Title

Object Recognition when Features Arrive Dynamically

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 40(0)

Authors
Harding, Samuel M
Shiffrin, Richard M

Publication Date
2018

Object Recognition when Features Arrive Dynamically

Samuel M. Harding (hardinsm@indiana.edu)

Richard M. Shiffrin (shiffrin@indiana.edu)
Department of Psychological and Brain Sciences; Cognitive Science Program
Indiana University, Bloomington, IN, 47405

Abstract

We report a model for object identification based on an experiment that varies the arrival times of different features of the objects. A single object, a circle with four spokes extending in different directions, is presented and must be classified as either one of four well trained target stimuli, or one of four well trained foil stimuli. The features (spokes) are presented either simultaneously or successively at intervals of 16, 33, or 50 ms., with target diagnostic features arriving first or last. All durations are short enough that the display appears simultaneous. The data show that individual decisions vary with both timing and diagnosticity. We apply a dynamic model based on one reported in (Cox & Shiffrin, 2017) for episodic recognition memory. Our model assumes features are perceived at varying times following presentation, possibly in error. At each moment the current features are compared to the well learned memory representations of the eight stimuli, producing a likelihood ratio for target vs foil. A decision is made when the log likelihood first exceeds a target decision boundary or falls below a foil decision boundary. The model implements a form of Bayesian optimal decision making given the assumptions concerning feature perception. It predicts the key findings quite well.

Keywords: response time modeling; dynamic stimuli; visual search; object recognition; feature sampling

Introduction

The time course and outcome of recognition decisions are frequently used to inform our understanding of perceptual, memory, and preference judgments. Many successful models of classification and recognition treat the choice and its time as the outcome of a dynamic process that samples and integrates stimulus information over time. For example, the drift diffusion (Ratcliff & Rouder, 1998) and Linear Ballistic Accumulator (LBA) (Brown & Heathcote, 2008) models successfully capture response patterns to static stimuli by continually sampling evidence at a constant rate, throughout the time that a single decision is being made. It is of considerable theoretical interest to explore the way these kinds of decisions vary when stimuli change over the course of a single trial, but few studies have explored this domain. In this report we model data from a study that manipulates the timing and amount of evidence available within each decision trial.

A wealth of existing models have explored how decision processes vary over time according to the influences of *internal* factors. For example primacy and recency effects in perceptual and preference judgments are thought to arise due to lateral inhibition between response options (Usher & McClelland, 2001) or by applying differential weight to evidence arriving early versus late during deliberation (Busemeyer & Townsend, 1993). Shifting attention to consider different aspects of response alternatives can lead to reversals in preference within the course of each trial (Diederich & Oswald, 2014). When responses are made under time pressure, choices made with only partial information may be reversed when given sufficient time to sample all relevant features (Lamberts, 1995; Cohen & Nosofsky, 2003). Such feature sampling may also account for response reversals in associative recognition, as initially context-based retrieval is bolstered by item, and finally associative features over time (Cox & Criss, 2017).

Less work, however, has focused on capturing how *external* changes in stimuli affect the decision process. Lamberts and Freeman (1999) asked participants to categorize items comprised of several discrete features; on some trials, the entire item was shown, while other trials presented single features in isolation. They found that responses to individual features could be used to predict those made to the entire object, suggesting that categorization decisions are made, in part, by integrating information across constituent parts. In another study, Holmes, Trueblood, and Heathcote (2016) changed the direction of coherent motion of a cloud of randomly moving dots, before a left/right motion discrimination response was made. Using a version of the LBA model with two distinct rates of processing the authors found evidence that participants noticed the change after a short delay and adjusted their decision process in light of the new information. Together, these studies highlight the novel insights provided by dynamically modifying stimuli during the course of an ongoing decision decision.

We present here an alternative dynamic model for object perception and categorization, based on a paradigm that changes the timing and order of the arrival of features of varying diagnosticity during a single decision. The data modeled are a subset of a larger experiment (other conditions are reported in Cousineau and Shiffrin (2004) and in Cousineau, Donkin, and Dumesnil (2015)). The task is one of visual search for well practiced targets and foils that do not change roles over 58 sessions of training. Some conditions in the study presented objects sequentially and others presented features of those objects sequentially. We apply our model to accuracy and reaction time data for three subjects who provided sufficient data in the feature sequential conditions that presented a single target or foil for a binary target-foil decision.

The model is an extension of one recently proposed to account for recognition decisions via perceptual sampling of features during storage and retrieval (Cox & Shiffrin, 2017). The present model compares the relative evidence in favor of either response, at each moment, derived from matching the perceptually sampled features against well-learned representations of the eight objects stored in memory. The response given, and its time, are determined by the point in time at which the evidence passes one of two decision boundaries.

Method

The three participants carried out visual search for 58 sessions approximately an hour in length, over several weeks. Four stimuli were defined as 'targets', four others were defined as 'foils' and the stimuli maintained these roles over the course of the experiment. Following the terminology of (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977) such training is termed consistent mapping, or CM. Each trial in the conditions we model presented a single stimulus to be classified as a 'target' or 'foil'. The four features of this test object appeared simultaneously, or sequentially with 16, 33, or 50 ms between each onset. This timing was chosen because even the slowest presentation rate was fast enough that the displays appeared simultaneous (albeit a bit 'flickery') making it unlikely that strategies would differ with presentation rate. Finally, the order of the sequentially presented features varied in diagnosticity, with the most diagnostic features presented either first or last (this is described in more detail below).

Participants

The three participants gave informed consent in accord with the Indiana University IRB. Monetary compensation was provided, and the participants were instructed to respond as quickly as possible without exceeding 5% error rate.

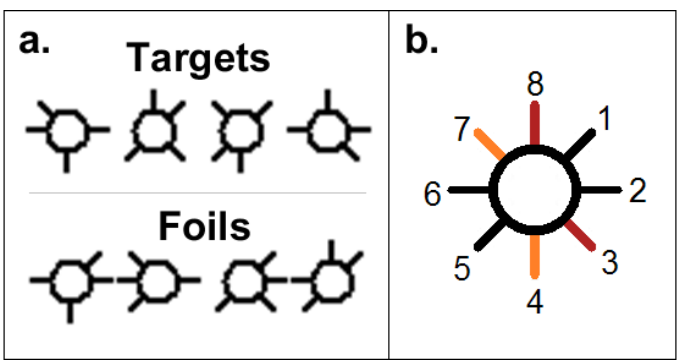

Figure 1: (a) Stimuli used throughout the course of the experiment were drawn from this set of eight objects, which were consistently mapped (CM) to either the "Target" or "Foil" category. (b) Each object was formed with four features (spokes) extending radially from a central black circle. *Target-diagnostic* features are highlighted; these pairs of objects only ever appeared, together, on items in the "Target" category.

Stimuli

A set of 8 novel objects was employed throughout the experiment. Stimuli were black circles with lines (spokes) extending outward from four out of eight potential locations per object (see Figure 1). There were four 'target' objects, and four 'foil' objects. No single feature distinguished targets from foils, but a particular two features could identify two of the targets uniquely, and a different set of two features could identify the other two targets uniquely. Generally, feature combinations differed in the degree to which they provided diagnostic evidence concerning a target versus foil decision.

Procedure

Displays alternated between 1, 2, and 4 objects on any given trial, but the interest here is in the single object test displays, which occurred on a random one third of the trials. At each of the three sequential presentation rates, the order in which features appeared varied by diagnosticity: for targets the two diagnostic features appeared either first or last; for foils, only one of the target-diagnostic features was shown, and it was either the first or the last to appear. Each trial began with the presentation of a fixation cursor in the center of the screen for 1000ms, which was followed by the appearance of a featureless circle for 500ms in the location of the subsequent test object. The features were then added one-by-one and remained visible until the subject made a response.

Results

The results of the experiment are broken down by accuracy and median response time in Figures 2 and 3. Of primary interest are the patterns of results when using sequential presentation of features, when compared to the simultaneous presentation condition (the data shown as grey squares).

Accuracy

When a target object was shown, accuracy stayed near ceiling if early-arriving features were strongly diagnostic of targets (diagnostic-first); however, when the first two features were instead diagnostic of foils, subjects were less likely to identify the object as a target. This pattern was reversed when the presented object was a foil: if the first feature to appear was highly-diagnostic of a target, participants tended to report that it was a target, while a late-appearing target feature did little to decrease performance[1]. Use of a generalized linear, mixed-effects model showed a significant two-way interaction that indicated that the number of correct responses was significantly related to the order of target-diagnostic feature presentation (first versus last) and the identity of the object (target versus foil), $F(1{,}2148) = 138.600$, $p < .001$[2].

The amount of delay between features played a role only when early features provided deleterious information, as the presence of early, useful information did little to improve the already near-ceiling performance. This was revealed via a 3-way interaction of including delay, object identity, and feature

[1]Note that foil objects only included one target-diagnostic feature.

[2]Binomial (link = logit) model, fit using MATLAB function, *fitglme* with Fixed Effects for *Order* (First, Last), *ISI* (16, 33, 50ms) and *Object Identity* (Target, Foil), and Random Effects for *Subject*.

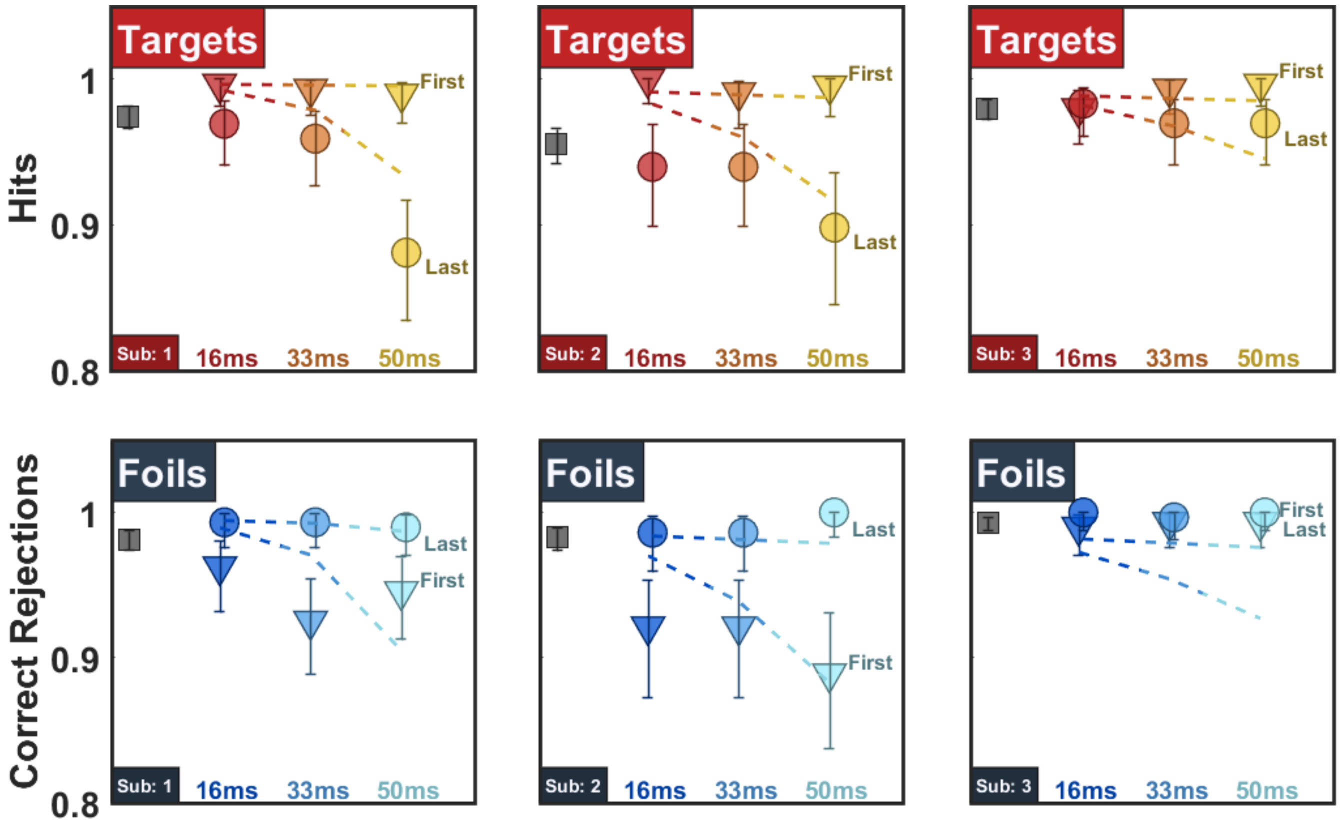

Figure 2: Mean accuracy data (shapes) and qualitative model fits (lines) for subjects (S1-S3) separated into 'Hits' (respond "Target" to targets; top row) 'and Correct Rejections' (respond "Foil" to foils; bottom row). Accuracy in the simultaneous condition is shown for reference (grey square), and each of the delay conditions is presented in the colored shapes (dark-to-light: 16ms, 33ms, 50ms). Target-diagnostic features appeared were the first (triangles) or last (circles) to appear.

order, $F(2,2148) = 3.4716$, $p < .05$. These results suggest that subjects began to accumulate evidence towards a decision before all of the features were presented, and that later-arriving features did not entirely mitigate this effect.

Response Times

Median response times were analyzed using a similar generalized linear mixed-effects model. Overall, response times were slowed when the features arrived sequentially. This makes sense, given that the amount of information available to the decision process was limited by the number of available features; this was supported by a significant main effect of ISI, $F(2,2148) = 42.015$, $p < .001$. The order of feature appearance was also important: response times were faster when early-arriving features aligned with the eventual object identity. For example, target-diagnostic features arriving first on a target object led to faster "target" responses than when these crucial features were the last to appear. Again, this pattern was reversed for foil objects. The 2-way interaction between stimulus and order was significant, $F(1,2148) = 36.982$, $p < .001$.

It is difficult to draw strong conclusions from the accuracy and response time data when analyzed separately, but these patterns suggest that the decision process operated continuously as features were sampled, and that early-arriving features tended to bias the eventual response. We therefore turn to a dynamic process model to account for both response times and accuracy as they vary with timing and order of appearance of diagnostic features.

The Model

The proposed model is based on one proposed to account for accuracy and response data in episodic recognition memory and reported in (Cox & Shiffrin, 2017). The model captures variability in response times and response proportions via a feature-sampling process that unfolds stochastically over time: as time passes following presentation of the test stimulus, the subject extracts features, which are entered into a probe and used to search long-term memory. At each moment this comparison yields a relative likelihood that the test stimulus matches objects from 'target' and ''foil' categories. A response is generated when the relative likelihood exceeds a "target" response criterion or drops below a "foil" response criterion. A key to the model's predictions is the differential arrival times of different feature types into the probe.

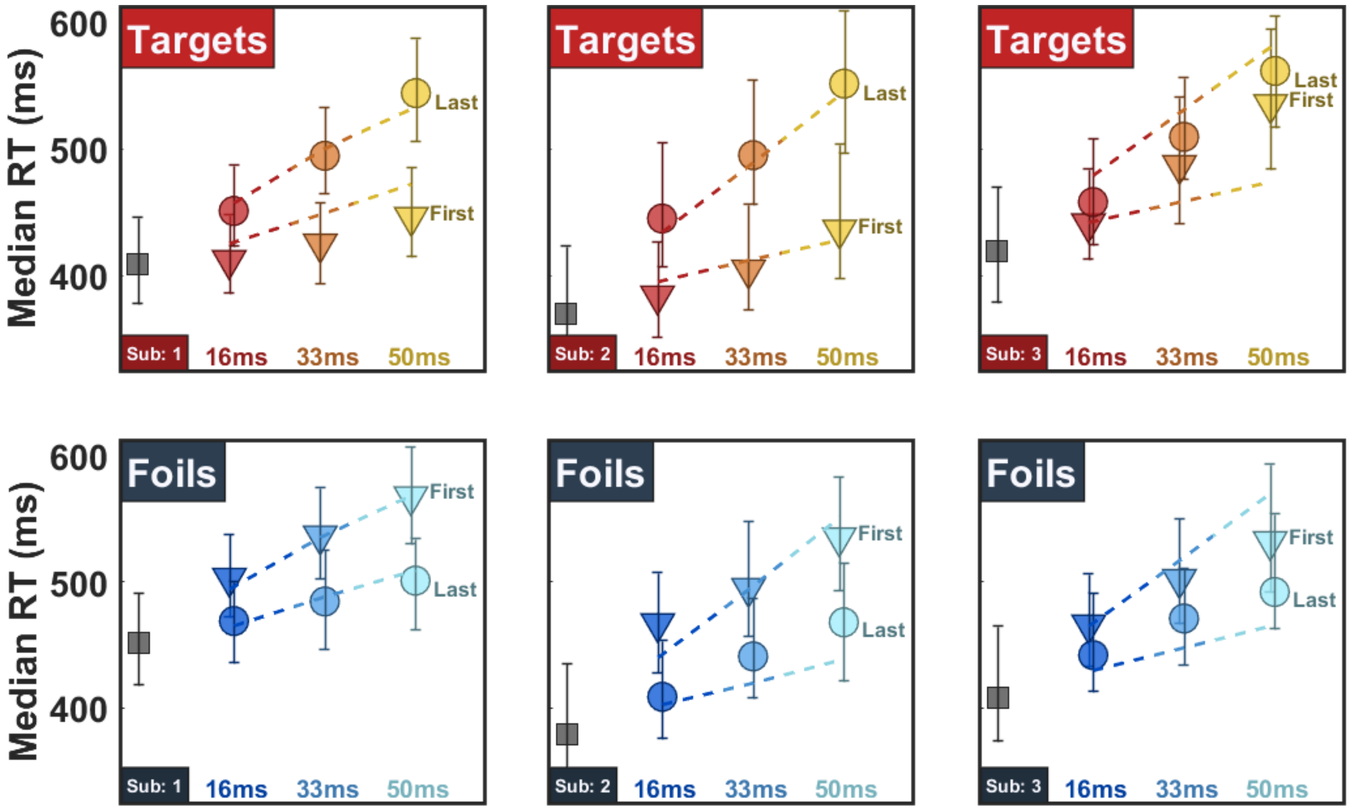

Figure 3: Median empirical (shapes) and predicted (lines) response times for correct responses made to Targets (top row) and Foils (bottom row). Responses were longer when the amount of delay between features was increased (timing conditions: 16, 33, or 50ms). When informative information appeared early (Targets, diagnostic-first; Foils, diagnostic-last), responses were faster than when this information was withheld until later in the trial. Error bars show the 25th and 75th quantile of the response time distributions.

Feature Sampling

The present model treats stimuli as comprised of discrete *features*, assumed to be the eight spokes of the stimuli. The features are not perceived immediately upon presentation, but are sampled probabilistically over time. Note that both the presence and absence of features provide information about the identity of the object: the presence of single diagnostic feature merely suggests a "Target" response. When paired with a second diagnostic feature, the two provide definitive evidence; the absence of this second feature conversely identifies the item as a foil. At each moment, the current belief about each feature is in one of three possible states: 'Present', 'Absent', or 'Unknown'. Each of the 8 features begins in the 'Unknown' state, and beliefs are updated via the accumulation of noisy perceptual evidence, modeled using a diffusion process (Ratcliff & Rouder, 1998), approximated via the matrix method given in Diederich and Busemeyer (2003). This feature sampling process is notably different than that proposed by Lamberts (1995), which assumes features are absent until sampled; our ternary framing is necessary when missing features actively contribute to the decision.

The diffusion process operates by integrating perceptual information about the stimulus over time. When a feature is absent, the mean rate of accumulation is controlled by the mean drift rate, μ_{Absent}; similarly, when a feature is added to the display, the rate of accumulation switches to a different rate, $\mu_{Present}$. Unlike the standard diffusion process, which terminates as soon as the evidence crosses a decision threshold, we utilize non-absorbing boundaries, which allows the process to continue monitoring for changes in the visible features. A delayed feature that is identified as "Absent" early in the trial can therefore be recognized as "Present," with sufficient information. The parameter that controls how much evidence is needed to move out of the "Unknown" state is governed by a parameter θ, which is a proportion of the total distance from the starting point of the process (assumed to be 0) to either non-absorbing boundary: smaller values correspond to more conservative identification. Regardless of timing of presentation, different features are perceived independently; the joint probability of any combination of 'Present/Absent' judgments, $\phi(t)$, is the product of the individual probabilities.

Memory Retrieval

At each moment in time the currently beliefs about the features, $\phi(t)$, are compared to the well-learned object representations stored in long-term memory. The probability that the presented object is a "target" is given by dividing the number of matching traces from the target class by the total number of retrieved items from either class, each count being augmented by a small constant to add some noise to the comparison process. $Pr(Target|\phi) = \frac{nT+\varepsilon}{nT+nF+\varepsilon}$. This framing highlights the role of feature diagnosticity; if a pair of target-diagnostic features have been perceived and sampled into $\phi(t)$, only traces from the target class will be retrieved from long-term memory, the small amount of noise excepted and thus $Pr(Target|\phi) = \frac{nT}{nT} = 1$.

Decision

In this framing, evidence is grounded in relative, rather than absolute terms (Cox & Shiffrin, 2012), and can be tracked using a log-likelihood ratio. Two decision boundaries are established such that decisions occur when the log-likelihood favoring one of the responses first crosses one of the boundaries. We generate a log-likelihood for each collection of features, which expresses the relative evidence in favor of the current object being a target versus a foil:

$$\beta_{Target}(\phi) = log\left[\frac{Pr(Target|\phi)}{Pr(Foil|\phi)}\right] \tag{1}$$

These log-likelihoods do not depend directly on time, only on the collection of features in short-term memory, ϕ, which changes as time passes. In order to find the distribution of evidence over time, $\beta_{Target}(t)$ we compute a weighted sum of log-likelihoods from each collection, according to the corresponding probability of having each such collection at that time. We utilize the method presented in Cox and Shiffrin (2017), to approximate this distribution as Gaussian, with mean and variance given by[3]:

$$\mu(t) = log(\beta)Pr(\phi,t) \tag{2}$$

$$\sigma(t) = log(\beta)^T\Sigma(t)log(\beta) \tag{3}$$

Responses are generated when the log-odds sufficiently favor one result, which is instantiated by tracking the process until it crosses one of two response boundaries. The two boundaries, are estimated to lie at a distance $(A/2)$ away from the starting point (b), and correspond to target and foil responses (B_{Target}, B_{Foil}). Computing the probability of crossing either boundary at time, t, is easily accomplished using the standard normal cumulative density function, evaluated at the boundary: $Pr(r = R) = \Phi(B_R, \mu(t), \sigma(t))$. Because feature sampling is independent across time, the Gaussians representing the distribution of evidence over time are also independent; the probability of first-passage time is thus given by the probability of having not yet crossed either boundary by time, t, times the immediate probability of crossing. The duration of the decision process is then added to a residual non-decision component, which is estimated separately for Target and Foil responses, allowing for e.g. greater response inhibition for Foils (NDT_T, NDT_F).

[3]The approximation utilizes the fact that the distribution over feature combinations, $\phi(t)$ is a multinomial distribution that must sum to 1. Multiplying by a weight vector (here, the βs for each state), projects the distribution onto a univariate subspace of familiarity, which is approximately normal for large N.

Model Fitting

There are three parameters associated with this perceptual feature sampling process: the two drift rates $(\mu_{Present}, \mu_{Absent})$, and the proportion of the decision space that corresponds to being in the 'Unknown' state (determined by θ). The decision process yielded an additional four parameters, the initial starting bias (b), the distance to the response boundaries (A), and the two non-decision components (NDT_T, NDT_F).

In spite of the relatively small number of parameters in the model, we were able to produce a high degree of match between the empirical data across three subjects providing somewhat heterogeneous data. Bayesian posterior estimates for the parameters were found using Differential Evolution (Turner & Sederberg, 2012). We fit, jointly, the response proportions and the complete response time distributions for correct responses, but only the response proportions for incorrect choices, as there were insufficient trials to estimate the shape of the response time distribution. Table 1 shows the priors for each parameter, as well as the posterior mode for each subject (across the columns).

The model successfully captured the patterns of all three subjects, with the minor exception of predicting too-few *Correct Rejections* for Subject 3. The parameters associated with feature sampling show the expected result, namely that visible features provide more evidence for a feature being 'Present' than when they were not yet visible in the display $(\mu_{Present} > \mu_{Absent})$. For subject 1, missing features provided evidence against its being in the stimulus, as expected; however, for subjects 2 and 3, missing features provided weak, positive evidence, though this was offset by requiring more evidence before deciding whether a feature was present (smaller θ values). One possible interpretation is that this 'head-start' allows for more rapid detection of new features when added to the display, which was consistent with their overall faster responding.

For all participants, the non-decision component associated with "Target" responses was lower than for "Foil" responses, which is justified within the context of the visual search task from which these conditions were drawn, in that rejecting a display and indicating that a target was "Absent" likely requires greater evidence than finding a target and responding "Present."

Name	Prior Distribution	Posterior Mode		
$\mu_{Present}$	$\sim N(0,5)$	1.96	3.22	3.76
μ_{Absent}	$\sim N(0,5)$	-0.82	0.77	0.99
θ	$\sim U(0,0.5)$	0.3	0.08	0.04
b	$\sim U(0,0.5)$	0.51	0.52	0.52
A	$\sim N(20,10)$	10.3	10.05	9.98
NDT_T	$\sim U(0,300)$	123	3	40
NDT_F	$\sim U(0,300)$	187	44	58

Table 1: Prior distributions and posterior modes (columns corresponding to individual subjects) of model parameters.

Discussion

Aside from technical details the model we employ is conceptually simple and coherent: features are sampled and accumulate as time passes. At each moment the collection of current features is matched to the well learned set of eight stimuli. The matching process produces a likelihood that the current collection matches one of the targets versus one of the foils. When this likelihood exceeds a target boundary or falls below a foil boundary a corresponding response is made.

It is clear from the data that a model like this is needed, because the timing and diagnosticity of features changes the pattern of results, for both accuracy and response time. We plan to pursue comparisons of our proposed model with existing approaches aimed at similar problems, such as EGCM-VS (Guest et al., 2017) and EBRW-PE (Cohen & Nosofsky, 2003), as well as extending the model to the conditions in which objects but not features arrive sequentially.

References

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*(3), 153–178. doi: 10.1016/j.cogpsych.2007.12.002

Busemeyer, 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. doi: 10.1037/0033-295X.100.3.432

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*(2), 150–165. doi: 10.1016/S0022-2496(02)00031-7

Cousineau, D., Donkin, C., & Dumesnil, E. (2015). Unitization of features following extended training in a visual search task. *Cognitive Modeling in Perception and Memory: a Festschrift for Richard M. Shiffrin*, 3–15.

Cousineau, D., & Shiffrin, R. M. (2004). Termination of a visual search with large display size effects. *Spatial Vision*, *17*(4-5), 327–352. doi: 10.1163/1568568041920104

Cox, G. E., & Criss, A. H. (2017). Parallel interactive retrieval of item and associative information from event memory. *Cognitive Psychology*, *97*, 31–61. doi: 10.1016/j.cogpsych.2017.05.004

Cox, G. E., & Shiffrin, R. M. (2012). Criterion Setting and the Dynamics of Recognition Memory. *Topics in Cognitive Science*, *4*(1), 135–150. doi: 10.1111/j.1756-8765.2011.01177.x

Cox, G. E., & Shiffrin, R. M. (2017). A dynamic approach to recognition memory. *Psychological Review*, *124*(6), 795–860. doi: 10.1037/rev0000076

Diederich, A., & Busemeyer, J. R. (2003). Simple matrix methods for analyzing diffusion models of choice probability, choice response time, and simple response time. *Journal of Mathematical Psychology*, *47*(3), 304–322. doi: 10.1016/S0022-2496(03)00003-8

Diederich, A., & Oswald, P. (2014). Sequential sampling model for multiattribute choice alternatives with random attention time and processing order. *Frontiers in Human Neuroscience*, *8*(September), 1–13. doi: 10.3389/fnhum.2014.00697

Guest, D., Kent, C., Adelman, J. S., Guest, D., Kent, C., & Adelman, J. S. (2017). Journal of Experimental Psychology : Learning , Memory , and Cognition The Relative Importance of Perceptual and Memory Absolute Identification Determining the Time Course of Absolute Identification.

Holmes, W. R., Trueblood, J. S., & Heathcote, A. (2016). A new framework for modeling decisions about changing information: The Piecewise Linear Ballistic Accumulator model. *Cognitive Psychology*, *85*, 1–29. doi: 10.1016/j.cogpsych.2015.11.002

Lamberts, K. (1995). Categorization Under Time Pressure. *Journal of Experimental Psychology: General*, *124*(2), 161–180. doi: 10.1037/0096-3445.124.2.161

Lamberts, K., & Freeman, R. P. (1999). Building object representations from parts: Tests of a stochastic sampling model. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(4), 904–926. doi: 10.1037/0096-1523.25.4.904

Ratcliff, R., & Rouder, J. N. (1998). Modeling Response Times for Two-Choice Decisions. *Psychological Science*, *9*(5), 347–356. doi: 10.1111/1467-9280.00067

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*(1), 1–66. doi: 10.1037/0033-295X.84.1.1

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*(2), 127–190. doi: 10.1037/0033-295X.84.2.127

Turner, B. M., & Sederberg, P. B. (2012). Approximate Bayesian computation with differential evolution. *Journal of Mathematical Psychology*, *56*(5), 375–385. doi: 10.1016/j.jmp.2012.06.004

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*(3), 550–592. doi: 10.1037//0033-295X.108.3.550