02
	\bar{T}_R						0	0.01	0.03	0	0.01	0.02	0.73	-0.01
	σ_R							0.01	-0.02	-0.04	0.01	-0.01	-0.02	0.77
Estimated	u								-0.16	0.05	0.02	0.02	0.01	0.03
	c_S									0.1	-0.05	-0.02	0.02	-0.01
	A_0										-0.23	-0.03	-0.01	-0.01
	b											0	0.02	0.02
	ρ												0.03	-0.02
	\bar{T}_R													0.01

target and foil trials are simulated, as shown in Table B3, although storage probability (u) remains the least correlated. Examining the lower right regions of Tables B2 and B3, we see that trade-offs between parameter estimates are generally weak. As mentioned above, the accuracy constraints placed on the simulated data introduce negative correlations between u and c_S and between boundary separation and bias, and both of these correlations are recovered in both 30-trial and 300-trial fits, indicating that these apparent trade-offs are a function of the true parameter values, rather than of uncertainty in parameter estimation.

Given that parameter estimates are not perfectly correlated with the true generating parameters in either the 30- or 300-trial fits, are the parameter estimates systematically biased? We computed the relative error in parameter estimates by subtracting the true values from the estimated values for each dataset and dividing by the corresponding range from which the parameter had been sampled (Table B1). The median of the resulting standardized estimation errors for each parameter

Table B4

Median standardized error in parameter estimates, where errors are computed by subtracting the true parameter value from the estimated parameter value for each dataset, then dividing by the corresponding parameter ranges given in Table B1.

Parameter	Number of simulated trials	
	30	300
u	0.019	-0.001
c_S	0.011	0.002
A_0	0.011	0.001
b	-0.005	0.000
ρ	-0.010	0.001
\bar{T}_R	0.026	0.002
σ_R	0.001	0.002

are given in Table B4. When only 30 trials are simulated, there is a very slight tendency (on the order of 1–3 percent of the range of the parameter) to overestimate encoding probability (u), encoding fidelity (c_S), boundary separation (A_0), and mean residual time (\bar{T}_R), while rate (ρ) goes underestimated. However, any systematic estimation bias all but disappears when estimating parameters from 300 simulated trials.

In general, model parameters can be estimated with moderate precision when only 30 target and 30 foil trials are observed, albeit with a tendency to overestimate initial boundary separation and residual time variability. With a greater number of trials available, estimation is considerably more precise and unbiased. Importantly, even with only 30 trials available, there are not strong negative correlations—trade-offs—among estimated parameters, at least relative to the correlations observed in the individual fits above. One consequence is that the correlations between parameters in the above fits likely represent genuine variation within the population. A second and more critical consequence is that the lack of trade-offs indicates that model parameters can be identified as reflecting independent processing components within the model, allowing for strong theoretical inferences to be drawn from fits to even relatively small amounts of data.

Appendix C

Comparison with diffusion model parameters

Although a good quantitative fit to the data reported by Starns et al. (2012), as shown in Section 1, is important, we note that in contrast to our model’s 31 free parameters (in model 5), the diffusion model reported by Starns et al. (2012) required 68 free parameters. The free parameters of the diffusion model included 16 mean drift rates and 16 drift distribution variances to account for each of the 16 combinations of word frequency (high/low), speed/accuracy emphasis, and number of study repetitions (0, 1, 2, or 4). Because the diffusion model is agnostic with respect to the source of the memory evidence in each of these cases, it cannot on its own provide an explanation for these differences between conditions. Our model, however, embodies particular mechanisms of retrieval and decision dynamics, so to the extent that the parameters of the diffusion model can be identified with those of our model, our model provides additional explanatory power.

In our model, memory evidence (“familiarity”; $x(t)$) is non-stationary, so to compare the

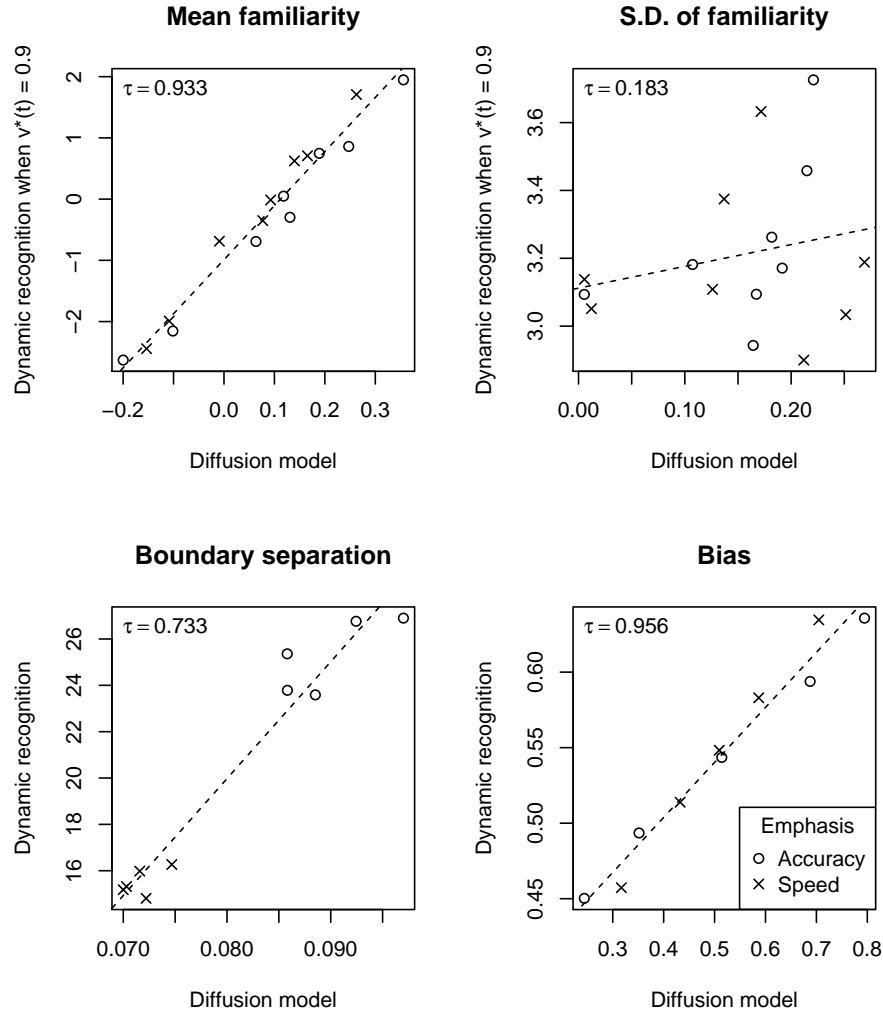


Figure C1. Comparisons between our dynamic recognition model's best-fitting parameters and best-fitting diffusion model parameters to the data from Starns et al. (2012). Dashed lines in each panel depict the least-squares regression line, while Kendall's τ correlation is given in the upper left of each panel. Dynamic recognition model values in the top two panels are the mean and standard deviation of the familiarity distributions for each item type at the time t at which $v^*(t) = 0.9$; these times differ between speed and accuracy emphasis owing to faster boundary collapse under speed emphasis.

memory evidence in our model with the stationary evidence distributions of the diffusion model, we computed the mean and variance of the accumulated change in familiarity at the time t at which decision boundaries had collapsed to 10% of their initial separation, i.e., the time at which $v^*(t) = 0.9$ (Equation 10). Because, according to model 5, decision boundaries collapse faster under speed emphasis, these times—and the associated familiarity mean and variance—will differ as a function of condition: for accuracy conditions, where $N_C^* = N_C = 30$, this time is 69 model time-steps; for speed conditions, where $N_C^* = 25.9$, this time is 60 model time-steps. As shown in the top two panels of Figure C1, mean familiarity at $v^*(t) = 0.9$ in our model and mean drift rate in the diffusion model are strongly correlated, whereas evidence variability in the two models shows only a weak systematic correlation. The lack of a strong correlation in evidence variability is likely due to the very different structure of the two models, as evidence variability plays somewhat different roles in each case. The strong correlation in mean evidence, however, indicates that our model provides a theoretically justified interpretation for differences in drift rates found in the diffusion model: As the number of study repetitions increases, more features are stored in the study trace (parameter u); low frequency words are comprised of more distinctive features than high frequency words (parameter g_L), leading to greater discriminability for low frequency words; speed instructions do not affect the contents of memory or the probe ($c_T = 1$), but rather lead participants to commit to a decision on the basis of fewer probe features than under accuracy instructions ($N_C^* < N_C$).

The model parameters governing more “decisional” aspects of processing, as opposed to memory evidence, are correlated between our model and the diffusion model, as shown in the lower two panels of Figure C1. Boundary separation decreases under speed instructions, and bias to say “old” increases as the proportion of targets on the test list increases; both models lead to the same interpretation of the effects of these independent variables. Both models also find that, although residual time is on average shorter under speed than accuracy emphasis (diffusion model mean residual times are .482 seconds in accuracy conditions, .431 seconds in speed conditions), residual time is more variable under speed emphasis (diffusion model residual time standard deviations are 0.0520 seconds in accuracy conditions, 0.0569 seconds in speed conditions¹⁰). Consistent with the group fits reported above, however, we find that the mean residual time in our model is less than that of the diffusion model; by explicitly modeling probe feature sampling, our model accounts for the dynamics of processes that separate-stage models of response time like the diffusion model must relegate to “residual” time.

Appendix D

Hierarchical Bayesian analysis of associative recognition SAT data

In this section, we present a set of analyses akin to those suggested by Liu and Smith (2009) that establish what properties of associative retrieval are affected by two manipulations: pair strength (Doshier, 1984a) and interference (Wickelgren & Corbett, 1977). These manipulations are critical because they could provide evidence for recall-like processing in associative recognition: Increasing pair strength should lead to faster recall while increasing interference should lead to slower recall, reflected in processing dynamics. We wished to re-analyze these data using more modern techniques with the aim of more firmly establishing what aspects of retrieval are affected by study time and repetition in associative recognition. This was made possible by the fact that Wickelgren

¹⁰The diffusion model fit reported by Starns et al. (2012) assumed a uniform distribution for residual time and, as such expressed variability in residual time as the range of this distribution; for ease of comparison with our model, we instead report the standard deviation, which for a uniform distribution is its range divided by $\sqrt{12}$.

and Corbett (1977) and Doshier (1984a) kindly reported their raw data—response probabilities for each participant and each pair type at each response signal lag—from all experiments.

We adopt a standard model of processing dynamics, namely, the exponential approach to asymptote:

$$d'(t) = \begin{cases} A [1 - \exp(-\frac{t-I}{R})] & \text{if } t > I \\ 0 & \text{if } t \leq I \end{cases} \quad (19)$$

where A is asymptotic d' , R is the processing rate, and I is the time at which processing starts. The likelihood function has a binomial form

$$\Pr(K_{i,j,k} | N_{i,j,k}, \theta_{i,j,k}) = \binom{N_{i,j,k}}{K_{i,j,k}} \theta_{i,j,k}^{K_{i,j,k}} (1 - \theta_{i,j,k})^{N_{i,j,k} - K_{i,j,k}} \quad (20)$$

where $K_{i,j,k}$ is the number of trials in which participant i gave a positive response to pair type j at lag k out of $N_{i,j,k}$ possible trials. The probability $\theta_{i,j,k}$ is derived from the model specified in Equation 19:

$$d'_{i,j,k} = \begin{cases} A_{i,j} [1 - \exp(-\frac{t_k - I_{i,j}}{R_{i,j}})] & \text{if } t > I_{i,j} \\ 0 & \text{if } t \leq I_{i,j} \end{cases} \quad (21)$$

$$\theta_{i,j,k} = \Phi \left[d'_{i,j,k} \left(\delta - \frac{1}{2} \right) + b_{i,k} \right] \quad (22)$$

where t_k is the time in seconds corresponding to the k th response signal lag, $A_{i,j}$ is the asymptotic d' for participant i to pair type j , $I_{i,j}$ is the start time (in seconds) for participant i to pair type j , $R_{i,j}$ is the processing rate for participant i of pair type j , δ is an indicator variable specifying whether the pair is intact or rearranged, and $b_{i,k}$ is a response bias for participant i at lag k . $\Phi[\cdot]$ is the standard normal CDF (we adopt the simplifying assumption that intact and rearranged pairs both have memory strength distributions with unit variance). Note that, as in our dynamic model, we assume that only the time between stimulus onset and the response signal t_k is used to process the stimulus and that the subsequent response time reflects only the detection of the signal and the execution of the response.

The free parameters $A_{i,j}$, $I_{i,j}$, $R_{i,j}$, and $b_{i,k}$ are estimated by drawing samples from the posterior of a hierarchical Bayesian model with vague priors and the structure depicted in Figure D1. The model was implemented in the JAGS Gibbs sampler (Plummer, 2013), using 10 parallel sampling chains each with 1000 steps of “burn-in” followed by 10,000 samples of which only every 10 were retained (this “thinning” procedure mitigates against autocorrelation in the samples). Thus, a total of 10,000 samples were drawn from the posterior distribution.

As described by Liu and Smith (2009), our Bayesian analysis affords several advancements over the model comparison approach that was originally available to Wickelgren and Corbett (1977) and Doshier (1984a). First, scaling and outlier effects resulting from the transformation of the raw data to d' scores—which can be difficult to estimate when response probabilities are very low or very high—are eliminated by using the binomial likelihood. Second, we need not impose specific and arbitrary penalties for model complexity because the parameter estimates themselves will tell us whether certain parameters need to have different values in order to account for the data (Kruschke, 2015). For practical purposes, we consider two parameters to be “credibly” different if the 95% credible interval of the posterior distribution of their difference excludes zero.

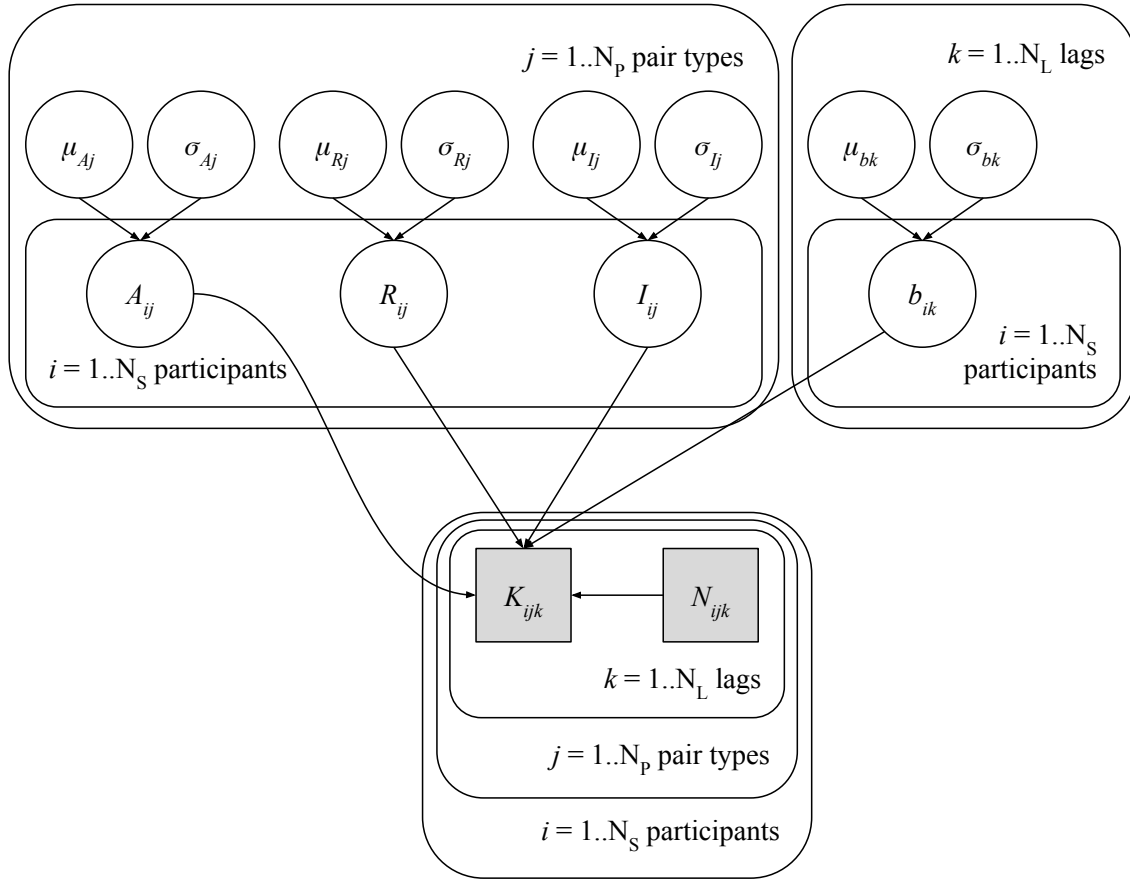


Figure D1. Hierarchical model used to estimate retrieval function parameters for speed-accuracy tradeoff experiments. See text for details.

Dosher (1984a)

To establish whether start time, processing rate, and/or asymptotic accuracy differed as a function of study time/repetition in associative recognition, Doshier (1984a) originally performed a series of model comparisons using the R^2_{Adj} statistic, which balances fit in terms of squared error against model complexity reflected by the number of free parameters. She found in her Experiment 2 that the R^2_{Adj} statistic favored a model in which only asymptotic accuracy A (not start time I or processing rate R) differed as a function of study time. In addition, she found in her Experiment 3 that R^2_{Adj} preferred a model in which asymptotic accuracy differed as a function of study type, but that the processing rate was affected not by total study time but by number of repetitions (with the exception of one participant, RS, who demonstrated exceptionally fast processing for pairs studied for 1 second).

Experiment 2. In Experiment 2, there were five participants ($i = 1..5$) and four pair types ($j = 1..4$) corresponding to four levels of study time: 1, 2, 4, or 6 seconds; intact and rearranged pairs from each group appeared with equal frequency. There were eight response signal lags at 0.2, 0.5, 0.7, 1, 1.5, 2, 3, and 4 seconds after stimulus onset. Based on the description of the original experiment, we inferred that there were an average of $N_{i,j,k} = 133$ trials of intact and rearranged

pairs of each type at each lag for each participant.

For our purposes, the critical question is what retrieval parameters are credibly affected by the study time manipulation. At the group level, asymptotic accuracy A increases with study time, with longer study times leading to credibly greater asymptotic accuracy across all levels (Figure D2a). This was not true of all participants, however: the asymptotic accuracies of participants BG (Figure D2b), DB (Figure D2c), and EG (Figure D2d) at the longest study times (2, 4, and 6 s) were credibly greater than that for 1 s but did not credibly differ from each other; meanwhile, asymptotic accuracies for 4 and 6 seconds of study did not credibly differ for participant JJ (Figure D2e) and those for 2 and 4 seconds of study did not credibly differ for participant WA (Figure D2f). In general, increased study time lead to greater asymptotic accuracy, but there are diminishing returns as study time is increased.

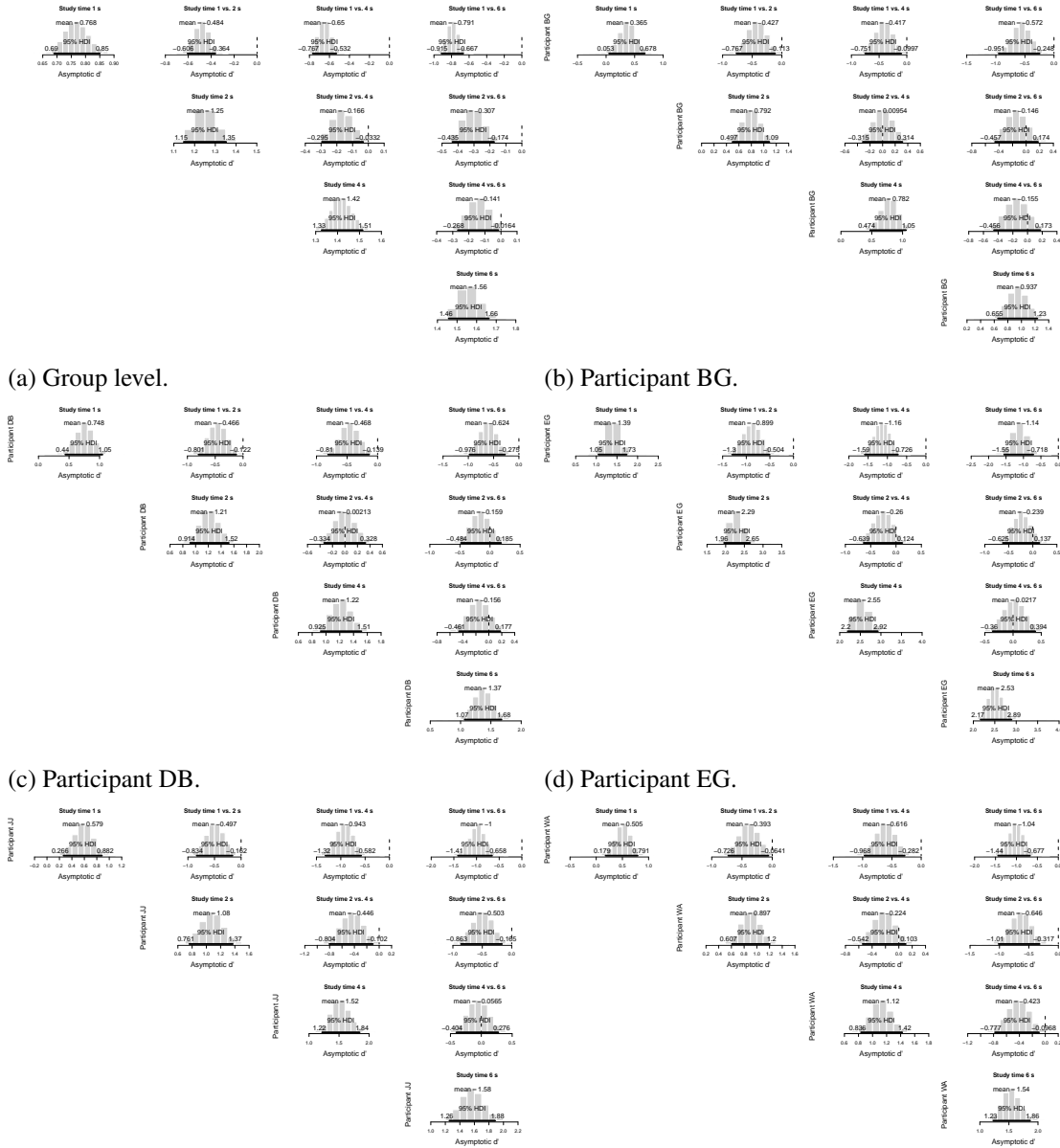
In accord with Doshier's original conclusions, there is no evidence that either start time I (Figure D3a) or processing rate R (Figure D4a) credibly differ as a function of study time at the group level, nor is there any evidence of an effect at the level of individuals (Figures D3b, D3c, D3d, D3e, and D3f for I ; Figures D4b, D4c, D4d, D4e, and D4f for R).

Experiment 3. In Experiment 3, participants studied pairs in four different study time conditions: once for 1, 2, or 5 seconds; or twice at 1 second each. Doshier concluded that although manipulating study time did not affect processing rate, repeated pairs were associated with faster retrieval. While plausible, this effect seemed to be small so we wished to determine its robustness at both the group and individual levels. Once again, there were five participants ($i = 1..5$) and four pair types ($j = 1..4$) for the four types of study and intact/rearranged pairs from each group appeared with equal frequency. There were seven response signal lags at 0.2, 0.5, 0.7, 1, 1.5, 2, and 3.5 seconds after stimulus onset and based on the description provided of the experiment, we inferred that there were an average of $N_{i,j,k} = 49$ trials of intact and rearranged pairs of each type at each lag for each participant.

Once again, we examine what retrieval parameters credibly differ as a function of the study time/repetition manipulation. At the group level, asymptotic accuracy (A) credibly increased with study time but only slightly different between a single 2 second presentation and two 1 second presentations (Figure D5a). Similar to Experiment 2, this qualitative ordering is mostly present in individuals as well, albeit with a different rate of diminishing returns as study time/repetitions increase (Figures D5b, D5c, D5d, D5e, and D5f).

In contrast to Doshier's original conclusions, however, we find no evidence that either processing rate (Figure D6a) or start time (Figure D7a) credibly differ as a function of study time/repetition, not even for any individuals (Figures D6b, D6c, D6d, D6e, and D6f for R ; Figures D7b, D7c, D7d, D7e, and D7f for I).

Discussion. Our Bayesian analysis of Doshier's data confirm that study time/repetition has a credible effect on associative recognition accuracy, but we also find that these data provide no evidence that processing dynamics (either rate or start time) differ as a function of study time. This is not to say such an effect does not exist or might not be detected in a different setting or with more observations. Our analysis does, however, lend credence to our claims and modeling assumptions in the main body of the paper that memory strength, as manipulated by study time/repetition, may not affect the dynamics of associative recognition.



(e) Participant JJ. (f) Participant WA.

Figure D2. Posterior distributions of means (diagonals) and subtractive contrasts (upper triangle) for asymptotic accuracy A in Doshier (1984a, Exp. 2).

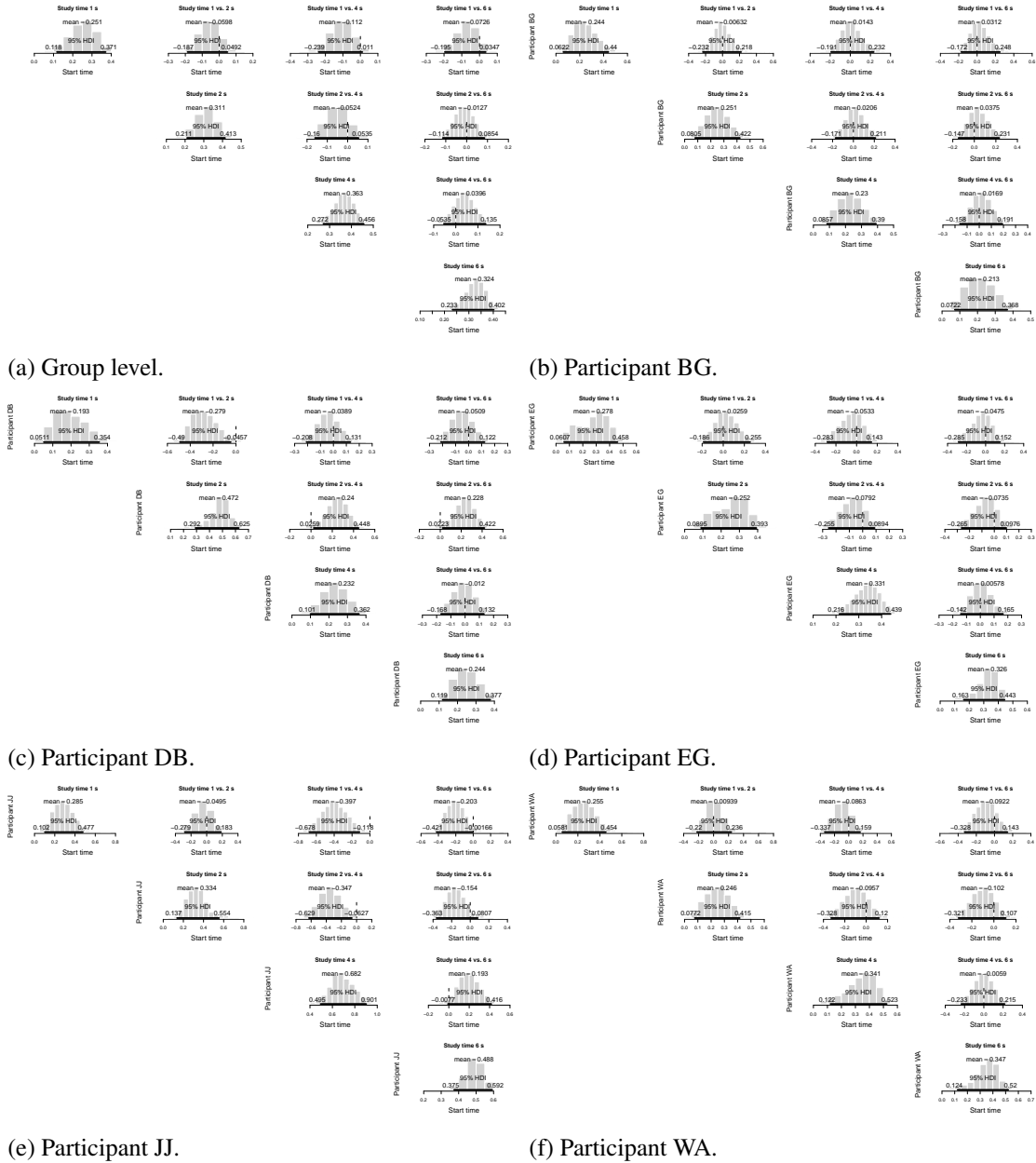


Figure D3. Posterior distributions of means (diagonals) and subtractive contrasts (upper triangle) for start time I in Doshier (1984a, Exp. 2).

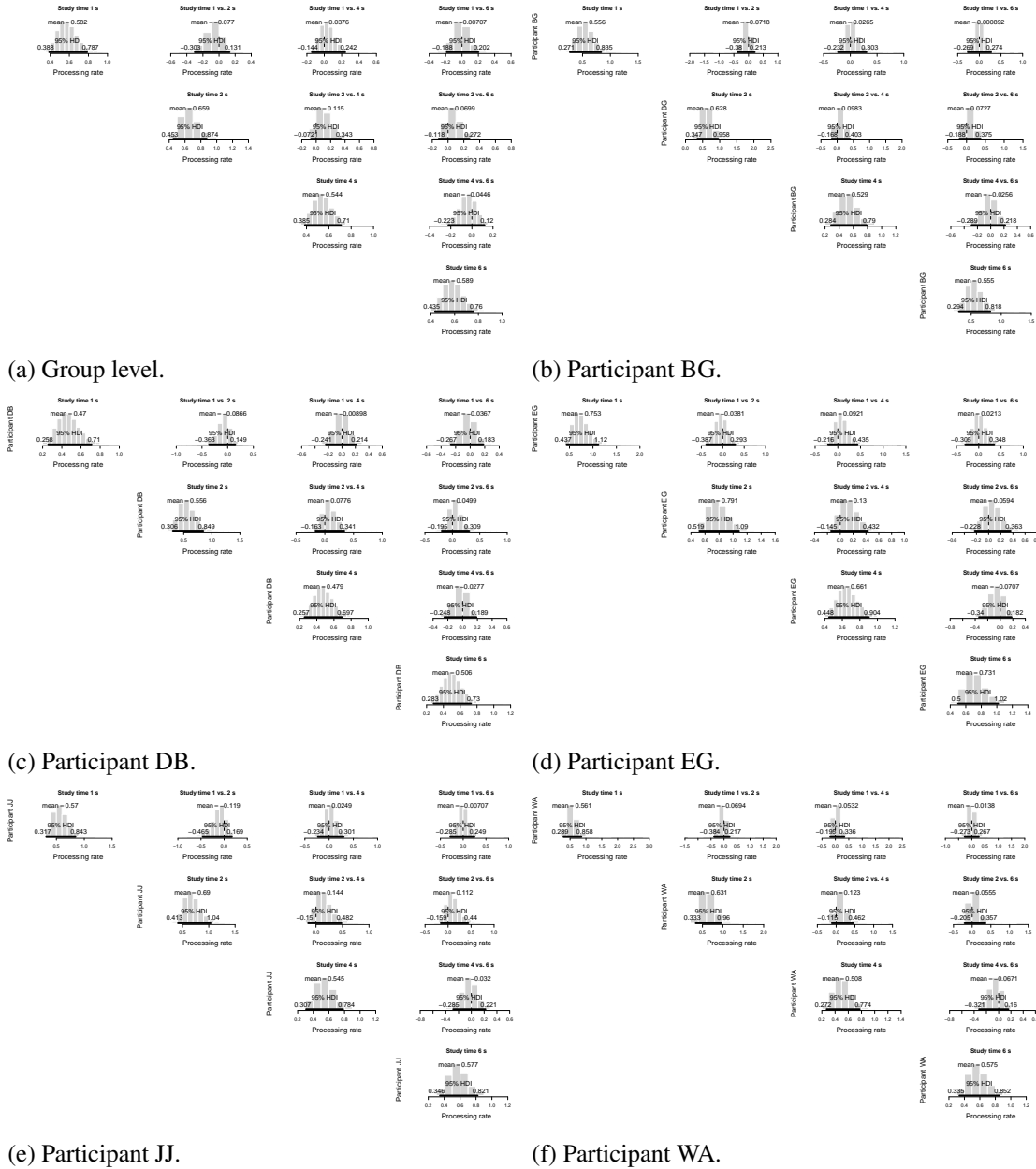


Figure D4. Posterior distributions of means (diagonals) and subtractive contrasts (upper triangle) for processing rate R in Doshier (1984a, Exp. 2).

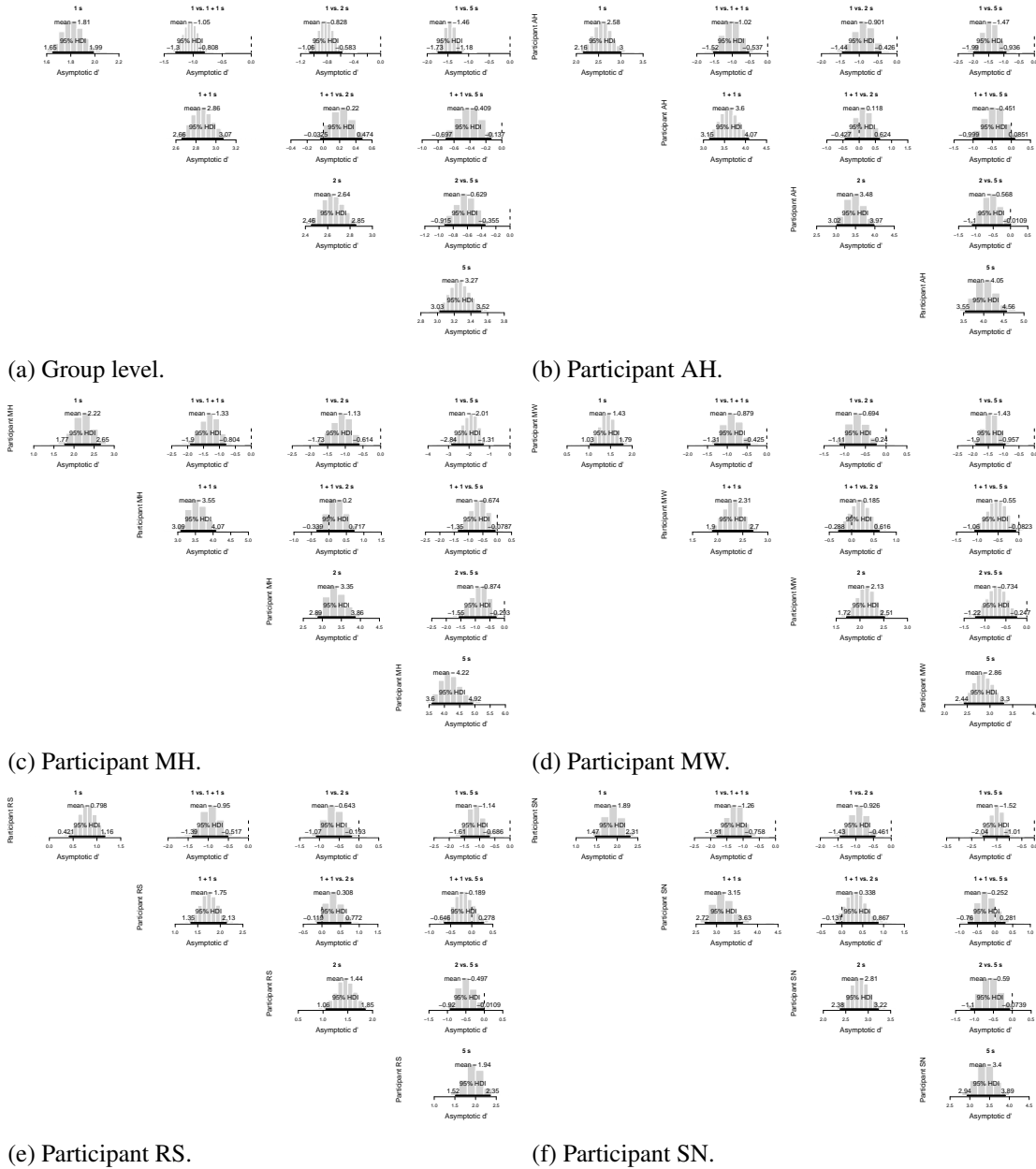


Figure D5. Posterior distributions of means (diagonals) and subtractive contrasts (upper triangle) for asymptotic accuracy A in Doshier (1984a, Exp. 3).

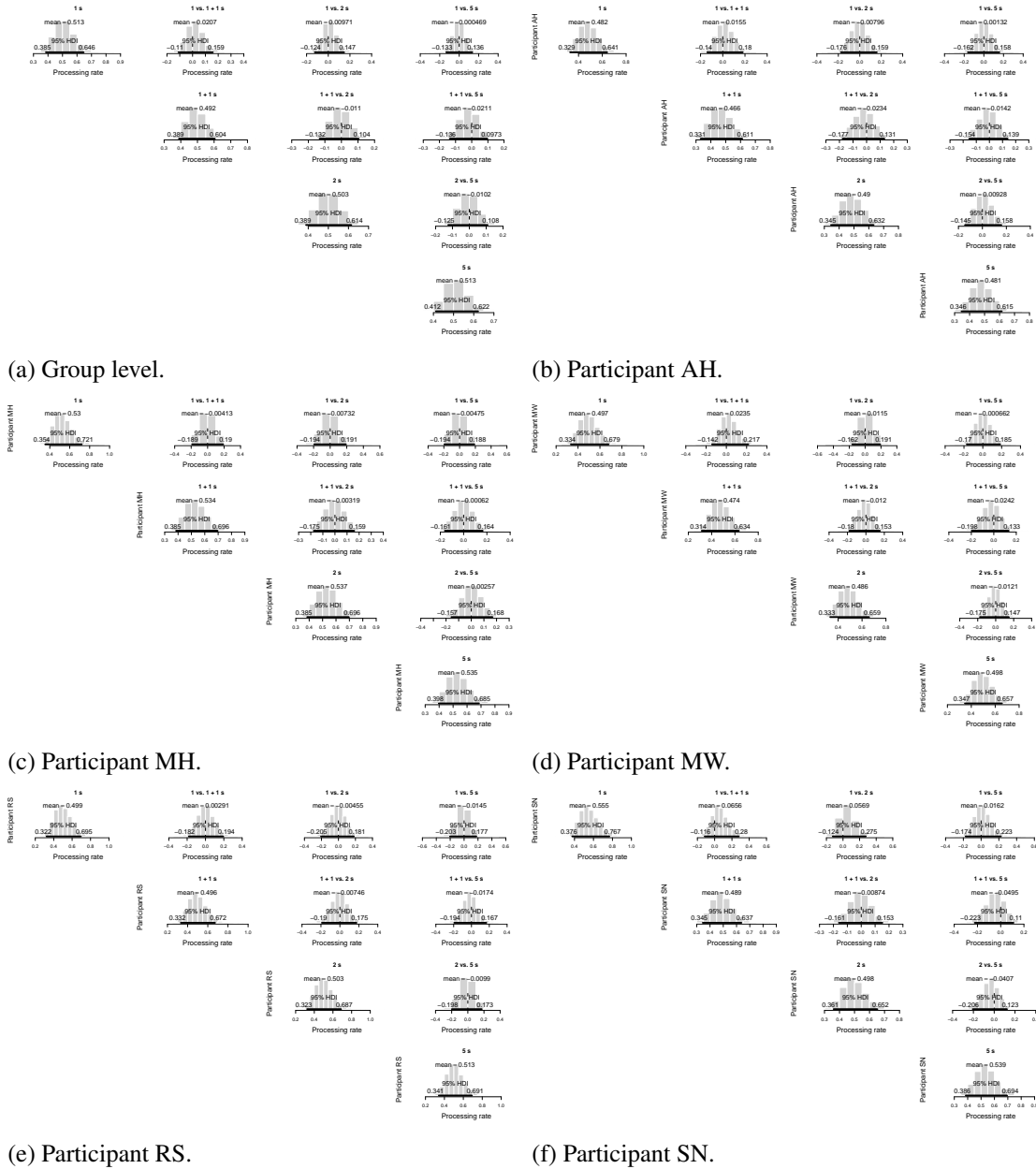


Figure D6. Posterior distributions of means (diagonals) and subtractive contrasts (upper triangle) for processing rate R in Doshier (1984a, Exp. 3).

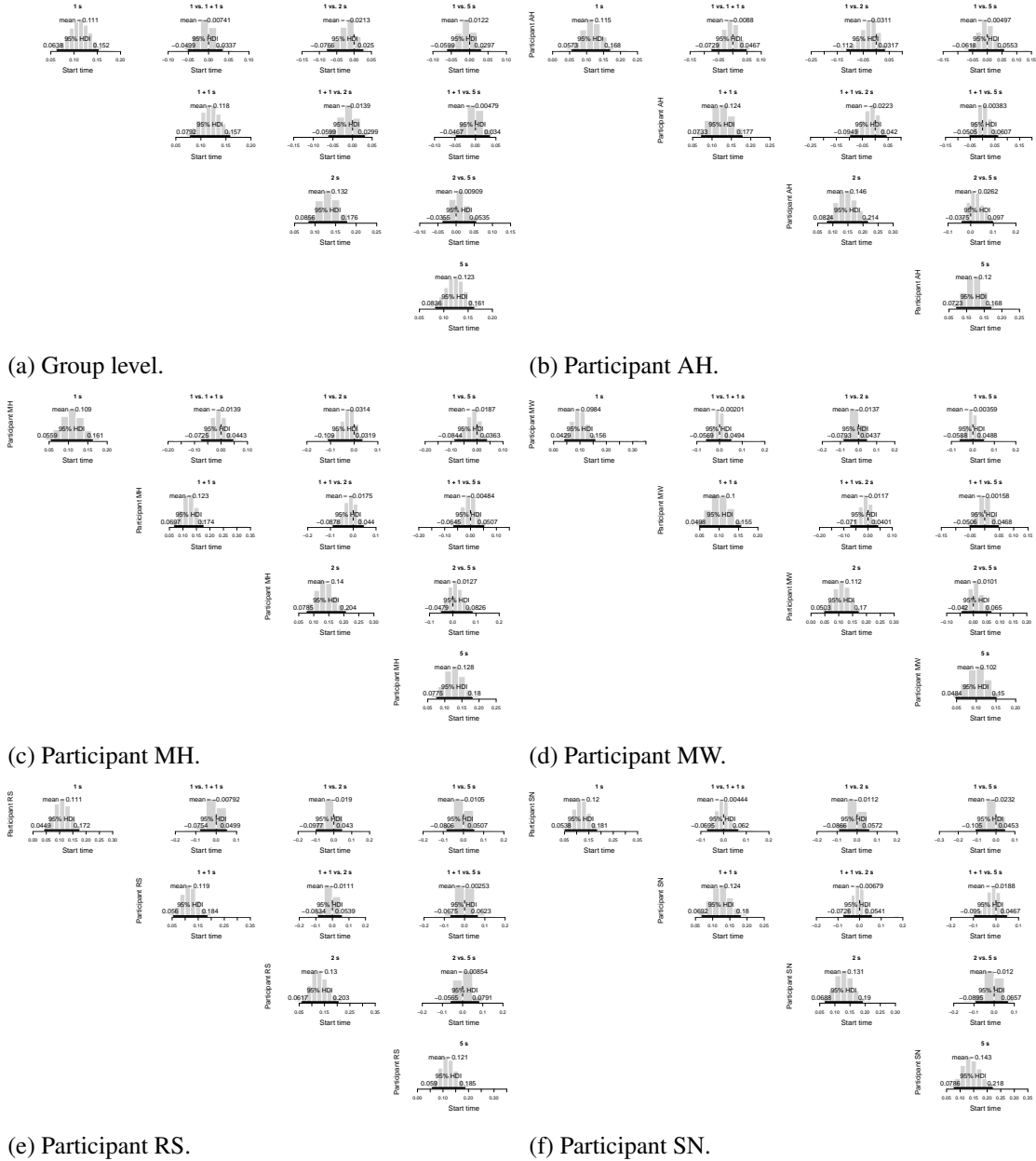


Figure D7. Posterior distributions of means (diagonals) and subtractive contrasts (upper triangle) for start time I in Doshier (1984a, Exp. 3).

Wickelgren and Corbett (1977)

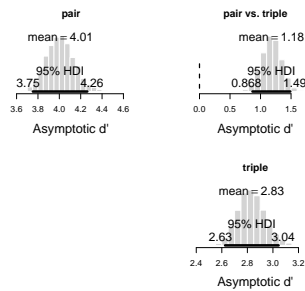
In the associative recognition study reported by Wickelgren and Corbett (1977), participants studied a mixed list of word pairs and triplets. At test, they had to distinguish between intact pairs (from either studied pairs or as part of a studied triplet) and rearranged pairs. Studying triplets potentially introduces additional associative interference. If associative recognition relies on a recall-like process, this interference should slow retrieval such that either the start time I is greater or processing rate R is slower for items studied as part of a triple relative to those studied as part of a pair. In their original model comparison, Wickelgren and Corbett (1977) reported moderate evidence of an effect of interference on retrieval dynamics, although it was not clear whether the effect was a delay in start time or a slowing of processing rate or neither. Our analysis will help adjudicate this question.

There were six participants ($i = 1..5$) and two pair types (studied as part of a pair or triple, $j = 1..2$) and intact/rearranged pairs from each group appeared with equal frequency. There were eight response signal lags at 0.2, 0.3, 0.4, 0.6, 0.9, 1.2, 2.5, and 4 seconds after stimulus onset and based on the description provided of the experiment, we inferred that there were $N_{i,j,k} = 72$ trials of intact and rearranged pairs of each type at each lag for each participant.

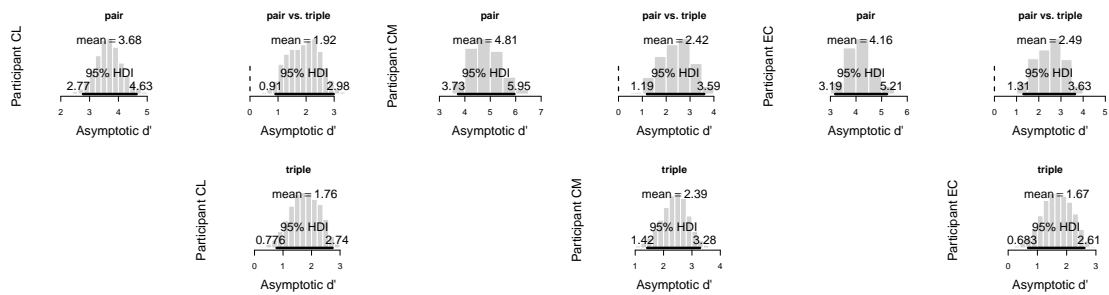
In accord with their original conclusions, we find that asymptotic accuracy is credibly greater for pairs than triples at the group level (Figure D8a) and for all but one participant (Figure D8e); this participant demonstrated the best performance of all participants, so this may be a ceiling effect.

There is no evidence for a credible difference in start time between pairs and triples at the group level (Figure D9a) or for most participants, however participant CM may be slower to start processing triples than pairs, albeit this difference only barely credible (Figure D9c). There is also slight evidence that processing rate differs between pairs and triples (Figure D10a), although this difference is not quite credible at either individual or group levels.

In sum, our analysis generally leads to the same conclusions found by Wickelgren and Corbett (1977). Associative interference (triples vs. pairs) leads to lower asymptotic memory performance, but may or may not have an effect on processing dynamics. It is unlikely that interference affects the start of processing, but it is possible that interference hampers retrieval speed. These data do not provide any strong evidence of an effect of interference on retrieval dynamics, suggesting that our modeling assumptions in the main body of the paper remain tenable.



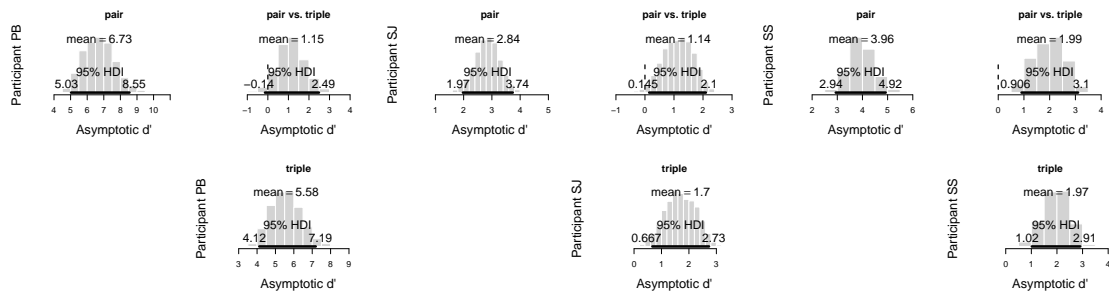
(a) Group level.



(b) Participant CL.

(c) Participant CM.

(d) Participant EC.



(e) Participant PB.

(f) Participant SJ.

(g) Participant SS.

Figure D8. Posterior distributions of means (diagonals) and subtractive contrasts (upper triangle) for asymptotic accuracy A in Wickelgren and Corbett (1977).

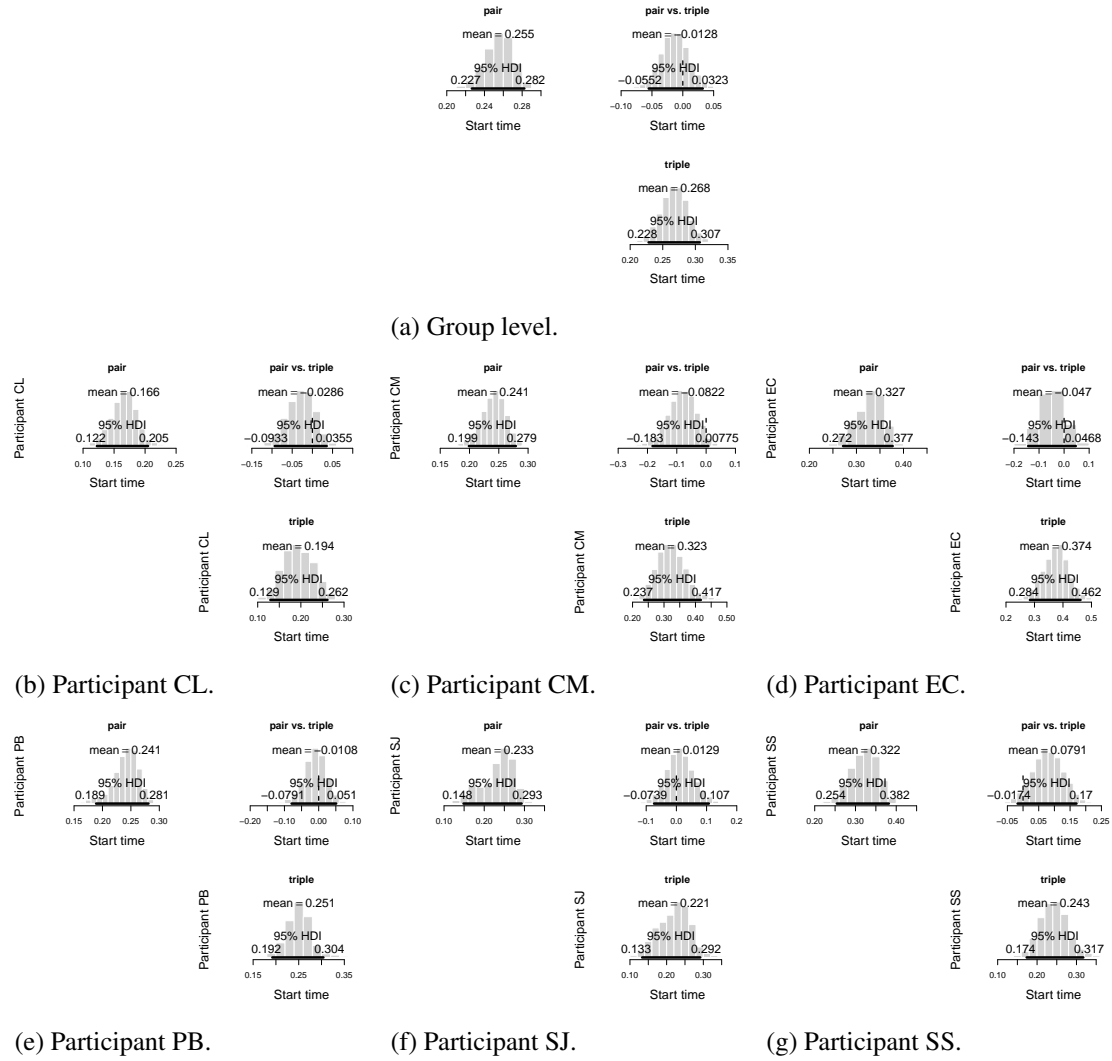


Figure D9. Posterior distributions of means (diagonals) and subtractive contrasts (upper triangle) for start time I in Wickelgren and Corbett (1977).

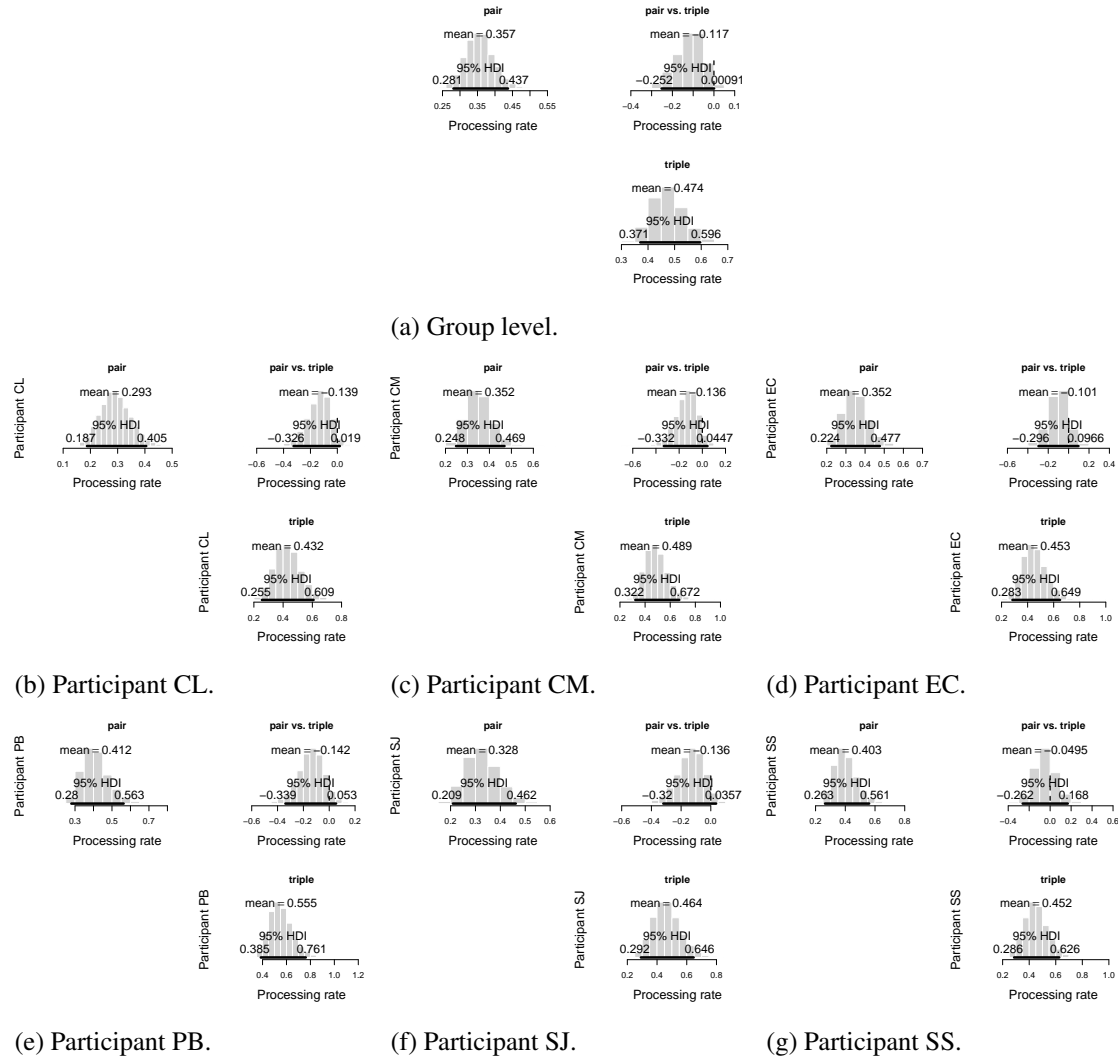


Figure D10. Posterior distributions of means (diagonals) and subtractive contrasts (upper triangle) for processing rate R in Wickelgren and Corbett (1977).