Exploring Decision Rules and Sampling Dynamics in Recognition Memory

Gregory E. Cox (grcox@indiana.edu)
Richard M. Shiffrin (shiffrin@indiana.edu)
Department of Psychological and Brain Sciences, Indiana University
1101 E. Tenth St., Bloomington, IN 47405 USA

Abstract

Cox and Shiffrin (2012) introduced a dynamic model of recognition memory that was capable of making simultaneous predictions for accuracy and mean response time. The present paper extends that work by investigating the assumptions underlying the model's decision process, in particular those pertaining to the process by which features are sampled at test and the processes by which evidence for an "old"/"new" recognition decision accumulates. These assumptions are tested against empirically collected response time distributions. Evidence is found that sampling dynamics can change in response to instructions, and that competition between accumulators for "old" and "new" evidence is not necessary to capture at least the recognition data presented here.

Keywords: Recognition memory; response time; memory models.

Introduction

Cognitive scientists have long understood the utility of timing information for making inferences about the cognitive processes underlying behavior. Yet, the field of recognition memory has only made limited use of this rich source of data. This is partially because of a lack of available data to constrain dynamic models of recognition, and partially because most theories of recognition are themselves static: Recognition decisions are presumed to operate on a fixed value of "familiarity", which is a signal indicating the presence or absence of an item in memory. If the familiarity is above a particular criterion, a participant declares the item "old", otherwise it is considered "new". Most extant models of memory are concerned purely with how this familiarity value is calculated for different items. Some of these models assume that an item is compared to many individual memory traces (e.g., Hintzman, 1988; Shiffrin & Steyvers, 1997; McClelland & Chappell, 1998), to an aggregate of several memory traces (Murdock, 1982), or to the present context (e.g., Dennis & Humphreys, 2001; Howard & Kahana, 2002).

We have recently proposed (Cox & Shiffrin, 2012) that taking a dynamic, as opposed to static, approach to recognition will shed light on many previously vexing issues. Our original work was concerned with the problem of criterion setting in the case of widely-varying stimulus materials, where the absolute value of familiarity may fluctuate wildly between different test items. Although, in principle, criteria might be learned over time (c.f., Turner, Van Zandt, & Brown, 2011), we showed that the problem of learning stimulus-specific criteria is bypassed entirely if one treats recognition decisions as inherently dynamic in nature. In this case, rather than basing a decision on a fixed asymptotic level of familiarity, recognition can make use of the *changes in* familiarity over time as information is gathered from the test stimulus. The shape of the "familiarity profile" generated by this gradual accrual of information is invariant to many of the factors that would result in different levels of asymptotic familiarity, e.g., amount of prior exposure to the stimulus and degree of encoding.

Although the model proposed by Cox and Shiffrin (2012) was originally motivated to address the criterion setting problem, it has potential application as a general model of recognition. We showed that the predictions of the model for accuracy and mean response times were in accord with published data on recognition memory, but left many questions unanswered. For example, we assumed that the decision took the form of a race between two parallel accumulators, one leading to an "old" response and the other to a "new" response. This decision mechanism was not explicitly compared to other possible mechanisms, e.g., those with non-independent accumulators like a random walk or diffusion process (e.g., Ratcliff, 1978). Further, we assumed that the times between sampling events—i.e., sampling a feature from the test stimulus—were drawn independently and identically from an exponential distribution, but made no attempt to investigate whether this sampling process might differ between study and test and between conditions.

At the level of mean RT, these sets of predictions would be difficult to disentangle from those arising from other assumptions regarding correlated accumulators or different sample time distributions. Thanks to the generous contributions of Chris Donkin and Andrew Heathcote, we are now able to test the assumptions of our original dynamic recognition model regarding the decision mechanism and sampling dynamics. In so doing, we have taken yet more steps toward developing a general theory of recognition that is able to account for both accuracy and response time.

The Model

The modeling framework used in this paper is a direct outgrowth of the dynamic recognition model proposed by Cox and Shiffrin (2012), which is itself based on the REM model (Shiffrin & Steyvers, 1997). The central tenet of the dynamic recognition model is that information about a test item is sampled over time and added to a memory probe. When the probe is compared to the contents of memory at time t, it produces a particular familiarity value, $\phi(t)$. How this value changes over time constitutes evidence for an old/new recognition decision: decrements in $\phi(t)$ (relative to $\phi(t-1)$) are evidence that the item is new, because additional item information *reduces* its familiarity. Contrariwise, increases in $\phi(t)$ are evidence that the item is old, because new item information increases its familiarity. We now give a more detailed exposition of the model, although the reader is directed to Cox and Shiffrin (2012) for additional information.

Representations and Encoding

An episodic memory trace for an item is represented as a vector of features, each of which is binary with an equal *a priori* probability of being "0" or "1". These features are divided into two kinds: *content* features which arise from the item itself, which may include information about its sensory characteristics or semantics, and *context* features which arise from the situation in which the item was encountered. Episodic memory is presumed to consist of a (very large) set of such traces from across the life span, although for practical purposes, we restrict ourselves to modeling just a subset of these traces (see below). In addition, we assume that two traces of the same item need not be encoded using the same features. For example, a particular apple might be encoded primarily with features pertaining to its physical appearance on one occasion, but on another the apple might be encoded primarily with features pertaining to its taste. This encoding variability is captured in the model by a parameter α, the probability that any one feature is shared between two traces of the same item (independent of whether a *value* is stored for that feature). Likewise, the parameter γ denotes the probability that a feature is shared between two traces of *different* items. The exact values of these parameters will depend on the kinds and varieties of items presented during study and test and are at least partially under the decision maker's control.

A memory trace is formed whenever an item is encountered, e.g., on a study list[1]. Memory traces are formed by copying into long-term memory a short-term memory representation of the item. The short-term memory representation is built up over time. Context features are presumed to be already present in the representation, since they are persistent in the environment. Content features are gradually sampled from the test item and added to the short-term memory representation. Such features are sampled at random, with replacement, and are copied into the STM representation correctly with probability c, otherwise a random value (either "0" or "1") is copied. Note that, because the sampling occurs with replacement, a new sampled value can replace one that was previously present in the STM representation, even if the new value was copied incorrectly. Finally, due to capacity limitations, we assume that at most N_c content features and N_x context features are available to enter into STM.

At study, the sampling process is terminated at stimulus offset after some specified time, T_s. We assume that the dynamics of the sampling process are governed by a homogeneous Poisson process, that is, the time between feature samples is exponentially distributed with rate ρ_s. Thus, the number of samples obtained at study (which may involve "overwriting" a previously sampled feature value) after T_s time units is Poisson distributed:

$$\Pr(k \text{ samples}; \rho_s, T_s) = \frac{(\rho_s T_s)^k}{k!} \exp(-\rho_s T_s).$$

At that point, the STM representation of the study item, comprised of the current context features plus whatever content features were sampled from the item in the time available, is copied into a long-term memory trace. Again, any one feature may be incorrectly copied from the STM representation into LTM with probability c, otherwise a random value is stored. Thus, a memory trace consists of N_x context features, some of which may have been copied incorrectly, and N_c content features, some of which have (correctly or incorrectly) sampled values and others of which are "blank", indicating that no value was sampled for that feature.

At test, the STM representation of the test item serves as a probe that is compared to the contents of memory. This comparison is made after each sampling event t, resulting in a familiarity value, $\phi(t)$. The change in $\phi(t)$ from one sampling event to the next constitutes the evidence for making a recognition decision. We now turn to the details of how the comparison between the probe and memory is made and how $\phi(t)$ is calculated.

Computing Familiarity

To compute a familiarity value, after each sampling event, the STM representation of the test item is used as a probe and is compared, in parallel, to all traces in episodic memory. Each comparison results in a measure of the similarity between the probe and a given trace, denoted $\lambda_i(t)$ (where i indexes the trace in memory). The similarity measure takes the form of a likelihood ratio: the likelihood that the probe and the trace encode the *same* item versus the likelihood that they encode *different* items.

Likelihood Calculation The evidence that goes into computing these likelihoods comes from the features that have been sampled thus far and added to the probe (by time t) and from the features that had been stored in the trace at study. For any one feature, there are five possible outcomes of the comparison:

1. The probe and trace both have a value stored, and the value matches (M).
2. The probe and trace both have a value stored, and the values do not match (N).
3. The probe has a value stored, but the trace does not (P).
4. The trace has a value stored, but the probe does not (T).
5. Neither the trace nor the probe have a value stored.

Although the full model described in Cox and Shiffrin (2012) makes use of all these possible outcomes, for present purposes we are only concerned with the first two. This is because outcomes P and T are only indicative of a non-match if traces of the same item are more likely to share features than are traces of different items (i.e., if $\alpha > \gamma$). In the simulations reported here, we deal only with items that are of the

---

[1] A memory trace would also be formed after each test trial. We do not attempt to model this here, but the formation of memory traces at test may be necessary to explain several phenomena in recognition memory (Criss, Malmberg, & Shiffrin, 2011).

same type (namely, nouns) and so we assume that the degree of encoding variability is uniform, i.e., $\alpha = \gamma$.

Although the reader is referred to Cox and Shiffrin (2012) for the details of these derivations, the following are the conditional probabilities needed to compute a match:

$$\begin{aligned} \Pr(M|\text{Same}) &= c + (1-c)\tfrac{1}{2} & \Pr(M|\text{Different}) &= \tfrac{1}{2} \\ \Pr(N|\text{Same}) &= (1-c)\tfrac{1}{2} & \Pr(N|\text{Different}) &= \tfrac{1}{2}. \end{aligned}$$

Because features are sampled independently, these probabilities are multiplied together for as many feature comparisons result in a M or N outcome, as appropriate to obtain the likelihood under each hypothesis, *same* or *different*. Letting N_M and N_N denote the number of feature-value matches and mismatches, respectively, the likelihood ratio for a trace i can thus be written:

$$\begin{aligned} \lambda_i(t) &= \left[\frac{\Pr(M|\text{Same})}{\Pr(M|\text{Different})}\right]^{N_M} \left[\frac{\Pr(N|\text{Same})}{\Pr(N|\text{Different})}\right]^{N_N} \\ &= (1+c)^{N_M}(1-c)^{N_N}. \end{aligned} \tag{1}$$

Familiarity Calculation A likelihood ratio $\lambda_i(t)$ is produced, in parallel, for all traces in memory. However, the vast majority of traces in memory will be an extremely poor match to the probe and produce very low likelihood ratios, either because these traces encode different items (differ in content features) or because they were encoded in vastly different situations (differ in context features). Thus, we assume that only some traces in memory are "active" at any one time. For a trace to be activated in response to a probe, it must produce a likelihood ratio greater than a threshold θ. For simplicity, we fix $\theta = 1$[2]. We denote the set of active traces, i.e., those for which $\lambda_i(t) > \theta$ by $A(t)$ (of size $|A(t)|$). Only those traces in $A(t)$ contribute to the familiarity value $\phi(t)$. Because traces that fail to match on either content (i.e., they encode the same item) or context (i.e., they were encoded under similar circumstances, e.g., in a memory study list) are extremely unlikely to pass threshold and entire $A(t)$, we only model storage of the N items from the study list plus K traces of the test item from different contexts.

Having computed a match value $\lambda_i(t)$ for each trace in memory and selected the set of activated traces $A(t)$, the familiarity value, $\phi(t)$, is simply the average likelihood over all active traces, i.e.,

$$\phi(t) = \frac{1}{|A(t)|}\sum_{i \in A(t)} \lambda_i(t). \tag{2}$$

Decision Mechanism

Our model assumes that recognition decisions are based not on the absolute value of familiarity, but rather on how this value changes over time. Because the likelihood ratio scale is highly skewed, we first take the logarithm $\log\phi(t)$ (this transformation does not alter the underlying logic of the model); the evidence for recognition decisions is then $\nabla\log\phi(t) = \log\phi(t) - \log\phi(t-1)$. Positive values of $\nabla\log\phi(t)$ are considered evidence that the test item is "old" while negative values are treated as evidence that the item is "new" (i.e., has not be encountered in the current context).

In our original model, positive and negative changes in familiarity feed into two parallel, non-interacting accumulators, $B^+(t)$ and $B^-(t)$. That is,

$$\begin{aligned} B^+(t) &= \sum_{\tau=0}^{t} \begin{cases} \nabla\log\phi(\tau) & \text{if } \nabla\log\phi(\tau) > 0 \\ 0 & \text{if } \nabla\log\phi(\tau) \le 0 \end{cases} \\ B^-(t) &= \sum_{\tau=0}^{t} \begin{cases} \nabla\log\phi(\tau) & \text{if } \nabla\log\phi(\tau) < 0 \\ 0 & \text{if } \nabla\log\phi(\tau) \ge 0 \end{cases}. \end{aligned}$$

The final recognition decision is, then, a race between these two accumulators: the predicted response is given by whichever accumulator reaches its threshold first (β^+ or β^- for $B^+(t)$ and $B^-(t)$, respectively).

The predicted response time is related to the number of sampling events t needed for the first of the accumulators to reach its threshold. In our previous work, we assumed, as is common in many models that posit sequential independent feature samples from a test stimulus (e.g., Brockdorff & Lamberts, 2000; Mewhort & Johns, 2005), that the times between sampling events were exponentially distributed with a fixed, uniform rate ρ_t (which may be different than the sampling rate at study, ρ_s). This has the consequence that the observed response time, given that t samples were needed for one of the accumulators to reach threshold, is drawn from a Gamma distribution (the convolution of t independent and identically distributed exponential distributions).

Simulations

Although our previous work on a dynamic model for recognition memory included predictions for mean response times, a much stronger test of the model is to compare its predictions to empirically collected *distributions* of response times. Unfortunately, empirical RT distributions in recognition memory are still somewhat rare (despite the fact that one of the leading models of response time, the diffusion model of Ratcliff, 1978, was originally developed to account for RT distributions in recognition memory!). Thus, we once again express our gratitude to Chris Donkin and Andrew Heathcote for providing us with empirical RT distributions against which to compare the predictions of our model. This comparison affords special insight into two features of our model that had to be simply assumed in our earlier work: first is the distribution of times between samples at test. Second is the assumption of independence between the two accumulators.

In obtaining a correspondence between the model and data, we fixed most of the parameters involved at the values given in Table 1. Our primary goal in these simulations was not

---

[2] The full model assumes that θ depends on the ratio $\frac{\alpha}{\gamma}$, but these are assumed equal in the subsequent simulations and so play no role in the setting of the activation threshold.

Table 1: Summary of the fixed parameters of the model, along with their values used in the present simulations.

Parameter	*Value*	*Description*
N_c	30	Maximum number of content features.
N_x	30	Number of context features.
c	0.85	Probability of correct feature copy.
K	200	Number of history traces of a test item available to be activated.
θ	1	Minimum likelihood needed to enter the set of active traces.
ρ_s	60	Feature sampling rate at study.

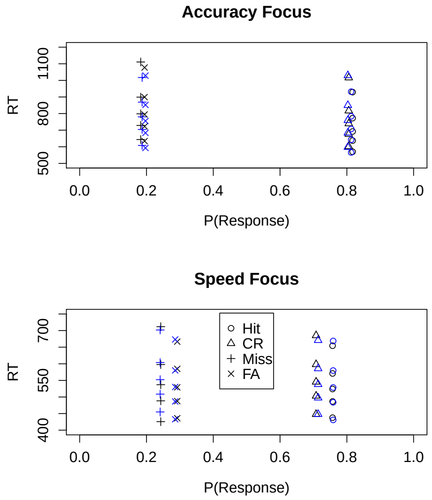

Figure 1: Observed group performance and RT quantiles (10%, 30%, 50%, 70%, 90%) are shown in black, with model predictions using the independent accumulator mechanism in blue.

to obtain the best quantitative fit possible, which could potentially require all model parameters to be freely varied. Rather, we wished to capture the qualitative features of the data whilst varying only a small number of parameters, thereby allowing a more direct interpretation of the model's predictions.

Sampling Dynamics

Heathcote and Donkin (2012) recently collected both accuracy and response times in a recognition memory paradigm. Participants studied lists of 25 words each. At the conclusion of each study list, participants were tested in a standard old/new recognition paradigm. The test lists were unbiased (i.e., composed of an equal number of studied and unstudied words). On half of the test lists, participants were instructed to try to be as accurate as possible without taking too long to make their response ("accuracy focus"), while on the other half, participants were instructed to be as fast as possible without sacrificing accuracy ("speed focus"). The resulting group mean accuracy and RT quantiles (along with model fits) are plotted in Figure 1.

Table 2: Best-fitting parameter values for the independent accumulator mechanism, given the Heathcote and Donkin (2012) recognition data.

Parameter	*Condition*	*Value*
β^+	Accuracy	13
	Speed	6
β^-	Accuracy	-17
	Speed	-9.4
ρ_t	Accuracy	158 samples per second
	Speed	105 samples per second
T_n	Accuracy	297 ms
	Speed	266 ms

To replicate this procedure in our model, we assumed that 25 items were studied for $T_s = 1$ time unit, fixed the other model parameters at the values given in Table 1, and simulated 1000 study/test lists. To fit the model to the observed RT distribution data, we wished to optimize four parameters, which we allowed to vary between the speed and accuracy focus conditions: The first two of these are the thresholds for the old and new accumulators, β^+ and β^-, respectively. We assume that participants can adjust their decision criteria in response to instructions, with lower thresholds leading to faster but potentially more error-prone responses. Two additional variables were allowed to vary between condition: the sampling rate at test, ρ_t, and the amount of "non-decision time" per trial, T_n. Non-decision time is introduced to account for any processes that are not being explicitly modeled, e.g., the time required to initially attend to the test stimulus and the time required to actually execute the motor response. Just as we assume that participants are able to adjust their decision bounds, we assume that instructions can induce participants to devote more resources toward particular components of the recognition process. For example, speed instructions may lead to faster execution of the motor response—leading to reduced non-decision time—but the added time pressure may result in less efficient extraction of information from the test stimulus—and therefore a lower average sampling rate.

Indeed, this is the pattern found in the fitted parameters, shown in Table 2 (which achieved an adjusted $R^2 = .98$ between predicted and observed RT quantiles). The thresholds are farther apart in the accuracy condition than in the speed condition, as one would expect. In addition, non-decision time is slightly lower in the speed condition, which may be attributed to a slight decrease in the time needed to execute the response resulting from practice in the blocks of speed trials. Most interesting, however, is that although response thresholds are lower—and responses correspondingly faster—in the speed focus condition, the sampling rate is estimated to be substantially lower in the speed condition[3]. In other words, although participants appear willing to make responses on the basis of less evidence when encouraged to produce fast responses, participants appear to be collect this evidence less efficiently. One possible explanation for this is, to paraphrase Starns, Ratcliff, and McKoon (2012), that the additional metacognitive demands in the speed condition (e.g., the need to monitor response time and avoid pure guessing)

---

[3] In addition, both estimated sampling rates at study are faster than the sampling rate assumed at study, which was fixed at 60 samples per second.

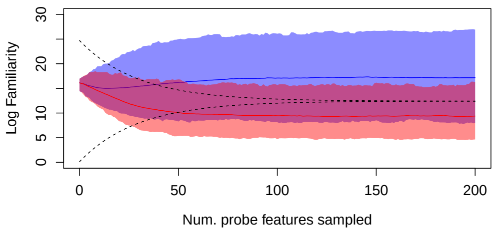

Figure 2: Collapsing thresholds (dashed lines) relative to the mean familiarity value for targets (blue line) and foils (red line). Light regions around the mean familiarity show the area between the 2.5% and 97.5% quantiles of the familiarity distribution at each time.

divert resources from processing the test stimulus itself.

Correlated Accumulators

Another feature of our model that was previously untested is the assumption of independence between the positive and negative accumulators. This independence places our model in a different class than random walk or diffusion models (e.g., Ratcliff, 1978), which assume that evidence in favor of one option (e.g., "old") is equivalent evidence *against* the other option (e.g., "new"). A random walk model may thus be seen as a model with two perfectly anti-correlated accumulators.

The primary reason for choosing an independent accumulator structure in our original work was the nature of the evidence on which we presume recognition decisions are based. In particular, because there is a maximum number of features that may be sampled, there is also a maximum (and minimum) value of familiarity that could result from a test item. Consider the case where, by a lucky happenstance, all the features sampled and added to the probe at time t perfectly match a single trace stored in memory. Further assume that this single trace is the only activated trace. Then, from equations 1 and 2, $\phi(t) = (1+c)^{N_c+N_x}$. Because N_c and N_x are fixed, this is the maximum possible familiarity value. Clearly, any subsequent feature samples can only lead to a zero or negative change. Thus, if the threshold on the positive counter has not yet been reached, and the positive and negative counters are anti-correlated, the "old" threshold will never be attained because no further positive changes are possible.

Thus, if the accumulators are to be perfectly anticorrelated, as in a random walk, their thresholds cannot be fixed across time. This is not merely a feature of this particular model, but any model that places a limit on the amount of evidence that may be accrued over time. Thus, we introduce a rule by which the accumulator thresholds may collapse over time (for examples of collapsing thresholds in other domains, see Balakrishnan & Macdonald, 2011; Frazier & Yu, 2008). This rule is but one of many possible rules, but is based on the principle that the thresholds should be reduced in proportion to the amount of information that remains to be sampled from the stimulus. On any feature sampling event, each of the N_c content features has an equal chance of being sampled. Thus, the probability that any one feature has *not* been sampled after t sampling events in $\left(1-\frac{1}{N_c}\right)^t$. Imagine that the decision maker has a number of features N_{max} that they consider sufficient for making a recognition decision (Diller, Nobel, & Shiffrin, 2001). Then, the probability that N_{max} features have been sampled after t sampling events is $\sigma(t) = \left[1-\left(1-\frac{1}{N_c}\right)^t\right]^{N_{max}}$. Thus, given initial thresholds $\beta^+(0)$ and $\beta^-(0)$, the thresholds collapse toward one another in a symmetric fashion:

$$\beta^+(t) = \beta^+(0) - \sigma(t)\left(\frac{\beta^+(0)-\beta^-(0)}{2}\right) \tag{3}$$

$$\beta^-(t) = \beta^-(0) + \sigma(t)\left(\frac{\beta^+(0)-\beta^-(0)}{2}\right). \tag{4}$$

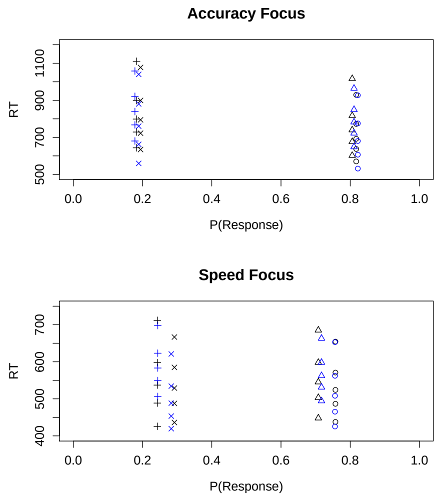

Figure 3: Observed group performance and RT quantiles (10%, 30%, 50%, 70%, 90%) are shown in black, with model predictions—using the correlated accumulators with collapsing thresholds—in blue.

Different choices of N_{max} can lead to very different threshold behavior, however both our own investigations (not reported here for space) and the empirical work of Balakrishnan and Macdonald (2011) suggest that thresholds begin rather far apart and thereafter converge relatively quickly. Such behavior is consistent with low values of N_{max}; in the absence of explicit evidence otherwise, we choose here $N_{max} = 1$. The resulting behavior of the thresholds relative to the mean familiarity value for targets and foils is shown in Figure 2.

Using this new decision rule, we again fit the model to Heathcote's data, the only difference being that now the *initial* threshold value, rather than a constant threshold, for each accumulator was varied. The resulting best-fitting parameter values are shown in Table 3, with model fits in Figure 3 (adjusted $R^2 = 0.89$ between predicted and observed RT quantiles). It is apparent that this new mechanism, more akin to

Table 3: Best-fitting parameter values for the correlated accumulator mechanism.

Parameter	*Condition*	*Value*
$\beta^+(0)$	Accuracy	8.6
	Speed	3.2
$\beta^-(0)$	Accuracy	-16
	Speed	-9
ρ_t	Accuracy	99 samples per second
	Speed	138 samples per second
T_n	Accuracy	357 ms
	Speed	367 ms

that of a random walk model, while capable of fitting accuracy just as well as the independent accumulator model does not fit RT quantile data as well, at least when assuming that the time between samples in exponentially distributed. As with the independent accumulator model, initial thresholds are have lower absolute value under speed focus relative to accuracy focus. However, in this case, non-decision time is estimated to be roughly equal between the two conditions, with increased sampling rate in the speed focus condition, as might be expected if speed instructions encouraged greater attention to the stimulus. Although the degree of fit is poorer when using correlated accumulators, the fit is certainly not too bad, not enough to rule out this mechanism as a plausible one for recognition decisions.

Discussion

Our original work on a dynamic model for recognition memory (Cox & Shiffrin, 2012) represents one of the few attempts to link a full-fledged model of memory (in this case, an extension of the REM model; Shiffrin & Steyvers, 1997) to a decision mechanism capable of predicting both accuracy and response time. By comparing the model's predictions against entire RT distributions, we have been able to show that while the original assumptions of the model are viable, there are other possible routes to explore. These include the effect of task instruction (speed vs. accuracy focus) on sampling dynamics at test as well as a correlated counter mechanism, although a more thorough investigation of these mechanisms and the meaning of their parameters is in order.

Our model is admittedly complex, however, and incorporates many sources of variability. A correlated accumulator mechanism could produce superior predictions given different parameters for the underlying memory process. Indeed, our ongoing work in fitting our own data and data from (Starns et al., 2012) strongly suggests correlation between accumulators, even if the current data of Heathcote and Donkin (2012) do not require them. Just as task demands may influence sampling dynamics, they may also lead to different ways of balancing evidence in favor of "old" and "new" responses. Further, the dynamics of the sampling process may themselves be non-stationary, with the sampling rate changing over time (Hockley & Murdock, 1987) or different features being detected at different rates (Brockdorff & Lamberts, 2000). Finally, although the current paper is primarily exploratory, we expect that additional models of the recognition process will be developed. As part of this venture, more attention must be paid to the flexibility of such models, and we believe that exercises such as those in this paper can help illuminate the space of possible mechanisms available to modelers of recognition memory. The complexity of the models presented here is balanced by the range of data they may be expected to explain, and the present work represents just one of many forays that will be necessary to develop a general theory of recognition memory.

References

Balakrishnan, J. D., & Macdonald, J. A. (2011). Performance measures for dynamic signal detection. *Journal of Mathematical Psychology*, *55*, 290–301.

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.

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., Malmberg, K. J., & Shiffrin, R. M. (2011). Output interference in recognition memory. *Journal of Memory and Language*, *64*, 316–326.

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

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.

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.

Heathcote, A., & Donkin, C. (2012). *Recognition data.* Unpublished.

Hintzman, D. L. (1988). Judgements of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, *95*(4), 528–551.

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.

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.

Mewhort, D. J. K., & Johns, E. E. (2005). Sharpening the echo: An iterative resonance model for short-term recognition memory. *Memory*, *13*(3/4), 300–307.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, *89*(3), 609–626.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM–retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*(2), 145–166.

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.

Turner, B. M., Van Zandt, T., & Brown, S. (2011). A dynamic stimulus-driven model of signal detection. *Psychological Review*, *118*(4), 583–613.