Maximizing masquerading as matching: Statistical learning and decision-making in choice behavior

Angela J. Yu He Huang

Department of Cognitive Science
University of California, San Diego
9500 Gilman Dr. MC 0515
La Jolla, CA 92103

Running title: Maximizing Masquerading as Matching

There has been a long-running debate over whether humans *match* or *maximize* when faced with differentially rewarding options under conditions of uncertainty. While maximizing, i.e. consistently choosing the most rewarding option, is theoretically optimal, humans have often been observed to match, i.e. allocating choices stochastically in proportion to the underlying reward rates. Previous models assumed matching behavior to arise from biological limitations or heuristic decision strategies; this, however, would stand in curious contrast to the accumulating evidence that humans have sophisticated machinery for tracking environmental statistics. It begs the questions of why the brain would build sophisticated representations of environmental statistics, only then to adopt a heuristic decision policy that fails to take full advantage of that information. Here, we revisit this debate by presenting data from a novel visual search task, which are shown to favor a particular Bayesian inference and decision-making account over other heuristic and normative models. Specifically, while subjects' first-fixation strategy appears to indicate matching in aggregate data, they actually maximize on a finer, trial-by-trial timescale, based on continuously updated internal beliefs about the spatial distribution of potential target locations. In other words, matching-like stochasticity in human visual search is neither random nor heuristics-based, but due specifically to fluctuating beliefs about stimulus statistics. These results not only shed light on the matching versus maximizing debate, but also more broadly on human decision-making strategies under conditions of uncertainty.

Keywords: matching versus maximizing, choice behavior, statistical learning, decision-making, Dynamic Belief Model

1 Introduction

There has been a long history of debate over whether humans and animals *match* (Herrnstein, 1961) or *maximize* (Hall-Johnson & Poling, 1984; Blakely, Starin, & Poling, 1988), when choosing among options with unequal rates or probabilities of reward. While maximizing, or consistently choosing the most rewarding option, is theoretically optimal (greatest cumulative accuracy or reward in the long-term), there is a substantial body of literature indicating a curious tendency for humans and animals to match (e.g. Herrnstein, 1961; Sugrue, Corrado, & Newsome, 2004), or to allocate their choices in approximate proportion to the underlying reward rates.

This apparently sub-optimal stochasticity in choice behavior has been interpreted as either a consequence of biological limitations or, relatedly, a fast and frugal heuristic for coping with hard decision problems. For example, one prominent account is *melioration theory* (Herrnstein, 1970) and the formally equivalent "Take-the-Best" (TTB) heuristic (Gigerenzer & Goldstein, 1996): it posits that humans and animals have a limited memory buffer and that they choose the best (maximizing) option based on only a few recent data points, so that on average the choice policy will appear stochastic due to fluctuations in empirical statistics. An even simpler proposal is that subjects base each decision entirely on the last trial, by persisting with the same choice when met with success, and switching otherwise; it is known as the "Win-Stay-Lose-Shift" (WSLS) algorithm (Rapoport & Chammah, 1965; Nowak & Sigmund, 1993; Randall & Zentall, 1997; Warren, 1966; Steyvers, Lee, & Wagenmakers, 2009; Lee, Zhang, Munro, & Steyvers, 2011) for binary choices, and Take-the-Last (TTL) when there are more than two choices (Gigerenzer & Goldstein, 1996). Separately, it has been suggested that matching can serve as a heuristic exploration strategy in a noisy and changeable environment, so that the decision-maker does not persevere with outdated choices (Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006).

The shared assumption of these heuristic accounts, that humans and animals are incapable of utilizing sophisticated decision strategies, stands in curious contrast with the converging behavioral and neurophysiological evidence that the brain possesses the machinery to near-optimally track evolving environmental statistics (Yu & Dayan, 2005b; Daw et al., 2006; Behrens, Woolrich, Walton, & Rushworth, 2007; Nassar et al., 2012; Ide, Shenoy, Yu*, & Li*, 2013). It begs the questions why the brain would build sophisticated representations of environmental statistics, only then to adopt a heuristic decision policy that fails to take full advantage of that information.

Here, we re-examine this matching vs. maximizing debate within a Bayesian ideal observer framework (Green & Swets, 1966), specifically proposing the hypothesis that humans continuously track

environmental statistics with an implicit assumption that the world can change at any moment; consequently, they choose where to search at any given time using an optimal, maximizing strategy but which is based on their dynamically evolving beliefs about environmental statistics. In other words, we hypothesize that the matching-like choice behavior is not due to random exploration or limitations of memory, but due to specific fluctuations in internal beliefs about environmental statistics, coupled with an optimal (maximizing) decision process. The hypothesis that humans have a natural tendency to extract statistical patterns while assuming such patterns are changeable over time is motivated by our previous work (Yu & Cohen, 2009), showing that a similar hypothesis can explain a classical sequential effect in 2-alternative forced choice (2AFC) tasks, in which responses are faster and more accurate if a stimulus extends a recent run of repetitions or alternations, and conversely slower and less accurate when a stimulus violates such a run, even if these runs arise purely by chance (Soetens, Boer, & Hueting, 1985). Subjects appear to act with the implicit assumption that the world is potentially changeable – giving more recent observations greater emphasis in predicting future outcomes, instead of giving uniform weights to the entire history of data (Yu & Cohen, 2009; Wilder, Jones, & Mozer, 2010). Here, we adopt a similar Bayesian modeling framework to examine whether subjects depend more on the recent trials to predict next target location (Dynamic Belief Model; DBM) or treat all previous data equivalently when making that prediction (Fixed Belief Model; FBM). These two candidate models for prior learning/updating reflect differing statistical assumptions that either do (DBM) or do not (FBM) allow the possibility of un-signaled, discrete changes in the statistical regularities in the environment.

To examine this hypothesis, we obtain behavioral data from a novel visual search paradigm, in which subjects can exploit statistical regularities in target location in order to improve the accuracy and efficiency of their search strategy. The key behavioral measure is how subjects allocate their first fixation choice: do they simply follow the last trials target location (WSLS/TTL), do they choose stochastically in proportion to the underlying target distribution (matching), or do they systematically choose the most probable target location (maximizing)? We adopt a visual search task because it has long been known that saccadic eye movements are influenced by various cognitive factors (Yarbus, 1967), such as prior knowledge about target location (He & Kowler, 1989; Einhäuser, Rutishauser, & Koch, 2008), temporal onset (Oswal, Ogden, & Carpenter, 2007), reward probabilities (Roesch & Olson, 2003), and general scene context (Ehinger, Hidalgo-Sotelo, Torralba, & Oliva, 2009). In daily tasks, saccadic patterns have been observed to be different among visual search, scene memorization (Henderson, 2007), reading (Rayner, 1998), tea and sandwich making (Land & Hayhoe, 2001), and driving (Land & Lee, 1994). It is poorly understood how such contextual knowledge is acquired and how it precisely modulates saccadic choices

and perceptual decisions – a scientific lacuna we address here. Compared to the rather abstract or artificial stimuli more commonly used in choice tasks, we expect human subjects to be particularly adept at internalizing and utilizing the spatial statistics of visual targets.

In this work, we compare human fixation choice behavior to the predictions of various models. We consider four Bayesian models, which differ in the assumptions they make about *statistical learning* and *decision-making*. Statistical learning refers to the observers internal representation of the target location statistics, and the sequential updating of the prior distribution over where the target lies based on experienced outcomes. We examine two variants of Bayesian learning models, DBM and FBM. We also examine two different decision processes: (1) *Match*, which produces saccade fixation locations in proportion to the internal predictive distribution of target location, and (2) *Max*, which always chooses to first search the currently most probable target location. Thus, there are four Bayesian models altogether, DBM+Match, DBM+Max, FBM+Match, FBM+Max. In addition, we also consider a heuristic algorithm, *melioration*, which does not require a sophisticated statistical representation. We note that melioration is equivalent to TTB, and subsumes TTL as a special case, where the memory buffer is just the last trial. In the following, we first describe the experimental design and present some basic data. We then use a series of data-model comparison to narrow down the best model for explaining human data.

2 Results

We first briefly describe the visual search task (see Methods for more details), before delving into the experimental findings and comparison to the various models. In the task (Fig. 1a), subjects must find a target stimulus (random-dot motion stimulus moving in a certain direction) in one of three possible locations, with the other two locations containing distractors (random-dot motion stimulus moving in the opposite direction). In the 1:3:9 condition, the target location is biased among the three options with 1:3:9 odds. In the 1:1:1 condition, the target appears in the three locations with equal probability on each trial. To eliminate the complications associated with the spatiotemporal dynamics of covert attention, which we cannot measure directly, the display is gaze-contingent: only the fixated stimulus is visible at any given time, with the other two stimuli being replaced by two small dots located at the center of the stimulus patches. Subjects receive feedback about true target location on each trial after making their choice, as well as their choice accuracy, search duration, and number of switches; they are encouraged through a point-based reward function to be fast, accurate, and efficient with the number of fixation switches (see Methods).

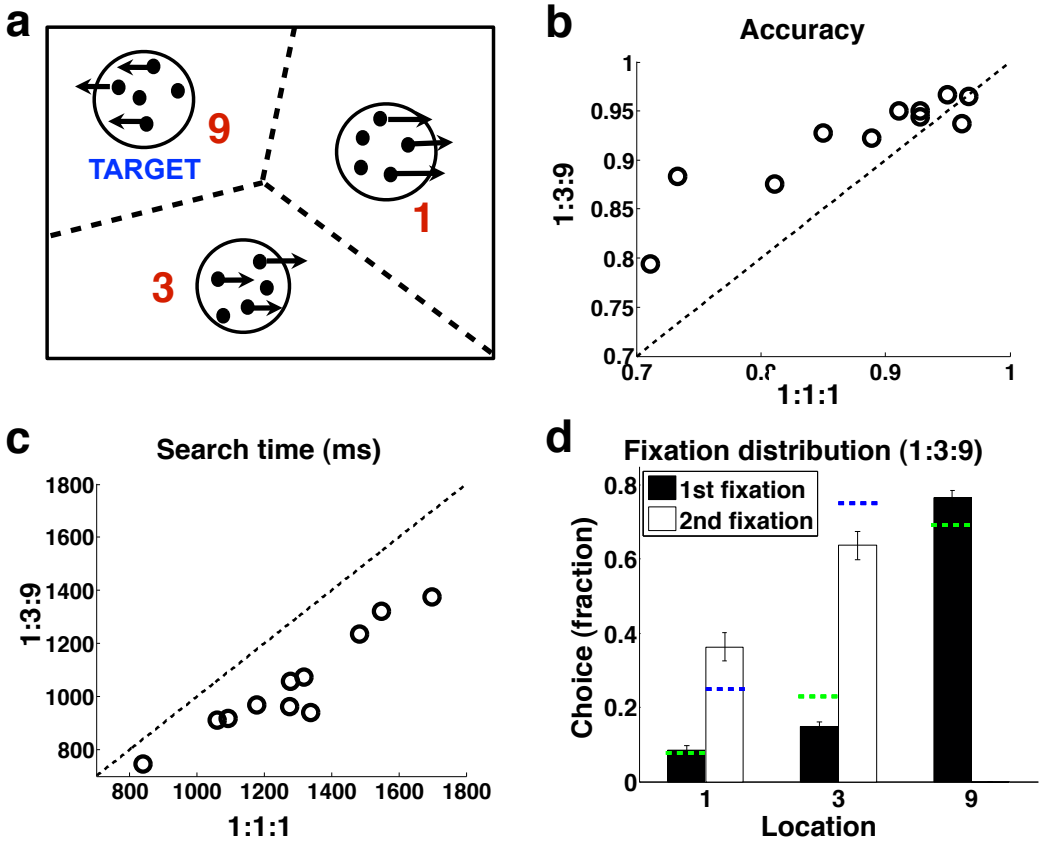

Figure 1: Experimental design and data. (a) On each trial, two of the random-dot stimuli are distractors, one is the target; subjects must find the target (see Methods). (b) Subjects are more accurate in finding the target in the 1:3:9 condition than the 1:1:1 condition, and (c) faster. (d) 1:3:9 condition, allocation of fixation location on first fixation (black), and second fixation when subjects first fixated the 9 location and found that it was not the target (white), averaged over all subjects. Green dashed lines indicate the matching probabilities on the first fixation, $(1/13, 3/13, 9/13)$; blue dashed lines indicate matching probabilities on the second fixation, $(1/4, 3/4)$. $n = 11$. Errorbars: s.e.m. across subjects.

We found that human subjects indeed internalized and exploited the spatial statistics to locate the target stimulus more accurately (Fig. 1b) and rapidly (Fig. 1c). Subjects were more accurate in finding the target in the 1:3:9 condition than the 1:1:1 condition (one-sided t-test, $p < 0.01$), and faster at finding the target in the 1:3:9 condition than the 1:1:1 condition ($p < 0.001$). Underlying this performance improvement was a prioritized search strategy that favors the more probable locations as a fixation choice. For the first fixation, subjects preferentially fixated the 9 location over the 3 location ($p < 0.0001$), which in turn was favored over the 1 location ($p < 0.01$). Similarly, for the second fixation, on trials in which the first fixation was at 9 and that was *not* the target, subjects then favored the 3 location over the 1 location ($p = 0.005$). Altogether, these results indicate that subjects not only knew where the most probable target location was, but had a graded representation of target probabilities at the different locations.

Aggregate statistics are coarse by nature, and much information is lost by averaging all trials together. In particular, it ignores the potential role of statistical learning. Subjects could be learning about spatial statistics based on each experienced trial, in a manner similar to DBM or FBM, and thus a Match or Max strategy should be defined with respect to their internal representation at each moment in time, instead of with respect to the "true" generative statistics (1:3:9 or 1:1:1), which they have no direct access to. Details of DBM and FBM can be found in Methods; here, we briefly describe their assumptions and properties. The versions of DBM and FBM used here are mult-alternative extensions of simpler models we previously developed for 2AFC tasks (Yu & Cohen, 2009). Although the true configuration of most probable, medium probable, and least probable target location does not change within a block (it is pseudo-randomized across blocks), and the relative odds for those locations remain at 1:3:9 for all blocks, DBM allows the possibility that subjects assume the underlying statistics to be changeable within a block. We entertain this hypothesis here, because we previously showed that, in 2AFC tasks, subjects act as though they assume the relative probability of stimulus type is predictable from recent trial history, consistent with DBM (Yu & Cohen, 2009), even though the true experimental statistics are constant (and random) throughout the experiment. Fig. 2a shows the generative model for DBM and FBM. Fig. 2b shows a sample run of DBM on an actual experienced sequence of trials for a subject. The predictive probability DBM assigns to each of the potential target locations on each trial fluctuates with the recent history of experienced trials. DBM+max produces fixation predictions that closely correspond to this subject's actual choices (top panel). The rare discrepancies occur when there are unexpected observations and the underlying probabilities are close in magnitude: for example, trial 80. Most of the time, even when the underlying probabilities are similar due to unexpected observations, DBM+max and the subject concur in switching (57, 78) or staying (64, 71, 75, 76,

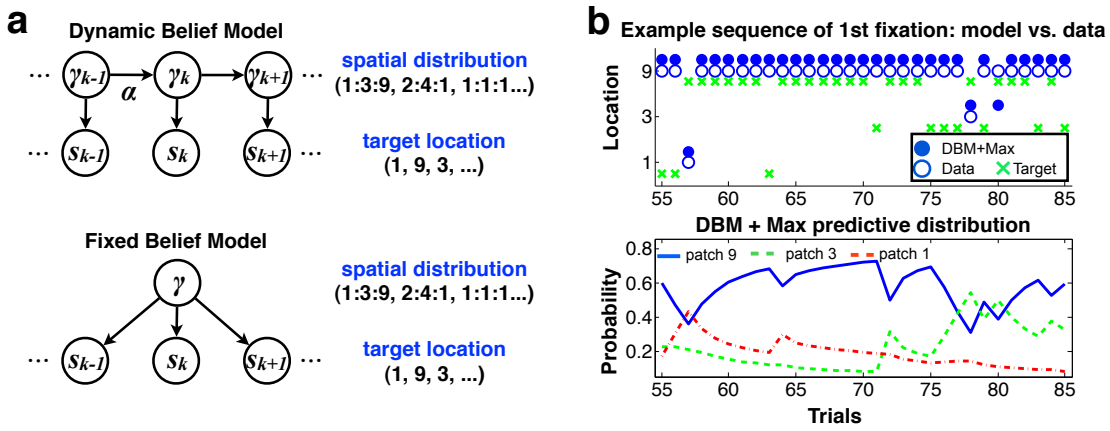

Figure 2: Dynamic Belief Model (DBM) and Fixed Belief Model (FBM) model architecture and comparison to behavioral data. (a) graphical model for DBM and FBM. FBM can be thought of as a special case of DBM, with $\alpha = 1$. (b) An example trial sequence (subject 2, block 4, trials 55-85) and corresponding DBM inference/prediction behavior ($\alpha = 0.92$). Given the sequence of target location experienced by a subject (green crosses in top panel), DBM computes the *predictive* probability on each trial k of each location containing the target (bottom panel), and the maximum is taken as the model prediction of choice (filled blue circle, top panel); compared to the subject's actual first fixation location (open blue circle, top panel).

77, 79, 83, 85).

In addition to the Bayesian models, we entertain the possibility that subjects employ melioration (Herrnstein, 1970), which keeps a limited and fixed memory of the last k trials, and picks the most frequent target location among these k trials as next trial's first fixation choice (ties are broken randomly). It is clear that subjects would show average "matching" behavior even if they always first searched in the last target's location ($k = 1$), because this choice distribution would exactly track the empirical target distribution, with a one-trial lag that would not be apparent in aggregate data. We fit the best memory size (number of recent trials kept in the buffer) for the melioration model, which was left as a free parameter in the original model (Herrnstein, 1970). Note that this subsumes the TTL/WSLS model as a special case with the memory size being 1 trial. By simulating the melioration model with different memory sizes (from 1 to 10), and comparing the choice predicted by the model and subjects' actual first fixation location, we found that the a memory size of 3 trials was the best at predicting subjects' choice.

To distinguish among the models, we examined the evolution of fixation choice pattern of humans, in comparison to the various models. Over the time course of a block, we found that subjects gradually learn to favor the more probable locations, in a manner well-matched by both FBM+Match and DBM+Max (Fig. 3a;b). In contrast, FBM+Max over-matched (Fig. 3a, solid) and DBM+Match

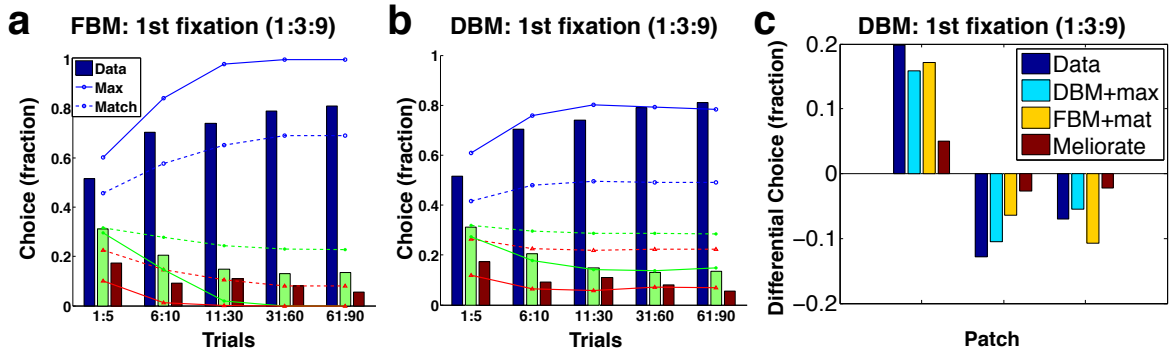

Figure 3: Model comparison of fixation distributions. (a) Averaged over subjects ($n = 11$) and blocks (6), first fixation location (colored bars: blue=9, green=3, red=1) increasingly favor the 9 location over 3, and in turn over 1, for different segments of the 90-trial blocks. FBM+max (solid line) predicts faster learning/over-matching compared to subjects; FBM+match produces predictions more similar to subjects' data. (b) Human data (colored bars) same as in (a). DBM+max produces predictions similar to subjects' data; DBM+match produces slower learning/under-matching compared to subjects. All predictions based on actual sequences of trials experienced by subjects. Errorbars = s.e.m. over blocks and subjects. (c) The average distribution of first-fixation choice, among the three patches, for the last ten trials of each block minus average choice distribution in the first ten trials.

(Fig. 3b, dashed) under-matched subjects' fixation choice distribution. Note that the fact that the human behavior and model traces reach asymptotic values pretty early (Fig. 3a;b) is not necessarily indicative of the cessation of learning after the curves flatten out. *Where* the curves asymptote reflect more on learning, such that if human behavior flattened out at global maximizing (as in FBM+Max in Fig. 3a), then that would indicate cessation of learning. But in fact, human behavior flattens out at a level well below global maximizing, indicating either significant asymptotic learning (DBM+max) or asymptotic random stochasticity (FBM+match).

This data also rule out melioration/TTB/TTL heuristic strategies, since they would produce choice statistics that are constant over the block (because they track empirical statistics, which are on average constant over the block). Fig. 3c shows the model comparison in a different way. When we look at the difference in subjects' average fixation distribution between the last ten trials and the first ten trials of each block, we find that subjects favor the 9 location much more relative to the 3 and 1 locations, toward the end of the block than at the beginning of the block. This trend is best captured by DBM+Max (on average 20% different from data in absolute units, as shown in Fig. 3c), second best by FBM+Match (39% absolute difference from data), and very poorly by melioration (74% absolute difference from data, memory size = 3, best-fitting parameter setting).

Left with two remaining model candidates, DBM+Max and FBM+Match, we next examine sub-

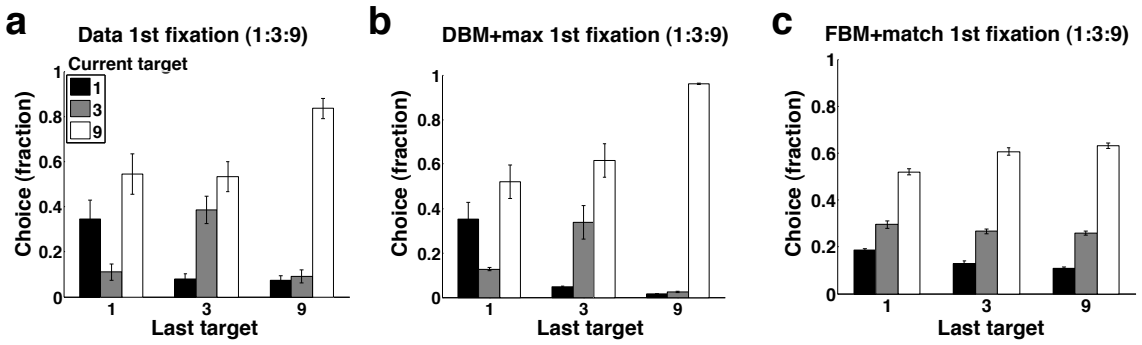

Figure 4: Data vs. model prediction of first fixation distribution conditioned on last target location. (a) In the 1:3:9 condition, human subjects generally prefer 9 (white) over 1 (black) and 3 (gray) regardless of last trial target location, but location 1 is particularly favored if the last target location was also 1 (left group), location 2 is particularly favored if the last target location was also 3 (middle group), and location 9 is particularly favored if the last target was also 9 (right group). (b) DBM+max produces a conditional distribution very similar to the behavioral data in (a). (c) FBM+match produces a conditional distribution dissimilar to the behavioral data in (a), in particular it is relatively insensitive to last trial target location. All model predictions based on actual sequences of stimuli subjects experienced. Errorbars = s.e.m. over blocks and subjects.

jects' fixation choice distribution *conditioned* on the last target location. In FBM, new data have less and less capacity to shift the posterior as the total amount of data builds and the posterior becomes more regid (Yu & Cohen, 2009), thus predicting that last trial's target location to have little to no effect on current trial fixation choice when averaged over the whole experimental session (Fig. 4c). In contrast, DBM continuously entertains the possibility of change, and thus allows its posterior to shift according to new data (Yu & Cohen, 2009), thus predicting that last trial's target location should have asymptotically non-trivial effect on current trial fixation location (Fig. 4b). Fig. 4a shows that human subjects behave as predicted by DBM: while the 9 location is most frequently the first fixation location in all cases, its advantage is much reduced if the last target was in location 1 or 3, and increased if the last target was in location 9. More specifically, following the target being in location 1, first fixation percentage is boosted in location 1 on the subsequent trial; similarly, following the target in in location 3, first fixation percentage is boosted in location 3 on the subsequent trial. This indicates that subjects are not adopting the same stochastic "matching" policy on every trial, but are exquisitely sensitive to recent trial history. While DBM+Max yields conditional distributions statistically indistinguishable from subjects' first fixation distributions (one-sided t-test of average Kullback-Leibler Divergence between subjects' conditional distributions and that of DBM+max, one-sided paired t-test, $p = 0.076$), FBM+match produces conditional distributions that are significantly different (one-sided paired t-test, $p < 0.001$).

A different analysis shows the impressive advantage DBM+max has over other models not only in characterizing gross empirical statistics, but also in predicting trial-to-trial first fixation choice. For each model, we compute on each trial the prior probability each model assigns to the subject's *actual* first-fixation choice. It is 1 or 0 for deterministic models (Follow-Last-Trial and Maximization), for correct and incorrect predictions, and somewhere in between for stochastic models (Match). For DBM, we first find the best-matching α parameter for each subject, which was on average 0.87, with a standard deviation of 0.15. We find that DBM+max has an average predictive accuracy of 0.81, far outperforming TTL or melioration with a 1-trial buffer (0.54), FBM+Match (0.58), DBM+match (0.59), FBM+max (0.78), $p < 0.01$ in each case (two-sided t-test). This big advantage is not due to the extra parameter, α, in DBM, which specifies an individual's belief about the probability of the target location statistics *not* changing from trial to trial (in contrast to FBM, which has no free parameters), as leave-one-block-out cross validation yields a predictive accuracy of 0.80, which is statistically indistinguishable from the training data predictive accuracy (two-tailed t-test, $p = 0.728$). That DBM is not over-fitting is hardly surprising, as we are in the realm of many more data points (540 trials per subject) than parameters (1 per subject).

While DBM+max can predict an individual's first fixation choice about 80% of the time, i.e. subjects choose the most probable target location 80% of the time, subjects do choose the other two locations about 20% of the time. We find that subjects favor the more probable of the remaining two options instead of choosing among equally often (12.2% versus 6.7%, $p = 0.016$). This raises the possibility that subjects may be applying some sort of softmax decision policy that is in between matching and maximizing. We therefore fit a choice distribution (q_1, q_2, q_3) that is a softmax function of the belief distribution (p_1, p_2, p_3):

$$q_i = \frac{p_i^\beta}{\sum_{j \in \{h,m,l\}} p_j^\beta} \, . \tag{1}$$

We use a polynomial form of softmax instead of an exponential form, seen in related work, (e.g. as in Daw et al., 2006), because the polynomial form has a natural interpretation in terms of matching ($\beta = 1$), under-matching ($\beta < 1$), and over-matching ($\beta > 1$). We find that the best-fitting β values across subjects are significantly greater than 1 ($p < 0.00002$), with a mean value of 3.27 (std = 1.00). $\beta = 3.27$ suggests a very strong tendency to maximize, as it would, for example, turn $(p1, p2, p3) = (9/13, 3/13, 1/13)$ into $(q1, q2, q3) = (0.973, 0.027, 0.001)$.

Finally, we note that a thorough analysis of the uniform condition behavior is omitted, because behavior in those blocks is completely unconstrained. Subjects can choose to employ whatever

idiosyncratic strategy they like (including, for example, always starting from the same location, or systematically rotate through them, A, B, C, A, B, C, ...), and they would on average perform exactly equally well. We find that, indeed, there is quite a bit of variability of first-fixation strategy among subjects on the uniform blocks.

3 Methods

3.1 Experimental Design

The data are from eleven subjects, recruited from the UCSD undergraduate students (five females). Subjects first performed a random-dot coherent motion direction discrimination task (Britten, Shadlen, Newsome, & Movshon, 1992) training session and achieved an accuracy exceeding 75% for 12%-coherence stimuli, before continuing onto the main experiment. In the main visual search experiment, subjects must identify one of the three random-dot stimulus patches as the target (left-moving for five subjects, right-moving for six subjects), the other two being distractor stimuli moving in the opposite direction. Subjects began each trial by fixating a central cross, then sequentially fixated one or more stimulus patches until pressing a space bar, which indicated that the last viewed stimulus was the chosen target. The three stimulus patches were circular and equidistant from the central cross, rotationally symmetrically positioned at non-cardinal angles. In the 1:1:1 condition (2 blocks), the target appeared in the three locations with equal likelihood on each trial; in the 1:3:9 locations (6 blocks, one of each possible configuration), the target appeared in the three locations with correspondingly biased probabilities. The order of the eight blocks (six biased blocks and two uniform ones, 90 trials per block) were randomized for each subject. Before the main experiment, subjects experienced 3 practice blocks: respectively, they consisted of 30, 40, and 40 trials, each with target location distribution drawn randomly from the configuration in the main experiment (2/8 probability of a 1:1:1 block, 1/8 probability of each of the 6 1:3:9 blocks). The random-dot motion coherence of the three blocks were 30%, 20%, and 12%, respectively. A target identification accuracy of 80% had to be reached in the first two practice two blocks, or else the same block has to be repeated; similarly, in the third practice block, an accuracy of 68% had to be reached before the subject can proceed to the main experiment. Other than experiencing practice blocks with similar statistics as in the main experiment, subjects did not receive explicit instructions on the spatial distribution of target location.

The gaze-contingent display only revealed a motion stimulus in the fixated location, with the re-

mainder replaced by a central dot; boundaries for fixation determining which stimulus was shown at any given time are as shown in Fig. 1a. Subjects' eye movements were monitored using a SR Research Eyelink 1000 eye tracker. A timing bar on the left side of the screen indicated time elapsed since onset of first stimulus fixation, first decreasing in length (green) until 8 seconds elapsed, and then growing in length in the opposite direction (red) at the same rate, though subjects were told that points were deducted indefinitely in proportion to their total response time (12.5/sec) even after the red bar reaches the edge of the display. At the end of each trial, subjects were shown the true target location and the total points gained/lost for that trial: 100 - 12.5 $\times$ (seconds taken to respond) + 25$\times$(number of fixation switches, 0 if only one patch fixated) $\pm$ 50 (+ if final response correct, - if incorrect). Subjects were told about the reward scheme at the start of the experiment, in addition to receiving detailed feedback, as explained above, during the experiment. Subjects were paid at the end of the experiment proportional to the total points earned, which were calibrated so as to award the average subject about 10 an hour.

Subjects were excluded from the study if they did not have normal or corrected vision, achieved less than 50% accuracy in the main experiment (lower than in the practice blocks), or showed unusually large first fixation spatial bias (> 2 standard deviations away in Kullback-Leibler divergence from the population mean distribution of first fixation, in the 1:3:9 condition).

3.2 Theoretical Modeling

To distinguish the two hypotheses that stochasticity in subjects' saccade choices arise from stable beliefs about target location statistics plus matching-like randomness in action selection, versus fluctuating beliefs about target location statics plus a maximizing strategy, we implement two distinct Bayesian models of trial-by-trial statistical learning: one that assumes that subjects use the entire history of observed data in the current block to infer about the target location statistics (Fixed Belief Model, FBM), and another that assumes that subjects use only recent history to make inferences about target location (Dynamic Belief Model, or DBM). The versions of DBM and FBM used here are multi-alternative extensions of simpler models we previously developed for 2AFC tasks (Yu & Cohen, 2009).

We first describe DBM, and then explain how FBM differs. In the DBM, we model subjects' trial-by-trial inference using a hidden Markov model. In the generative model (Fig. 2a), the target location on trial k, denoted by $s_k \in \{1, 2, 3\}$, depends on the configuration b_k and multinomial

parameters $\boldsymbol{\gamma}_k := (\gamma_h, \gamma_m, \gamma_l)$, where $\gamma_h + \gamma_m + \gamma_l = 1$:

$$P(s_k|b_k, \boldsymbol{\gamma}_k) = \begin{cases} (\frac{1}{3}, \frac{1}{3}, \frac{1}{3},) & b_k = 1 \\ (\gamma_h, \gamma_m, \gamma_l), & b_k = 2 \\ (\gamma_h, \gamma_l, \gamma_m), & b_k = 3 \\ (\gamma_m, \gamma_h, \gamma_l), & b_k = 4 \\ (\gamma_m, \gamma_l, \gamma_h), & b_k = 5 \\ (\gamma_l, \gamma_h, \gamma_m), & b_k = 6 \\ (\gamma_l, \gamma_m, \gamma_h), & b_k = 7 \end{cases} \tag{2}$$

At the start of each experimental block, the prior distribution over b_1 and $\boldsymbol{\gamma}_1$ on the first trial are $p_0(\boldsymbol{\gamma})p_0(b)$, where $p_0(\boldsymbol{\gamma})$ is a Dirichlet distribution $Dir(\boldsymbol{\gamma}; \frac{9}{13}, \frac{3}{13}, \frac{1}{13})$, and

$$p_0(b) = \begin{cases} 1/4 & b = 1 \\ 1/8 & b = 2 \dots 7 \end{cases} \tag{3}$$

In order to capture the assumption of a non-stationarity, we assume $(b_k, \boldsymbol{\gamma}_k)$ to be subject to change: on each trial, it has probability α of remaining the same as the last trial, and probability $1 - \alpha$ of being redrawn from the prior distribution $p_0(b_k, \boldsymbol{\gamma}_k)$. [1]

$$P(b_k, \boldsymbol{\gamma}_k|b_{k-1}, \boldsymbol{\gamma}_{k-1}) = \alpha\delta((b_k, \boldsymbol{\gamma}_k) - (b_{k-1}, \boldsymbol{\gamma}_{k-1})) + (1 - \alpha)p_0(b_k, \boldsymbol{\gamma}_k) \tag{4}$$

The recognition model inverts the above generative process to *infer* the current parameter values $\boldsymbol{\gamma}_k$ and configuration b_k based on observed target locations $\mathbf{s}_k := (s_1 \dots s_k)$. This inference of the joint distribution over $(b_k, \boldsymbol{\gamma}_k)$ can be computed iteratively as follows:

$$P(b_k, \boldsymbol{\gamma}_k|\mathbf{s}_k) \propto P(s_k|b_k, \boldsymbol{\gamma}_k)P(b_k, \boldsymbol{\gamma}_k|\mathbf{s}_{k-1}) \tag{5}$$

$$P(b_k, \boldsymbol{\gamma}_k|\mathbf{s}_{k-1}) = \alpha P(b_{k-1}, \boldsymbol{\gamma}_{k-1}|\mathbf{s}_{k-1}) + (1 - \alpha)p_0(b_k, \boldsymbol{\gamma}_k) \tag{6}$$

The *predictive distribution* over the upcoming target location can be computed by marginalizing

[1] We also explored an alternative model in which the parameters $\boldsymbol{\gamma}$ were *not* subject to unsignaled reset, only b_k is. However, the choices made by this model under a maximizing strategy were statistically indistinguishable from those of the DBM described in the main text.

out model variables:

$$P(s_k|\mathbf{s}_{k-1}) = \sum_{b_k} \int_{\boldsymbol{\gamma}_k} P(s_k|b_k, \boldsymbol{\gamma}_k) P(b_k, \boldsymbol{\gamma}_k|\mathbf{s}_{k-1}) d\boldsymbol{\gamma}_k \tag{7}$$

FBM differs from DBM in that it assumes that the target statistics are stable over time. It can be thought of a special case of DBM, in which $\alpha = 1$.

3.3 Data analysis: model predictive accuracy

We use the average predictive probability of subjects' first fixation under each model for comparing among the various models. For DBM+max and FBM+max, the predictive probability on a trial is 1 if the model successfully predicted the subject's first fixation choice on that trial, and 0 otherwise. For DBM+match and FBM+match, the predictive probability refers to the probability assigned by each model to the subject's choice. These predictive probabilities are then averaged across subjects and blocks of trials.

For DBM, we use leave-block-out cross-validation performance to assess any potential over-fitting due to the α parameter, the sole free parameter in the model. For the 1:3:9 condition, for each subject, the average predictive probability on 5 nonuniform blocks was minimized by selecting α from a range of values covering the range [0, 1]. The model's performance was then evaluated, for the chosen value of α, on the held-out 6th nonuniform block. This procedure was repeated with each block as hold-out set, and the average "test-data" predictive accuracy is reported in the main text.

4 Discussion

The experimental and theoretical results in this study shed new light on the debate of matching versus maximizing in choice behavior. As evidenced by our careful model-based analysis of human fixation choice behavior in a novel visual search task, apparently matching-like behavior actually arises from a dynamically evolving Bayesian learning process, combined with a maximizing decision policy. We find that humans readily internalize spatial statistics after just a handful of exemplars, and use that information to improve accuracy and efficiency in target search by biasing saccadic choice in a systematic manner. This work shows that the control of eye movements is

not only sensitive to low-level sensori-motor factors previously identified, such as saliency (Itti & Koch, 2000) or long-term ones such as visual acuity map (Najemnik & Geisler, 2005), but also to dynamically changing contextual factors such as evolving spatial knowledge about the spatial distribution of target location.

Our results contrast with most previous work on choice behavior under conditions of uncertainty, which attributed matching-like behavior of subjects solely to inherent stochasticity in decision policy (Sugrue, Corrado, & Newsome, 2005; Daw et al., 2006; Vul, Goodman, Griffiths, & Tenenbaum, 2009); it builds on recent work suggesting that humans are continuously learning about environmental statistics due to an implicit assumption of a changing world (Yu & Cohen, 2009) and that matching-like behavior may arise from mis-tuned prior probabilistic beliefs rather than a truly sub-optimal decision policty (Grene, Benson, Kersten, & Schrater, 2010). The specific contribution we make here is that matching-like behavior in part arises from a maximizing (and optimal) choice strategy based on stochastic beliefs about stimulus statistics, which are driven by chance fluctuations in empirical statistics over the experimental session. While the non-stationary assumption seems sub-optimal in our experimental context, it would be a valuable asset in natural environments where statistical regularities do change over time, such as seasonal weather patterns, rise and fall in predator and prey populations, financial and economic markets, and so on. We hypothesize that the apparently irrational matching behavior is an adaptive response to the inherent non-stationarity in natural environments, and that the variability in how close subjects act like a "matcher" versus a "maximizer" may arise from implicit assumptions about the stability of environmental statistics in a particular behavioral context.

We found in the study that as a population, human subjects tend to act as though they believe environmental statistics could be changing on the order of once every seven to eight trials (corresponding to the population mean of the best-fittingα value of 0.87), though there was significant variability across the population. We also found that subjects' choice distribution significantly over-matched with respect to their prior probability distribution (with an average polynomial exponent of 3.27), again with significant variability across the population. It is possible that subjects may not be exactly maximizing, but injecting a certain amount of stochasticity into their choice policy that is still highly over-matching. However, with the current experimental design, it is difficult if not impossible to distinguish whether there is true stochasticity in their choice policy, or whether there is *apparent* noise due to DBM still not being quite the correct learning model. Indeed, it has been shown in a two-alternative forced choice task that a modification to the DBM learning model, termed DBM2 (Dynamic Belief Mixture Model), captures certain minor but systematic aspects of sequential adjustment not captured by DBM (Wilder et al., 2010). Future experimental

and modeling work will certainly be helpful to further unravel the precise nature of learning and decision-making in these tasks.

While subjects behave as bounded rational observers, operationalized as iterative Bayesian inference, we note that this work does not necessarily imply that doing the task requires explicit representation of probabilities. The brain evolved under the selective pressure to approach statistical optimality, but may do so mechanistically without any explicit representation or understanding of probabilities. Indeed, previously we have shown that the predictive probabilities yielded by the Bayesian Model, DBM, can be well approximated by simple leaky accumulating neuronal dynamics, as long as the parameters of the dynamics are tuned just right to reflect the statistics of the problem (Yu & Cohen, 2009). Relatedly, although at first glance, FBM appears to need to keep track of all past observations at all times, and thus may impose an unrealistic demand on an arbitrary large memory, it can be actually be implemented exactly by keeping track of a running total of the number of times the target appears in each of the stimulus locations, as these provide the sufficient statistics for the Dirichlet posterior distribution (Bishop, 2006). Thus, the implementation of neither FBM nor DBM requires an explicit representation of probabilities or particularly complex computations, and the issue of computational complexity and neural plausibility is not a particularly pertinent one for dismissing one model in favor of the other, or both altogether.

While not explicitly addressed here, our results do not preclude the possibility that humans can learn about the true nature of the stability, or lack thereof, of statistical regularities in the environment. However, as we showed previously (Yu & Cohen, 2009), it can take even an ideal Bayesian learner a surprising number of trials to exchange a prior bias toward changeable, statistical regular world for a random, stationary world – thus, even if humans are capable of adapting to the rate of statistical change in the world, the length of our experimental session may be insufficient for that type of learning. Future work is needed to clarify the extent to which humans can adapt their internal assumptions about the rate of change of the world in different behavioral contexts.

Our results are also relevant for the study of attention. Our findings demonstrate that overt attention, mediated by purposeful eye movements, complement covert attention to play a critical role in the brain's selection and filtering process. While traditionally attentional selection was thought of arising from limited neuronal resources at perceptual, cognitive, and motor levels (Broadbent, 1958; Deutsch & Deutsch, 1963; Treisman, 1969; Eriksen & St James, 1986), more recently formal Bayesian statistical models have suggested covert attentional selection to be computationally desirable beyond any resource limitation considerations (Dayan & Yu, 2002; Dayan & Zemel, 1999; Yu & Dayan, 2005a; Yu & Cohen, 2009). The current work adds to this "selection-for-

computation" principle of attentional selection (Dayan & Zemel, 1999; Yu & Dayan, 2005b; Yu, Dayan, & Cohen, 2009), the concept that attentional processes support the optimization of the inductive process (Helmholtz, 1878) inherent in sensory and perceptual processing (see Yu, 2013 for a longer discussion). Specifically, this work demonstrates that eye movements contribute to sensory processing efficiency by specifically favoring sensing locations in a manner that is sensitive to environmental statistics and task objectives. Understanding the precise manner in which covert and overt attention interact to mediate efficient sensory processing is an important direction of future research.

5 Acknowledgments

We thank the assistance from R. Jackson and C. Carper, and J. Schilz in writing the experimental code, J. Schilz for help with running subjects, members of H. Poizner's lab for help in operating the eyetracker, and P. Shenoy for help with comparing per-trial model accuracy. We also thank Peter Dayan for helpful comments on the paper. This work was in part supported by NSF to the Temporal Dynamics of Learning Center, by a UCSD academic senate research grant to A.J.Yu, and by ARL to a CAN CTA consortium including A.J.Yu.

References

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neurosci*, *10*(9), 1214–21.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Blakely, E., Starin, S., & Poling, A. (1988). Human performance under sequences of fixed-ratio schedules: Effects of ratio size and magnitude of reinforcement. *Psychological Record*, *38*, 111-20.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: a comparison of neuronal and psychophysical performance. *J. Neurosci.*, *12*, 4745-65.

Broadbent, D. (1958). *Perception and communication*. Elmsford, New York: Pergamon.

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*(7095), 876-9.

Dayan, P., & Yu, A. J. (2002). ACh, uncertainty, and cortical inference. In T. G. Dietterich, S. Becker, & Z. Ghahramani (Eds.), *Advances In Neural Information Processing Systems 14* (p. 189-196). Cambridge, MA: MIT Press.

Dayan, P., & Zemel, R. S. (1999). Statistical models and sensory attention. *ICANN Proceedings*.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, *87*, 272-300.

Ehinger, K. A., Hidalgo-Sotelo, B., Torralba, A., & Oliva, A. (2009). Modeling search for people in 900 scenes: A combined source model of eye guidance. *Visual Cognition*, *17*, 1366-1378.

Einhäuser, W., Rutishauser, U., & Koch, C. (2008). Task demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision*, *8*(2), 1-19.

Eriksen, C., & St James, J. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics*, *40*(4), 225-40.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*(4), 650-669.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Los Altos, CA: Peninsula Publishing.

Grene, C. S., Benson, C., Kersten, D., & Schrater, P. (2010). Alterations in choice behavior by manipulations of world model. *Proc Natl Acad Sci U S A*, *107*, 16401-6.

Hall-Johnson, E., & Poling, A. (1984). Preference in pigeons given a choice between sequences of fixed-ratio schedules: Effects of ratio values and duration of food delivery. *Jounral of the*

Experimental Analysis of Behavior, *42*, 127-35.

He, P. Y., & Kowler, E. (1989). The role of location probability in the programming of saccades: implications for "center-of-gravity" tendencies. *Vision Res*, *29*(9), 1165-81.

Helmholtz, H. L. F. v. (1878). The facts of perception. In R. Kahl (Ed.), *Selected writings of hermann von helmholtz*. Middletown, CT: Wesleyan University Press, 1971. (Translated from German original Die Tatsachen in der Wahrnehmung.)

Henderson, J. M. (2007). Regarding scenes. *Current Directions in Psychological Science*, *16*(4), 219-22.

Herrnstein, R. J. (1961). Relative and absolute strength of responses as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behaviour*, *4*, 267-72.

Herrnstein, R. J. (1970). On the law of effect. *Journal of the Experimental Analysis of Behaviour*, *13*, 243-66.

Ide, J. S., Shenoy, P., Yu*, A. J., & Li*, C.-R. (2013). Bayesian prediction and evaluation in the anterior cingulate cortex. *Journal of Neuroscience*, *33*, 2039-2047. (*Yu and Li contributed equally as senior authors)

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10-12), 1489-506.

Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research*, *41*(25-26), 3559-65.

Land, M. F., & Lee, D. N. (1994). Where we look when we steer. *Nature*, *369*(6483), 742-4.

Lee, M. D., Zhang, S., Munro, M., & Steyvers, M. (2011). Psychological models of human and optimal performance in bandit problems. *Cogn Syst Res*, *12*, 164-74.

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, *434*(7031), 387-91.

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience*, *15*(7), 1040-1046.

Nowak, M., & Sigmund, K. (1993). A strategy of win-stay, lose-shift that outperforms Tit-for-Tat in the Prisoner's Dilemma game. *Nature*, *364*, 56-8.

Oswal, A., Ogden, M., & Carpenter, R. H. S. (2007). The time course of stimulus expectation in a saccadic decision task. *J Neurophysiol*, *97*, 2722-30.

Randall, C. K., & Zentall, T. R. (1997). Win-stay/lose-shift and win-shift/lose-stay learning by pigeons in the absence of overt response mediation. *Behavioural Processes*, *41*(3), 227-36.

Rapoport, A., & Chammah, A. M. (1965). *Prisoner's dilemma*. Ann Arbor, MI: University of

Michigan Press.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372-422.

Roesch, M. R., & Olson, C. R. (2003). Impact of expected reward on neuronal activity in prefrontal cortex, frontal and supplementary eye fields and premotor cortex. *J Neurophysiol*, 1766-89.

Soetens, E., Boer, L. C., & Hueting, J. E. (1985). Expectancy or automatic facilitation? separating sequential effects in two-choice reaction time. *J. Exp. Psychol.: Human Perception & Performance*, *11*, 598-616.

Steyvers, M., Lee, M. D., & Wagenmakers, E. J. (2009). A bayesian analysis of human decision-making on bandit problems. *J Math Psychol*, *53*, 168-79.

Sugrue, L. P., Corrado, G. S., & Newsome, W. T. (2004). Matching behavior and the representation of value in the parietal cortex. *Science*, *304*(5678), 1782-7.

Sugrue, L. P., Corrado, G. S., & Newsome, W. T. (2005). Choosing the greater of two goods: Neural currencies for valuation and decision making. *Nature Reviews Neuroscience*, *6*, 263-75.

Treisman, A. (1969). Strategies and models of selective attention. *Psychol. Rev.*, *76*, 282-99.

Vul, E., Goodman, N. D., Griffiths, T. L., & Tenenbaum, J. B. (2009). One and done? optimal decisions from very few samples. In *Proceedings of the 31st annual conference of the cognitive science society.* Amsterdam, Netherlands.

Warren, J. M. (1966). Reversal learning and the formation of learning sets by cats and rhesus monkeys. *Journal of Comparative and Physiological Psychology*, *61*(3), 421-8.

Wilder, M. H., Jones, M., & Mozer, M. C. (2010). Sequential effects reflect parallel learning of multiple environmental regularities. In P. P. J. Zemel R. S. Bartlett & K. Q. Weinberger (Eds.), *Advances in neural information processing systems* (Vol. 24, p. 1791-9). La Jolla, CA: NIPS Foundation.

Yarbus, A. F. (1967). *Eye movements and vision*. New York: Plenum Press.

Yu, A. J. (2013). Bayesian models of attention. In S. Kastner & K. Nobre (Eds.), *Handbook of attention.* Oxford, UK: Oxford University Press.

Yu, A. J., & Cohen, J. D. (2009). Sequential effects: Superstition or rational behavior? *Advances in Neural Information Processing Systems*, *21*, 1873-80.

Yu, A. J., & Dayan, P. (2005a). Inference, attention, and decision in a Bayesian neural architecture. In L. K. Saul, Y. Weiss, & L. Bottou (Eds.), *Advances in Neural Information Processing Systems 17.* Cambridge, MA: MIT Press.

Yu, A. J., & Dayan, P. (2005b). Uncertainty, neuromodulation, and attention. *Neuron*, *46*, 681-92.

Yu, A. J., Dayan, P., & Cohen, J. D. (2009). Dynamics of attentional selection under conflict: Toward a rational Bayesian account. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 700-17.