Linking Memory Models and Decision Models: Insights on Frequency and Speed/Accuracy Trade-off

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

Abstract

Many models used to explain accuracy and response time in recognition memory have separated retrieval from decision: Retrieval produces a value of memory strength that drives a decision process characterized by evidence that does not change until a decision is reached. Cox and Shiffrin (2012) have used an alternative dynamic approach that assumes these stages interact: As time passes after presentation of a test item, more information joins a probe of memory, changing the response from memory over time; the evidence that drives a decision thus changes from moment to moment. This model is consistent with many findings in recognition memory and has been used to explain puzzling "fluency" effects (Cox, Lewis, & Shiffrin, 2013). Here, we apply the model to two large-scale studies of accuracy and response time in recognition memory, showing that the model performs comparably with existing separate-stage models while affording richer conceptual interpretations of findings concerning word frequency effects and speed/accuracy trade-offs.

Keywords: Recognition memory; memory models; response time.

Introduction

Developing an account of a cognitive process is complicated by the fact that, in addition to the process(es) of interest, participants must always make a decision about what observable outcome to generate (e.g., what button to push, what word to say, what motor trajectory to follow, etc.). Thus, a cognitive model must always consist of at least two components: a model of the cognitive process(es) needed to make a final choice, and a model for how those processes are transformed into that choice. Models of recognition memory have traditionally assumed that these components operate in successive stages. In the first stage, a test item is compared to the contents of episodic memory to produce a "familiarity" value that is then used as unvarying input to a decision process. Although the decision process may introduce additional variability, such as via a noisy accumulation process, the input it receives from the memory process is assumed to be static within a trial (e.g., Ratcliff, 1978; Brown & Heathcote, 2008).

A variety of models exist within this framework, all of which specify how the familiarity signal is generated before being used to make a decision. Typically, these models adopt a signal-detection approach wherein familiarity is compared to a criterion value and the participant calls a test item "old" if the item's familiarity exceeds this criterion. For example, TODAM (Murdock, 1982) models the familiarity signal as an inner product between a composite feature vector representing the list items and another feature vector representation of the test item. Gillund and Shiffrin (1984) posited familiarity to be the sum of study-list trace activations, where activation was determined by similarity between the test probe and traces of context and item components in both. REM (Shiffrin & Steyvers, 1997) and SLiM (McClelland & Chappell, 1998) consider familiarity to be the log-odds that the test item matches at least one of the study items. BCDMEM (Dennis & Humphreys, 2001) views familiarity as the likelihood that an item was encountered in the study context versus a different pre-experimental context.

These theories allow predictions to be made for the way experimental manipulations affect accuracy, but not response time. Response time is, however, a rich source of evidence when it comes to building cognitive models and is useful in distinguishing between different accounts of cognitive phenomena (e.g., Lamberts, Brockdorff, & Heit, 2003); models that make identical predictions about accuracy may make wildly different predictions about timing. There has been a long tradition of collecting response time in recognition memory, most often in short-term memory studies, but sometimes in long-term memory as well (e.g., Ratcliff & Murdock, 1976), including with signal-to-respond paradigms (e.g., Dosher, 1984). Most modeling of joint accuracy and RT data has assumed, as above, that memory and decision processes are separate, with familiarity acting as input to a subsequent decision process, such as a diffusion (Ratcliff, 1978), linear ballistic accumulator (LBA; Brown & Heathcote, 2008), or random walk (Hockley & Murdock, 1987; Nosofsky & Palmeri, 1997). These decision processes are dynamic, with accumulation of memory evidence to an "old" or "new" decision threshold producing both observed accuracy and response times. Diffusion and LBA models in particular have achieved great success as measurement tools, when the estimated model parameters are used to draw inferences about cognitive processes.

An alternative to the two-stage approach is to assume that the memory process is itself dynamic, with familiarity varying over time as more and different information is retrieved about the test item. In such a model, retrieval and decision operate concurrently and are inextricably intertwined. This dynamic approach to recognition has been shown to predict results qualitatively consistent with extant results in recognition memory (Cox & Shiffrin, 2012) and has offered a new perspective on dynamic presentation and short-term priming in recognition (Cox et al., 2013). In this paper, we investigate whether this dynamic approach can *quantitatively* fit recognition data as well as two-stage models. More importantly, we ask whether this approach provides new insights into the processes of recognition memory. We therefore apply this model to two datasets, one by Rae, Heathcote, Donkin, Averell, and Brown (in press) and another by Starns, Ratcliff, and McKoon (2012), looking especially at the effects of word frequency and speed- accuracy trade-off in recognition.

A Dynamic Approach to Recognition

The dynamic approach to recognition taken in this paper was first introduced in Cox and Shiffrin (2012), itself an outgrowth of previous models that elucidate the dynamics of memory access and decision making (Brockdorff & Lamberts, 2000; Diller, Nobel, & Shiffrin, 2001; Nosofsky & Palmeri, 1997). In this approach, the decision process leading to a final recognition decision is based not on a constant rate of evidence accumulation, but a time-varying interaction of retrieval of features from knowledge, growth of an episodic memory probe, retrieval that varies as the probe varies, and evidence accumulation that varies with what is retrieved. When a test stimulus is presented, its physical form plus context are used to access knowledge (for a word, this would be a lexical trace in semantic memory). Features are extracted from knowledge gradually over time and are added to the features of the current context to form an accumulating probe in short-term memory. At each epoch of time, the probe is compared in parallel to traces in episodic memory, producing similarity-based activations of each trace characterized as likelihood ratios. The average likelihood ratio at time t is termed the familiarity at time t. Familiarity changes over time as the probe changes; changes in familiarity are accumulated as time passes, with a response occurring when the accumulated change reaches an "old" or "new" boundary.

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

Representation and Encoding

Study events lead to the formation of separation traces in episodic memory. Traces (as well as the memory probe) consist of a vector of feature values, each value being missing or having one of two values. Features can arise from the context of an event (including information about the time, location, and internal state of the participant) or the content of an event (for words, these include semantic, orthographic, and phonological information). We assume that there are a fixed number N_X of context features, each with a stored value (justified partly by the fact that study times in the present application were always at least one second; Malmberg & Shiffrin, 2005). We also assume that there are N_C content features, stored with probability u. We assume there is noise at encoding, such that a feature value is stored correctly with probability c_S and with probability $1 - c_S$ its value is sampled at randomly from the value base-rates for that feature.

Probe Formation

At the beginning of a test trial, before any stimulus is presented, the probe consists only of the N_X context features. At some time after the presentation of the test stimulus, content features begin to be sampled from knowledge. Each new value is added to the previously sampled feature values, producing a probe that grows as time passes. Sampling is a homogeneous Poisson process, where any of the N_C content features may be sampled at each time step (when a feature is re-sampled nothing changes and the sampling process continues). Specifically, the average proportion of content features sampled grows exponentially with time:

$$v(t) = 1 - (1 - 1/N_C)^t . \tag{1}$$

In most situations, we assume that the correct value is sampled for each feature. We will, however, consider model variants in which the correct value is stored in the probe with probability c_T and with probability $1 - c_T$ a value is sampled at random from the base-rates for that feature.

Likelihood

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

The comparison between the probe and a given trace i results in a likelihood ratio $\lambda_i(t)$, reflecting the relative probability that the probe and trace encode the same item versus the probability that they encode different items. To determine these probabilities, we need only the number of features that match between the probe and trace and the number that mismatch. If a value is not stored in the trace or has not yet been sampled into the probe, that feature provides no evidence[1]. The probability of a feature match between the probe

---

[1] In cases where items are of different types and may be represented with qualitatively different features, a missing feature in either the probe or the trace can, in fact, be evidence that the probe and trace encode different items (Cox & Shiffrin, 2012).

and trace, given that they encode the same item, is

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

which simplifies to $p_{M|S} = c_S + (1-c_S)g$ when $c_T = 1$. Here, g denotes the probability of a match occurring by chance. If feature values are equiprobable, $g = \frac{1}{2}$. However, if features differ in their base-rate probability, then an unlikely value is less likely to be sampled by chance, so $g < \frac{1}{2}$; if the matching feature value is more common, $g > \frac{1}{2}$. The probability of a match if the probe and trace encode different items is just this base-rate probability: $p_{M|D} = g$. The probability of a mismatch given that probe and trace encode the same item is

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

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 $p_{N|D} = (1-g)$. The likelihood ratio for a single trace is the product of the likelihood ratios for each of the $N_M(t)$ matching and $N_N(t)$ mismatching features at time t:

$$\lambda_i(t) = \left(p_{M|S}/p_{M|D}\right)^{N_M(t)} \left(p_{N|S}/p_{N|D}\right)^{N_N(t)}. \qquad (2)$$

Activation and Familiarity

Familiarity, denoted $\phi(t)$, is the average likelihood ratio among those traces that are active at time t. A trace is considered "active" if its likelihood ratio is greater than one:

$$\phi(t) = \langle \lambda_i(t) : \lambda_i(t) > 1 \rangle. \qquad (3)$$

Initially, when the probe consists of context features only, all list traces and few history traces tend to be activated. As time passes, non-target list traces tend to deactivate and history traces tend to activate.

Making a Decision

A decision is based on the accumulated changes in log-familiarity as successive samples are taken, equivalent to the difference in log-familiarity between the current time t and the start time 0 (since $\sum_{\tau=0}^{t-1}\left[\log\phi(\tau+1) - \log\phi(\tau)\right] = \log\phi(t) - \log\phi(0)$). As features are sampled randomly, some matching some traces and some matching others, familiarity takes a trajectory that is describable as a time-varying random walk. Mean familiarity trajectories for different item types are depicted in Figure 1. Familiarity for targets tends to decrease at first, since the first few content features tend not to match most traces. Familiarity for targets then tends to rise as the many matching content features produce a high likelihood ratio for the target trace. Foil trajectories continue to drop, unless well-matching history traces are activated, in which case foil familiarity might rise again. In all cases, however, the probe will eventually saturate with features, causing familiarity to reach a stable asymptote.

Collapsing Boundaries Response boundaries that are constant with time might result in some trajectories failing to result in a decision if their asymptotic values fell between the two bounds. There are various ways to address this issue, but the most straightforward is to assume that the "old" and "new" decision boundaries collapse over time as a function of probe completeness (Equation 1)[2]:

$$r(t) = 1 - \left[1 - (1 - 1/N_C^*)^t\right]^2. \qquad (4)$$

The boundaries begin at time 0 a certain distance apart, A_0, with boundary separation $A(t)$ diminishing over time: $A(t) = A_0 r(t)$. The "old" boundary at time t is $(1-b)A(t)$ and the "new" boundary is $-bA(t)$. Thus, boundaries are symmetric if $b = \frac{1}{2}$, "old"-biased if $b > \frac{1}{2}$, and "new"-biased if $b < \frac{1}{2}$. This situation is depicted in Figure 1.

The rate of boundary collapse is governed by the probability of sampling any one feature at a given time, $\frac{1}{N_C}$, where N_C is the number of available content features. A decision maker may, particularly in conditions emphasizing speed, choose to make do with fewer content features. In this case, boundaries would collapse more quickly, and so the value N_C may be replaced by a different value N_C^* reflecting the number of content features the decision maker is willing to consider.

Relationship to Real Time When familiarity hits either the "old" or "new" boundary, the corresponding response is made. The number of time-steps t needed to reach that boundary governs the response time. We assume a simple linear mapping between model time-steps and "real" time $\hat{t}$:

$$\hat{t} = T_R + \rho t \qquad (5)$$

where ρ is the time taken per time-step of the model and T_R is a residual time that includes the time taken to detect the stimulus and execute the appropriate motor response. We simplify and assume T_R and ρ are constant; at present, this suffices to allow the model to fit data reasonably well and provide useful insights into underlying cognitive processes.

Model Fitting

To fit the above model to the following two datasets, we first developed an analytic approximation to the model that enables fast computation of predictions for response probabilities and RT distributions. Space does not permit this approximation to be described here in detail; suffice it to say that the full model (which requires many thousands of simulations) produces predictions that are nearly identical to those of the approximation, given the same parameters.

We fit the model to group data using quantile maximum likelihood (QML; Heathcote, Brown, & Mewhort, 2002), finding the maximum with the Nelder-Mead simplex algorithm (several start points were used to avoid local maxima).

---

[2] Quantitatively, exponents nearly equal to 2 best fit the available data, and prevent boundaries from collapsing too quickly at the beginning of the trial.

The data to be fit consisted of a set of bins for each condition containing the number of responses of each type (e.g., hit, false alarm, etc.), that were observed in the RT quantile ranges: 0-0.1, 0.1-0.3, 0.3-0.5, 0.5-0.7, 0.7-0.9, 0.9-1. Fitting via QML enables us to jointly fit accuracy and RT while using the Akaike and Bayesian Information Criteria (AIC, BIC) for model comparison. By comparing different parameterizations of our model, we gain additional insight into the mechanisms involved in recognition; by comparing our model with standard diffusion models, we see where the dynamic approach might diverge from a traditional two-stage approach.

Rae et al. (in press)

Rae et al. (in press) report results from a recognition memory experiment with 47 participants and a total of 18,001 trials; participants studied lists containing 56 words, of which 50 appeared as targets on a subsequent test along with 50 foils. For each study/test block, participants were instructed to emphasize either accuracy or speed in their responses. Although Rae et al. (in press) did not analyze and report effects of normative word frequency, we wanted to facilitate comparison with the experiment by Starns et al. (2012), and therefore divided their results based on a median split in word frequency (using the MRC database; Coltheart, 1981). For further details, the reader is directed to their paper.

Frequency High- and low-frequency (HF and LF) words can differ along many dimensions, two of which depend directly on frequency: HF words occur more often in life experience, and HF words are *comprised of features* that occur more often in life experience (Malmberg, Steyvers, Stevens, & Shiffrin, 2002). Within our model, these two aspects of frequency are captured by two parameters: K, the number of traces of an item stored from prior life history; and g, the base-rate of a random feature match. The effects of frequency might be due to a higher value for K for HF words than LF words ($K_L < K_H$), a higher value for g for HF that LF words ($g_L < g_H$), or both. In fitting different numbers of history traces, we allowed K_L and K_H to vary freely (they were not constrained to be ordered). To account for potential differences in feature frequency, we fit the single parameter g_L and let $g_H = 1 - g_L$.

As evidenced by both AIC and BIC (Table 1, rows 1–3), and by the optimization routine hitting the minimum K values of 1 for both HF and LF items, the model (#2) is preferred where history traces play no role ($K_L = K_H = 0$) and frequency effects are accounted for solely in terms of different base-rates between HF and LF items ($g_L < g_H$). This suggests that feature frequency, a factor based on traces stored in knowledge, plays a more important role than the number of historical occurrences, which is based on traces in episodic memory. Potential implications of this result are left for the general discussion.

Speed-Accuracy Trade-offs Rae et al. (in press) assessed the possibility that instructions to focus on speed, rather than accuracy, affect not just response boundaries, but decision evidence as well. Starns et al. (2012) argue that a focus on response speed may impair participants' ability to form effective retrieval cues. In our model, this concept can be embodied in two ways. Speed instructions may 1) cause sampling of test item features to be less accurate, represented by a value of $c_T < 1$; 2) cause a more rapid collapse of response boundaries, represented as a lower value of N_C^* in Equation 4. If boundaries collapse faster, decisions will tend to be made on the basis of fewer features, before the evidence has matured.

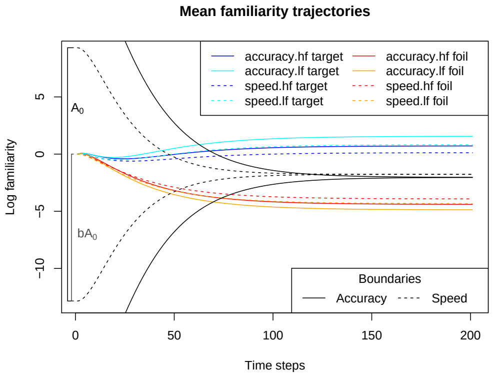

Figure 1: Mean familiarity trajectories and response boundaries for the preferred model fit (#8) to the data from Rae et al. (in press). The initial boundary separation (A_0) and bias (b) parameters are labeled for the speed condition.

As shown in rows 4-8 of Table 1 the preferred model (#8) allows both c_T and N^* to vary. This is in accord with Rae et al. (in press) and Starns et al. (2012): participants under speed instructions make decisions both with lower *quality* evidence—reflected in decreased boundary separation and increased sampling noise c_T—and lower *quantity* of evidence—reflected in faster boundary collapse. This situation is depicted graphically in Figure 1.

Quality of Fit We fit a diffusion model to these data using the above QML method, allowing the mean and variance of the drift distribution to vary freely as a function of word frequency and speed/accuracy focus; bias, boundary separation, and residual time each varied with speed/accuracy focus (18 total free parameters). Because our model does not incorporate variability in start point or residual time, we did not allow them to vary in the diffusion model either (equivalent to the model originally introduced by Ratcliff, 1978). The diffusion model could only achieve a log-likelihood of -1186 (AIC=2407, BIC=2548), worse than any version of our model (our worst fit, model 1, has a log-likelihood of -828), despite having more parameters than any version of our model. Implications are discussed below; for the moment, we argue that our dynamic model of recognition provides both a good quantitative and conceptual account of these data.

Starns et al. (2012)

In the experiment reported by Starns et al. (2012), four participants each completed 20 sessions of a recognition memory task for a total of 66,986 trials. Within each study/test block, study words were shown 1, 2, or 4 times and words at both

Table 1: Best-fitting parameter values and fit criteria for the data from Rae et al. (in press). The number of content and context features were fixed for all models at $N_C = N_X = 30$ and $c_T = 1$ for all accuracy conditions. Other values fixed for each model are marked with an asterisk. Models 1-3 only vary frequency-related parameters; the preferred model of the first three is indicated by italics. Models 4-8 vary parameters related to speed/accuracy trade-off; the overall preferred model is indicated in boldface.

						Accuracy					Speed							
	u	c_S	K_L	K_H	g_L	A_0	b	T_R	ρ	N_C^*	A_0	b	T_R	ρ	c_T	N_C^*	AIC	BIC
1	0.380	0.930	1	1	0.5∗	51.2	0.539	243	5.54	30∗	17.4	0.602	325	3.53	1∗	30∗	1680	1774
2	0.368	0.937	0∗	0∗	0.486	46.6	0.541	258	5.56	30∗	17.6	0.597	327	3.48	1∗	30∗	*1628*	*1714*
3	0.366	0.936	1	1	0.481	46.6	0.542	257	5.58	30∗	18.0	0.595	326	3.43	1∗	30∗	1646	1747
4	0.379	0.940	0∗	0∗	0.477	44.3	0.545	264	5.71	30∗	17.1	0.609	321	3.41	0.959	30∗	1541	1635
5	0.387	0.934	0∗	0∗	0.478	44.4	0.546	275	5.50	30∗	20.4	0.579	287	5.79	1∗	19.5	1397	1490
6	0.376	0.943	0∗	0∗	0.476	42.1	0.547	283	5.52	30∗	16.4	0.607	314	4.30	0.967	24.8	1453	1554
7	0.413	0.927	0∗	0∗	0.478	38.4	0.554	271	7.12	24.6	22.2	0.577	268	6.72	1∗	17.3	1334	1435
8	0.406	0.937	0∗	0∗	0.477	40.4	0.550	264	7.30	24.2	22.4	0.579	267	6.00	0.970	18.9	**1302**	**1411**

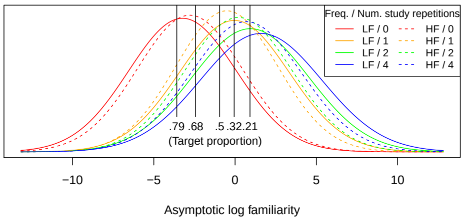

Figure 2: Asymptotic familiarity distributions and decision bounds (black) produced by model 7 of the data from Starns et al. (2012).

study and test could be of either high or low frequency. Between blocks, participants were instructed to focus on either speed or accuracy. Unlike the Rae et al. (in press) study, test lists also varied in the proportion of targets, either .21, .32, .5, .68, or .79, thus inducing a bias (participants were informed about this proportion prior to testing). The reader is directed to their paper for further detail.

We allowed u (the probability of encoding a content feature at study) to vary with number of repetitions; across all model fits (Table 2), u increases with the number of repetitions. In all model fits, b (not shown in Table 2, for space) strictly increases with the proportion of targets, as in Figure 2 which shows the asymptotic boundaries ($A_0\left(\frac{1}{2} - b\right)$) for model 7. As above, the best account for word frequency effects assumes that HF/LF words differ in feature-frequency base rates rather than the number of episodic history traces (rows 1-3 of Table 2). Finally, speed instructions are found to affect both boundary separation and the rate of boundary collapse, but allowing c_T to vary did not provide a substantially better fit to these data (rows 4-8 of Table 2).

Target/Foil Variance A main argument in Starns et al. (2012)—expanded in Starns and Ratcliff (2014)—is that the distribution of target evidence has greater variance than that for foils. Although the evidence distributions in our model are non-stationary, the asymptotic familiarity distributions predicted by the preferred model (#7; Figure 2) show greater target than foil variance in all but one case: HF targets that appear only once at study. In that case, the very low u value means that the single matching target trace cannot take many values even when the probe is saturated, thereby decreasing the variance of the familiarity distribution when this single target is averaged with the likelihoods of the many non-matching list traces (which have higher variance as their feature values bear only a chance resemblance to the target). Given the small size of this effect, we do not make a strong theoretical claim about it, but note that the ratio of target to foil variance is often estimated to be lower for weak HF words (Starns et al., 2012; Starns & Ratcliff, 2014).

Quality of Fit The best-fitting diffusion model reported by Starns et al. (2012) achieved a BIC of 7886, considerably lower than our best BIC of 12261 (Table 2, row 7). However, Starns et al. (2012) incorporated several additional sources of variability, including freely varying evidence parameters across conditions and variability in start point, drift criterion, and residual time (here denoted T_R). When start-point and drift criterion variability are removed, the diffusion model achieves a still-superior BIC of 10219 (56 free parameters); however, when only residual time variability is removed, the fit of the diffusion model is drastically reduced, yielding a BIC of 21028 (63 free parameters, with constant T_R allowed to differ between each focus/bias condition, as in our model). Thus, it would appear that the fit of the diffusion model is primarily due, not to the model's evidence accumulation process, but to the considerable variability assumed in non-decision time. Indeed, our model can be seen as an explicit model of the perceptual processes (e.g., feature sampling) that are considered part of the two-stage model's "non-decision" time, explaining why our model provides a better account even when T_R is assumed constant.

General Discussion

Our prior research showed how use of a dynamic model could explain in qualitative fashion findings such as fluency that lay outside the standard approaches for modeling recognition memory (Cox et al., 2013). In this article, we fit the dynamic model quantitatively to accuracy and response time distribution data from two studies with rich datasets. The dynamic model fits these data sets well qualitatively, and fits the Rae et al. (in press) dataset at least as well quantitatively as the standard two-stage approach. A standard diffusion model achieves a superior quantitative fit to the Starns et al. (2012) dataset only by allowing residual time to be a random variable. Just as our dynamic model seeks to "unpack" processes, like the encoding of the memory probe, that remained obscure in previous two-stage approaches, this finding suggests that there are many other sources of variability in recognition to explore.

Our modeling results contribute to the growing body of evidence that speed instructions do not have a selective influence on response caution, but can affect evidence processes as well (Rae et al., in press). In both cases investigated here,

Table 2: Best-fitting parameter values and fit criteria for the data from Starns et al. (2012). $N_C = N_X = 30$ for all models and $c_T = 1$ for all accuracy conditions. Other values fixed for each model are marked with an asterisk. $\bar{A}_0$, $\bar{T}_R$, and $\bar{\rho}$ are averaged across the five bias conditions; c_T and N_C^* are equal across bias conditions. Models 1-3 only vary frequency-related parameters; the preferred model of the first three is indicated by italics. Models 4-8 vary parameters related to speed/accuracy trade-off; the preferred model is indicated in boldface.

								Accuracy				Speed						
	u_1	u_2	u_4	c_S	K_L	K_H	g_L	$\bar{A}_0$	$\bar{T}_R$	$\bar{\rho}$	N_C^*	$\bar{A}_0$	$\bar{T}_R$	$\bar{\rho}$	c_T	N_C^*	AIC	BIC
1	0.304	0.373	0.408	0.936	174	208	0.5*	22.6	277	4.90	30*	15.2	283	4.07	1*	30*	12703	13122
2	0.300	0.371	0.409	0.937	0*	0*	0.468	22.2	277	4.99	30*	15.3	283	4.10	1*	30*	*12525*	*12935*
3	0.295	0.365	0.402	0.940	172	211	0.468	22.5	283	4.88	30*	15.3	287	4.03	1*	30*	12930	13358
4	0.299	0.366	0.408	0.940	0*	0*	0.468	21.6	283	5.00	30*	15.1	264	4.02	0.943	30*	12728	13147
5	0.302	0.372	0.414	0.937	0*	0*	0.470	22.1	279	5.00	30*	15.5	248	6.32	1*	21.6	11984	12403
6	0.302	0.371	0.408	0.941	0*	0*	0.472	22.2	279	5.00	30*	15.9	222	7.00	0.958	19.9	12089	12518
7	0.304	0.379	0.425	0.940	0*	0*	0.470	19.9	297	5.15	29.5	15.8	249	6.64	1*	20.9	**11833**	**12261**
8	0.302	0.371	0.412	0.944	0*	0*	0.471	22.0	279	5.23	28.9	16.3	225	6.33	0.950	21.4	12143	12580

speed instructions induce participants to make use of fewer features when making recognition decisions, as manifested in an increased rate of boundary collapse. In at least one experiment (Rae et al., in press), speed instructions also result in an impairment of perceptual processes, increasing noise in the memory probe. A focus on response speed might divert attentional resources that would otherwise be used to maintain an accurate representation of the memory probe in STM. It may also be that sampling noise itself changes over time, such that probe features are more likely to be sampled incorrectly early on and take time to "resolve" to their correct values; such an effect would be apparent early in a trial, when speeded responses are more likely to be made.

An interesting possibility raised by our modeling concerns the source of word frequency effects. In both studies, such effects were better predicted by variations in feature frequency, a factor stored in lexical knowledge, than by numbers of event traces, a factor stored in episodic memory (cf., Malmberg et al., 2002). According to the dynamic model, history traces only match in content, not context, so they become active only when sufficient content features join the memory probe. This occurs at a later point in the decision process; if delayed activation of history traces does occur, and history traces are indeed the main source of word frequency effects, then such effects should be reduced in speed conditions because there is less time to activate the interfering traces. This is not the case in the data modeled here, providing additional evidence that word frequency effects are more strongly determined by feature frequency than number of history traces.

The result bears on the question of how much interference in recognition is due to "context noise" versus "item noise" (Dennis & Humphreys, 2001; Criss & Shiffrin, 2004). The c_S parameter in all our model fits is high (around 0.94), indicating that items tend to be quite discriminable, hence allowing only a small role for item noise. We argue that a feature-frequency account of frequency effects is itself a form of context noise. This owes to the fact that feature base-rates—which makes LF items more discriminable—must be learned over time *in contexts prior to that of the experiment*. It is just that the relevant preexperimental experience resides in general knowledge (e.g., semantic memory) rather than episodic memory, with this knowledge subsequently exploited during episodic recognition and other memory and perception tasks (Nelson & Shiffrin, 2013).

Finally, we remark that these new insights into the mechanisms of recognition memory were made possible by use of a dynamic modeling approach in which retrieval and decisions are integrated mechanisms, rather than separated successive stages. The model makes explicit links between the underlying processes that determine storage, retrieval, and decision making, extending prior modeling approaches and suggesting new interpretations of extant results.

References

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, S., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*, 153–178.

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology*, *33A*(4), 497–505.

Cox, G. E., Lewis, N. J., & Shiffrin, R. M. (2013). On the dynamics of information accumulation in recognition. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual conference of the Cognitive Science Society* (pp. 346–351). Austin, TX: Cognitive Science Society.

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., & Shiffrin, R. M. (2004). Context noise and item noise jointly determine recognition memory: A comment on Dennis and Humphreys (2001). *Psychological Review*, *111*(3), 800–807.

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.

Dosher, B. A. (1984). Discriminating preexperimental (semantic) from learned (episodic) associations: A speed-accuracy study. *Cognitive Psychology*, *16*, 519–555.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*(1), 1–67.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2002). Quantile maximum likelihood estimation of response time distributions. *Psychonomic Bulletin & Review*, *9*(2), 394–401.

Hockley, W. E., & Murdock, B. B. (1987). A decision model for accuracy and response latency in recognition memory. *Psychological Review*, *94*(3), 341–358.

Lamberts, K., Brockdorff, N., & Heit, E. (2003). Feature-sampling and random-walk models of individual-stimulus recognition. *Journal of Experimental Psychology: General*, *132*(3), 351–378.

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.

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.

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

Nelson, A. B., & Shiffrin, R. M. (2013). The co-evolution of knowledge and event memory. *Psychological Review*, *120*(2), 356–394.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*(2), 266–300.

Rae, B., Heathcote, A., Donkin, C., Averell, L., & Brown, S. (in press). The hare and the tortoise: Emphasizing speed can change the evidence used to make decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

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

Ratcliff, R., & Murdock, B. B. (1976). Retrieval processes in recognition memory. *Psychological Review*, *83*(3), 190–214.

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. (2014). Validating the unequal-variance assumption in recognition memory using response time distributions instead of ROC functions: A diffusion model analysis. *Journal of Memory and Language*, *70*, 36–52.

Starns, J. J., Ratcliff, R., & McKoon, G. (2012). Evaluating the unequal-variance and dual-process explanations of zROC slopes with response time data and the diffusion model. *Cognitive Psychology*, *2012*(1), 1–34.