Optimal policy for attention-modulated decisions explains human fixation behavior

Optimal policy for attention-modulated decisions
explains human fixation behavior

by

Anthony Injoon Jang

Harvard-M.I.T. Division of Health Sciences and Technology

Submitted in Partial Fulfillment of
the Requirements for the M.D. Degree

February 2021

Area of Concentration:	Neuroscience
Project Advisor:	Jan Drugowitsch, PhD
Prior Degrees:	B.S. (Neuroscience)

I have reviewed this thesis. It represents work done by the author under my guidance/supervision.

Thesis Advisor

Optimal policy for attention-modulated decisions explains human fixation behavior

Thesis by

Anthony Injoon Jang

Submitted in Partial Fulfillment of

the Requirements for the M.D. Degree

Harvard Medical School

Harvard-MIT Division of Health Sciences and Technology

Boston, Massachusetts, USA

2021

Acknowledgments

I would like to thank my adviser Jan Drugowitsch for welcoming me into his lab as a medical student without a strong background in theoretical neuroscience. This work would not have been possible without his patience and enthusiasm in helping me succeed. Jan's doors were always open, and he was happy to answer any questions or go through a derivation step-by-step. I have no doubt the skills and knowledge I gained working with Jan will help me tremendously as an independent investigator in the future.

I would also like to thank members of the Drugowitsch lab including Johannes Bill, Emma Krause, Luke Rast and Mohammadmehdi Kafashan for their friendship, advice and collaboration. Spending time with them served as an awesome respite from the medical school and hospital rotations.

I would also like to thank Ian Krajbich for sharing the behavioral data which allowed us to compare model simulations to human behavior, which is a crucial aspect of the project.

Declaration of authorship

I, Anthony Injoon Jang, declare that this thesis titled, ‘Optimal policy for attention-modulated decisions explains human fixation behavior’ and the work presented in it are my own. I have personally completed the mathematical derivations and computational simulations that generated the results presented here. However, I acknowledge that the derivations presented here were a learning process for me, and would not have been possible without the guidance from my adviser, Jan Drugowitsch, who is an expert in the field. As a part of this acknowledgement, I used the pronoun ‘we’ rather than ‘I’ to describe the work.

Abstract

Traditional accumulation-to-bound decision-making models assume that all choice options are processed with equal attention. In real life decisions, however, humans alternate their visual fixation between individual items to efficiently gather relevant information [1]. These fixations also causally affect one's choices, biasing them toward the longer-fixated item [2]. We derive a normative decision-making model in which attention enhances the reliability of information, consistent with neurophysiological findings [3]. Furthermore, our model actively controls fixation changes to optimize information gathering. We show that the optimal model reproduces fixation-related choice biases seen in humans and provides a Bayesian computational rationale for this phenomenon. This insight led to additional predictions that we could confirm in human data. Finally, by varying the relative cognitive advantage conferred by attention, we show that decision performance is benefited by a balanced spread of resources between the attended and unattended items.

Contents

CHAPTER 1

Introduction

Would you rather have a donut or an apple as a mid-afternoon snack? If we instantaneously knew their associated rewards, we could immediately choose the higher-rewarding option. However, such decisions usually take time and are variable, suggesting that they arise from a neural computation that extends over time [4, 5]. In the past, such behavior has been modeled descriptively with accumulation-to-bound models that continuously accumulate noisy evidence from each choice option, until a decision boundary is reached in favor of a single option over its alternatives. Such models have been successful at describing accuracy and response time data from human decision makers performing in both perceptual and value-based decision tasks [6, 7]. Recently, we and others showed that, if we assume these computations to involve a stream of noisy samples of each item's perceptual feature (for perceptual decisions) or underlying value (for value-based decisions), then the normative strategy could be implemented as an accumulation-to-bound model [8–10]. Specifically, the normative strategy could be described with the diffusion decision model [6] with time-varying decision boundaries that approach each other over time.

Standard accumulation-to-bound models assume that all choice options receive equal attention during decision-making. However, the ability to drive one's attention amidst multiple, simultaneous trains of internal and external stimuli is an integral aspect of everyday life. Indeed, humans tend to alternate between fixating on different items when making decisions, suggesting that control of overt visual attention is intrinsic to the decision-making process [11, 12]. Furthermore, their final choices are biased towards the item that they looked at longer, a phenomenon referred to as the choice bias [2, 13–15]. While several prior studies have developed decision-making models that incorporate attention [2, 16–19], our goal was to develop a normative framework that incorporates control of attention as an intrinsic aspect of the decision-making process in which the agent must efficiently gather information from all items while minimizing the deliberation time, akin to real life decisions. In doing so, we hoped to provide a computational rationale for why fixation-driven choice biases seen in human behavior may arise from an optimal decision strategy. For example, the choice bias has been previously replicated with a modified accumulation-to-bound model, but the model assumed that fixations are driven by brain processes that are exogenous to the computations involved in decision-making [2]. This stands in contrast to studies of visual attention where fixations appear to be controlled to extract choice-relevant information in a statistically efficient manner, suggesting that fixations are driven by processes endogenous to the decision [1, 20–23].

We asked if the choice bias associated with fixations can be explained with a unified framework in which fixation changes and decision-making are part of the same process. To do so, we endowed normative decision-making models [10] with attention that boost the amount of information one collects about each choice option, in line with neurophysiological findings [3, 24–26]. We furthermore assumed that this attention was overt [27, 28], and thus reflected in the decision maker's gaze which was controlled by the decision-making process.

We first derive the complex normative decision-making strategy arising from these assumptions and characterize its properties. We then show that this strategy featured the same choice bias as observed in human decision makers: it switched attention more frequently when deciding between items with similar values, and was biased towards choosing items that were attended last, and attended longer. It furthermore led to new predictions that we could confirm in human behavior: choice biases varied based on the amount of time spent on the decision and the average desirability across both choice items. Lastly, it revealed why the observed choice biases might, in fact, be rational. Overall, our work provides a unified framework in which the optimal, attention-modulated information-seeking strategy naturally leads to biases in choice that are driven by visual fixations, as observed in human decisions.

CHAPTER 2

Results

2.1 An attention-modulated decision-making model

Before describing our attention-modulated decision-making model, we will first briefly recap the attention-free value-based decision-making model [10] that ours builds upon. This model assumes that for each decision trial, a true value associated with each item (z_1,z_2) is drawn from a normal prior distribution with mean $\bar{z}$ and variance σ_z^2. Therefore, $z_j \sim (\bar{z}, \sigma_z^2)$ for both $j \in \{1, 2\}$. The smaller σ_z^2, the more information this prior provides about the true values. We assume the decision maker knows the shape of the prior, but can't directly observe the drawn true values. In other words, the decision maker a-priori knows the range of values associated with the items they need to compare, but doesn't know what exact items to expect nor what their associated rewards will be. For example, one such draw might result in a donut and an apple, each of which has an associated value to the decision maker (i.e., satisfaction upon eating it). In each nth time step of length δt, they observe noisy samples centered around the true values, called *momentary evidence*[1], $\delta x_{j,n} | z_j \sim (z_j \delta t, 2\sigma_x^2 \delta t)$. The variance σ_x^2 here controls how informative the momentary evidence is about the associated true value. A large σ_x^2 implies larger noise, and therefore less information provided by each of the momentary evidence samples. While the model is agnostic to the origin of these samples, they might arise from computations to infer the items' values (e.g., how much do I currently value the apple?), memory recall (e.g., how much did I previously value the apple?), or a combination thereof [5]. As the decision maker's aim is to choose the higher-valued item, they ought to accumulate evidence for some time to refine their belief in the items' values. Once they have accumulated evidence for $t = N\delta t$ seconds, their posterior belief for the value associated with either item is

$$z_j | \delta x_{j,1:N} \sim \left(\frac{\sigma_x^2 \sigma_z^{-2} \bar{z} + \frac{1}{2} x_j(t)}{\sigma_x^2 \sigma_z^{-2} + \frac{1}{2} t}, \frac{\sigma_x^2}{\sigma_x^2 \sigma_z^{-2} + \frac{1}{2} t} \right), \tag{2.1}$$

where $x_j(t) = \sum_{n=1}^{N} \delta x_{j,n}$ is the accumulated evidence for item j [10]. The mean of this posterior (i.e., the first fraction in brackets) is a weighted sum of the prior mean, $\bar{z}$, and the accumulated evidence, $x_j(t)$. The weights are determined by accumulation time, t, and the variances of the prior,

[1] In [10] the variance of the momentary evidence was $\sigma_x^2 \delta t$ rather than $2\sigma_x^2 \delta t$. We here added the factor 2 without loss of generality to relate it more closely to the attention-modulated version we introduce further below.

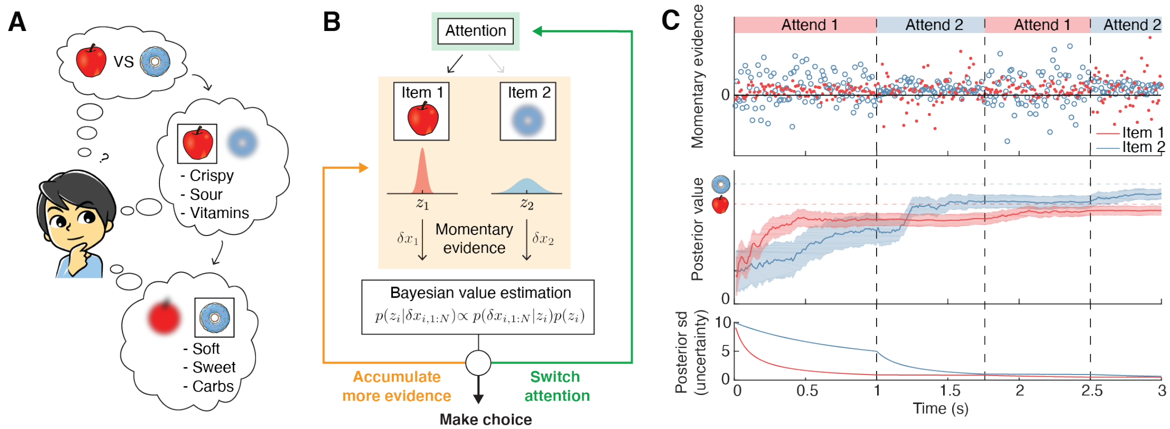

FIGURE 2.1 Attention-modulated evidence accumulation. (A) Schematic depicting the value-based decision-making model. When choosing between two snack items (e.g., apple versus donut), people tend to evaluate each item in turn, rather than think about all items simultaneously. While evaluating one item, they will pay less attention to the unattended item (blurred item). (B) Schematic of the value-based decision process for a single trial. At trial onset, the model randomly attends to one item (green box). At every time step, it accumulates momentary evidence (orange box) that provides information about the true value of each item, which is combined with the prior belief of each item's value to generate a posterior belief. Note that the momentary evidence of the attended item comes from a tighter distribution. Afterwards, the model assesses whether to accumulate more evidence (orange), make a choice (black), or switch attention to the other item (green). (C) Evolution of the evidence accumulation process. The top panel shows momentary evidence at every time point for the two items. Note that evidence for the unattended item has a wider variance. The middle panel shows how the posterior estimate of each item may evolve over time (mean $\pm$ 1SD). The dotted lines indicate the unobserved, true values of the two items. The bottom panel shows how uncertainty decreases regarding the true value of each item. As expected, uncertainty decreases faster for the currently attended item compared to the unattended one. For this descriptive figure, we used the following parameters: $z = [13, 10]$, $\sigma_x^2 = 5$, $\sigma_z^2 = 10$, $\gamma = 0.1$, $\delta t = 0.01$.

σ_z^2, and the momentary evidence, σ_x^2, which control their respective informativeness. Initially, $t = 0$ and $x_j(t) = 0$, such that the posterior mean equals that of the prior, $\bar{z}$. Over time, with increasing t, the influence of $x_j(t)$ becomes dominant, and the mean approaches $x_j(t)/t$ (i.e., the average momentary evidence) for large t, at which point the influence of the prior becomes negligible. The posterior's variance (i.e., the second fraction in brackets) reflects the uncertainty in the decision maker's value inference. It initially equals the prior variance, σ_z^2, and drops towards zero once t becomes large. In this attention-free model, uncertainty monotonically decreases identically over time for both items, reflecting the standard assumption of accumulation-to-bound models that, in each small time period, the same amount of evidence is gathered for either choice item.

To introduce attention-modulation, we assume that attention limits information about the unattended item (Figure 2.1). This is consistent with behavioral and neurophysiological findings showing that attention boosts behavioral performance [3, 29, 30] and the information encoded in neural populations [26, 31, 32]. To achieve this, we first assume that the total rate of evidence across both items, as controlled by σ_x^2, is fixed, and that attention modulates the relative amount of information gained about the attended versus unattended item. This 'attention bottleneck' is controlled by κ $(0 \leq \kappa \leq 1)$, such that κ represents the proportion of total information received for the unattended item, versus $1 - \kappa$ for the attended item. The decision maker can control which item to attend to, but has no control over the value of κ, which we assume is fixed and known. To limit information, we change the momentary evidence for the attended item j to $\delta x_{j,n} \sim \left(z_j \delta t, \frac{1}{1-\kappa}\sigma_x^2 \delta t\right)$, and for the unattended item $k = 3 - j$ to $\delta x_{k,n} \sim \left(z_k \delta t, \frac{1}{\kappa}\sigma_x^2 \delta t\right)$. Therefore, if $\kappa \leq \frac{1}{2}$, the variance of the unattended item increases (i.e., noisier evidence) relative to the attended item. This makes the momentary evidence less informative about z_k, and more informative about z_j, while leaving the overall amount of information unchanged (see Methods). Setting $\kappa = \frac{1}{2}$ indicates equally informative momentary evidence for both items, and recovers the attention-free scenario [10].

Lowering information for the unattended item impacts the value posteriors as follows. If the decision maker again accumulates evidence for some time $t = N\delta t$, their belief about item $j = 1$'s value changes from Eq. (2.1) to

$$z_1 | \delta x_{1,1:N} \sim \left(\frac{\sigma_x^2 \sigma_z^{-2} \bar{z} + (1-\kappa) X_1(t)}{\sigma_x^2 \sigma_z^{-2} + (1-\kappa) t_1 + \kappa t_2}, \frac{\sigma_x^2}{\sigma_x^2 \sigma_z^{-2} + (1-\kappa) t_1 + \kappa t_2} \right), \tag{2.2}$$

where t_1 and t_2, which sum up to the total accumulation time ($t = t_1 + t_2$), are the durations that items 1 and 2 have been attended, respectively. The accumulated evidence $X_1(t)$ now isn't simply the sum of all momentary pieces of evidence, but instead down-weights them by $\frac{1-\kappa}{\kappa}$ if the associated item is unattended (see Methods). This prevents the large inattention noise from swamping the overall estimate [33]. An analogous expression provides the posterior $z_2 | \delta x_{2,1:N}$ for item 2 (see Appendix).

The attention modulation of information is clearly observable in the variance of the value's posterior for item 1 (Eq. (2.2)). For $\kappa < \frac{1}{2}$, this variance, which is proportional to the decision maker's uncertainty about the option's value, drops more quickly over time if item 1 rather than item 2 is attended (i.e., if t_1 rather than t_2 increases). Therefore, it depends on how long each of the two items have been attended to, and might differ between the two items across time (Figure 2.1C). As a result, decision performance depends on how much time is allocated to attending to each item.

The decision maker's best choice at any point in time is to choose the item with the larger expected value, as determined by the value posterior. However, the posterior by itself does not determine when it is best to stop accumulating evidence. In our previous attention-free model, we addressed the optimal stopping time by assuming that accumulating evidence comes at cost c per second, and found the optimal decision policy under this assumption [10]. Specifically, at each time step of the decision-making process, the decision maker could choose between three possible actions. The first two actions involve immediately choosing one of the two items, which promises the associated expected rewards. The third action is to accumulate more evidence that promises more evidence, better choices, and higher expected reward, but comes at a higher cost for accumulating evidence. We found the optimal policy using dynamic programming that solves this arbitration by constructing a value function that, for each stage of the decision process, returns all expected rewards and costs from that stage onward [34, 35]. The associated policy could then be mechanistically implemented by an accumulation-to-bound model that accumulates the difference in expected rewards, $\Delta = \langle z_2 | \delta x_{2,1:N} \rangle - \langle z_1 | \delta x_{1,1:N} \rangle$, and triggers a choice once one of two decision boundaries, which collapse over time, is reached [10].

Once we introduce attention, a fourth action becomes available: the decision maker can choose to switch attention to the currently unattended item (Figure 2.1B). If such a switch comes at no cost, then the optimal strategy would be to continuously switch attention between both items to sample them evenly across time. We avoid this physically unrealistic scenario by introducing a cost c_s for switching attention. This cost may represent the physical effort of switching attention, the temporal cost of switching [18, 36], or a combination of both. Overall, this leads to a value function defined over a four-dimensional space: the expected reward difference Δ, the evidence accumulation times t_1 and t_2, and the currently attended item $y \in \{1, 2\}$ (see Appendix). As the last dimension can only take one of two values, we can equally use two three-dimensional value functions. This results in two associated policies that span the three-dimensional *state space* (Δ, t_1, t_2) (Figure 2.2).

2.2 Features of the optimal policy

At any point within a decision, the model's current state is represented by a location in this 3D policy space, such that different regions in this space designate the optimal action to perform (i.e., choose, accumulate, switch). The boundaries between these regions can be visualized as contours in this 3D state space (Figure 2.2A). As previously discussed, there are two distinct policy spaces for when the decision maker is attending to item 1 versus item 2 that are symmetric to each other (Figure 2.2B).

Within a given decision, the deliberation process can be thought of as a particle that drifts and diffuses in this state space. The model starts out attending to an item at random ($y \in 1, 2$), which determines the initial policy space (Figure 2.2B). Assume an example trial where the model attends to item 1 initially ($y = 1$). At trial onset, the decision maker holds the prior belief, such that the particle starts on the origin ($\Delta = 0$, $t_1 = t_2 = 0$) which is within the "accumulate" region. As the model accumulates evidence, the particle will move on a plane perpendicular to $t_2 = 0$, since t_2 remains constant while attending to item 1 (Figure 2.2C, first column). During this time, evidence about the true values of both items will be accumulated, but information regarding item 2 will be significantly noisier (as controlled by κ). Depending on the evidence accumulated regarding both items, the particle may hit the boundary for "choose 1", "choose 2", or "switch (attention)". Assume the particle hits the "switch" boundary, indicating that the model is not confident enough to make a decision after the initial fixation to item 1. In other words, the difference in expected rewards between the two items is too small to make an immediate decision, and it is deemed advantageous to collect more information about the currently unattended item. Now, the model is attending to item 2, and the policy space switches accordingly ($y = 2$). The particle, starting from where it left off, will now move on a plane perpendicular to the t_1 axis (Figure 2.2C, second column). This process is repeated until the particle hits a decision boundary (Figure 2.2C, third column). Importantly, these shifts in attention are endogenously generated by the model as a part of the optimal decision strategy — it exploits its ability to control how much information it receives about either item's value.

The optimal policy space shows some notable properties. As expected, the "switch" region in a given policy space is always encompassed in the "accumulate" region of the other policy space, indicating that the model never switches attention or makes a decision immediately after an attention switch. Furthermore, the decision boundaries in 3D space approach each other over time, consistent with previous work that showed a collapsing 2D boundary for optimal value-based decisions without attention [10]. The collapsing bound reflects the model's uncertainty regarding the difficulty of the decision task [9]. In our case, this difficulty depends on how different the true item values are, as items of very different values are easier to distinguish than those of similar value. If the difficulty is known within and across choices, the boundaries will not collapse over time, and their (fixed) distance will reflect the difficulty of the choice. However, since the difficulty

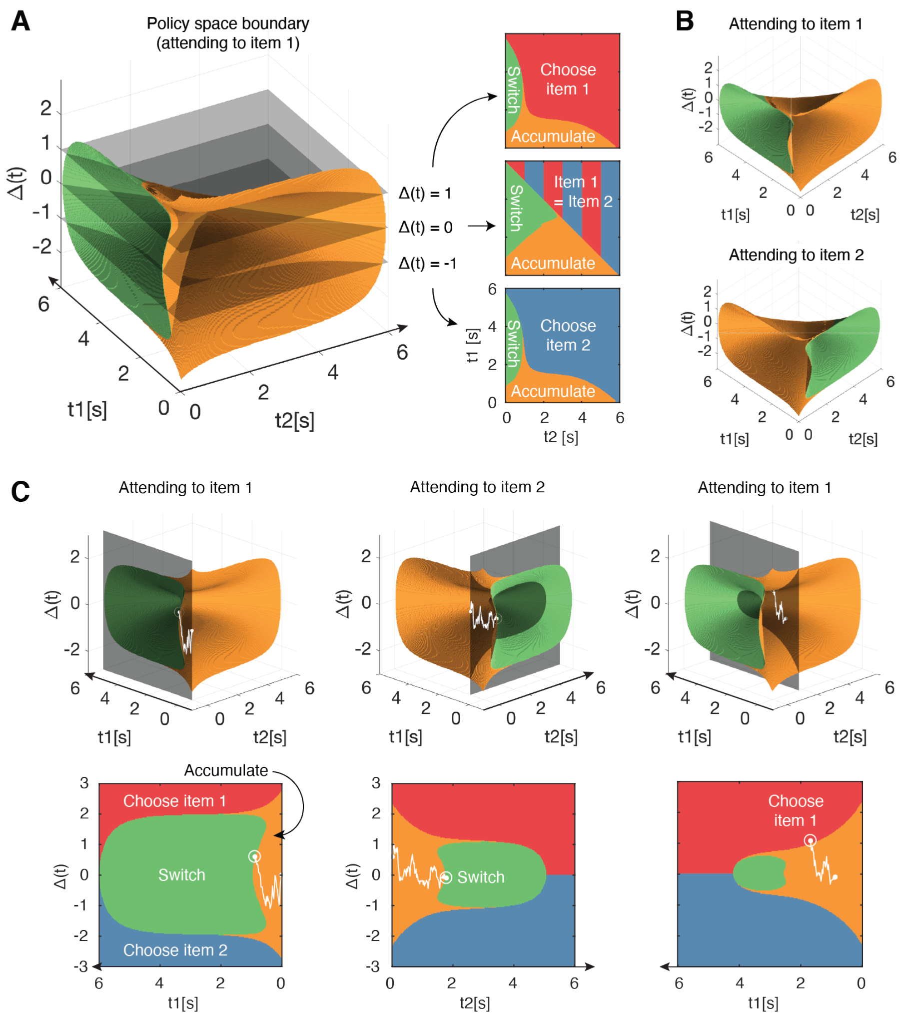

FIGURE 2.2 Navigating the optimal policy space. (A) The optimal policy space. The policy space can be divided into regions associated with different optimal actions (choose item 1 or 2, accumulate more evidence, switch attention). The boundaries between these regions can be visualized as contours in this space. The three panels on the right show cross-sections after slicing the space at different Δ values, indicated by the gray slices in the left panel. Note that when $\Delta = 0$ (middle panel), the two items have equal value and therefore there is no preference for one item over the other. (B) Optimal policy spaces for different values of y (currently attended item). The two policy spaces are mirror-images of each other. (C) Example deliberation process of a single trial demonstrated by a particle that diffuses across the optimal policy space. In this example, the model starts by attending to item 1, then makes two switches in attention before eventually choosing item 1. The bottom row shows the plane in which the particle diffuses. Note that the particle diffuses on the (grey, shaded) plane perpendicular to the time axis of the unattended item, such that it only increases in t_j when attending to item j. Also note that the policy space changes according to the item being attended to, as seen in (B). See results text for more detailed description. See Figure 2.3 to view changes in the optimal policy space depending changes to model parameters.

of individual choices varies and is a priori unknown to the decision maker in our task, the decision boundary collapses so that the model minimizes wasting time on a choice that is potentially too difficult.

The optimal model had five free parameters that affect its behavior: 1) variance of evidence accumulation (σ_x^2), 2) variance of the prior distribution (σ_z^2), 3) cost of evidence accumulation ($c[s^{-1}]$), 4) cost of switching attention (c_s), and 5) relative information gain from the attended vs. unattended items (κ). The contour of the optimal policy boundaries changes in intuitive ways as these parameters are adjusted (Figure 2.3). Increasing the noisiness of evidence accumulation (σ_x^2) causes an overall shrinkage of the evidence accumulation space. This allows the model to reach a decision boundary more quickly under a relatively higher degree of uncertainty, given that evidence accumulation is less reliable but equally costly. Similarly, increasing the cost of accumulating evidence ($c[s^{-1}]$) leads to a smaller accumulation space so that the model minimizes paying a high cost for evidence accumulation. Increasing the switch cost c_s leads to a smaller policy space for the "switch" behavior, since there is an increased cost for switching attention. Similarly, decreasing the inattention noise by setting κ closer to $\frac{1}{2}$ leads to a smaller "switch" space because the model can obtain more reliable information from the unattended item, reducing the necessity to switch attention. To find a set of parameters that best mimic human behavior, we performed a random search over a large parameter space and selected the parameter set that best demonstrated the qualitative aspects of the behavioral data (see Appendix).

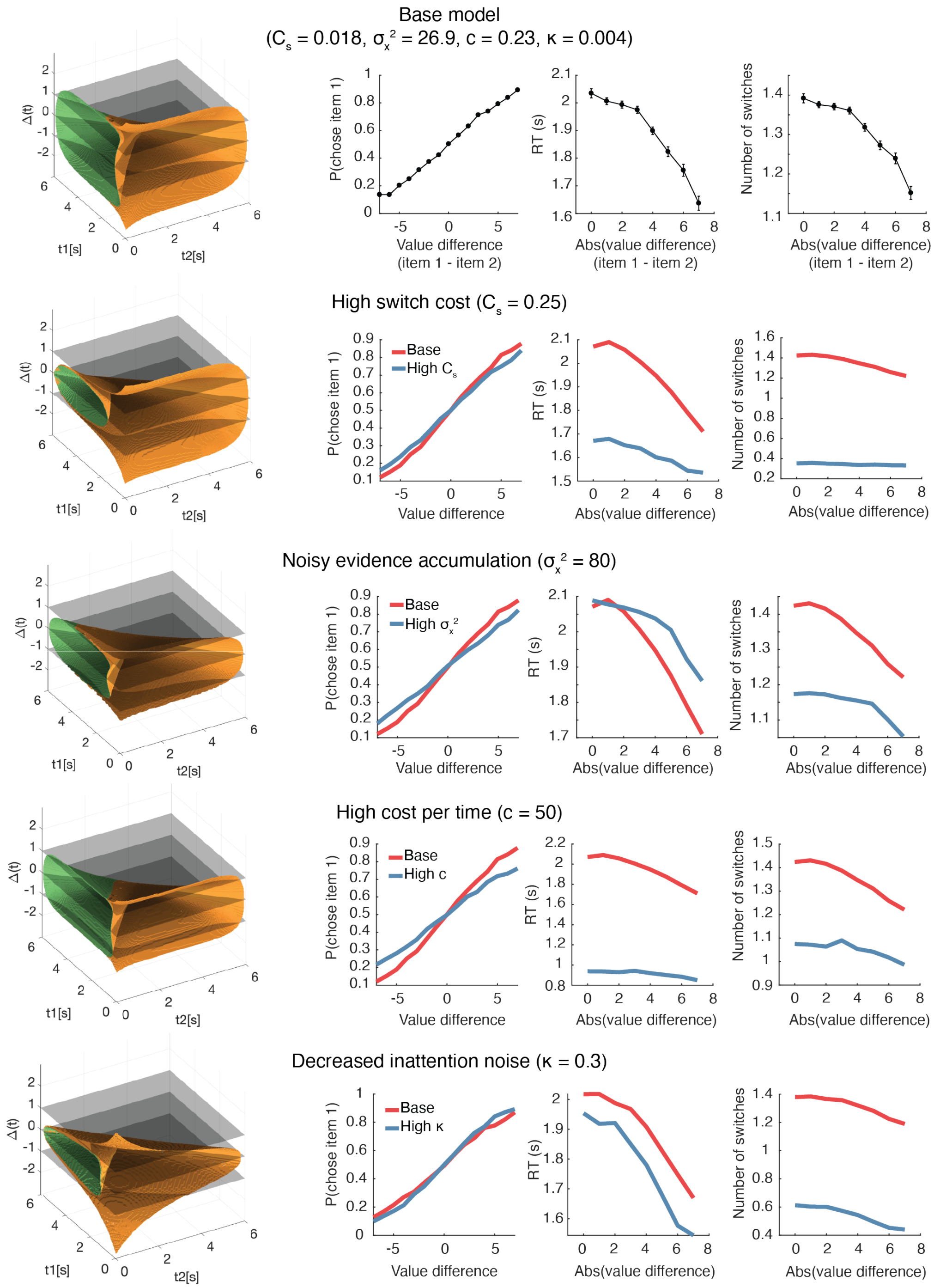

FIGURE 2.3 Changes in the optimal policy space and model behavior with adjustments in free model parameters. The optimal policy space and its associated psychometric curves from the base model is shown in the top row. The policy space and psychometric curves corresponding to changes in single free parameters are shown in subsequent rows. In rows 2-4, psychometric curves from he base model on row 1 is shown in red for comparison. P(choose item 1), probability of choosing item 1; RT, response time.

2.3 The optimal policy replicates human behavior

To assess if the optimal policy features the same decision-making characteristics as human decision makers, we used it to simulate behavior in a task analogous to the value-based decision task performed by humans in Krajbich et al (2010) [2]. Briefly, in this task, participants first rated their preference of different snack items on a scale of -10 to 10. Then, they were presented with pairs of different snacks after excluding the negatively rated items and instructed to choose the preferred item. While they deliberate on their choice, the participants' eye movements were tracked and the fixation duration to each item was used as a proxy for visual attention.

We simulated decision-making behavior using value distributions similar to those used in the human experiment (see Methods), and found that the model behavior qualitatively reproduce essential features of human choice behavior (Figure 2.4). As expected in value-based decisions, a larger value difference among the compared items made it more likely for the model to choose the higher-valued item (Figure 2.4A; $t(38) = 105.7, p < 0.001$). Furthermore, the model's mean response time (RT) decreased with increasing value difference, indicating that less time was spent on trials that were easier (Figure 2.4B; $t(38) = -11.1, p < 0.001$). The model also made less attentional switches for easier trials, indicating that difficult trials required more evidence accumulation from both items, necessitating multiple switches in attention (Figure 2.4C; $t(38) = -8.10, p < 0.001$). Since the number of switches is likely correlated with response time, we also looked at switch rate (number of switches divided by response time). Here, although human data showed no relationship between switch rate and trial difficulty, model behavior showed a positive relationship, suggesting an increased rate of switching for easier trials. However, this effect was absent when using the same number of trials as humans, and did not generalize across all model parameter values (Figure 2.5).

The model also reproduced the biasing effects of fixation on preference seen in humans [2]. An item was more likely to be chosen if it was the last one to be fixated on (Figure 2.4D), and if it was viewed for a longer time period (Figure 2.4E; $t(38) = 5.32, p < 0.001$). Interestingly, the model also replicated a particular fixation pattern seen in humans, where a short first fixation is followed by a significantly longer second fixation, which is followed by a medium-length third fixation (Figure 2.4F). We suspect this pattern arises due to the shape of the optimal decision boundaries, where the particle is more likely to hit the "switch" boundary in a shorter time for the first fixation, likely reflecting the fact that the model prefers to sample from both items at least once. Consistent with this, Figure 2.4C shows that the "accumulate" space is larger for the second fixation compared to the first fixation. Of note, the attentional drift diffusion model (aDDM) that was initially proposed to explain the observed human data [2] did not show this fixation pattern (Figure 2.8D)

One feature that distinguishes our model from previous attention-based decision models is that attention only modulates the variance of momentary evidence without explicitly down-weighting the value of the unattended item [2,37]. Therefore, at first glance, preference for the more-attended item is not an obvious feature since our model does not appear to boost its estimated value. How-

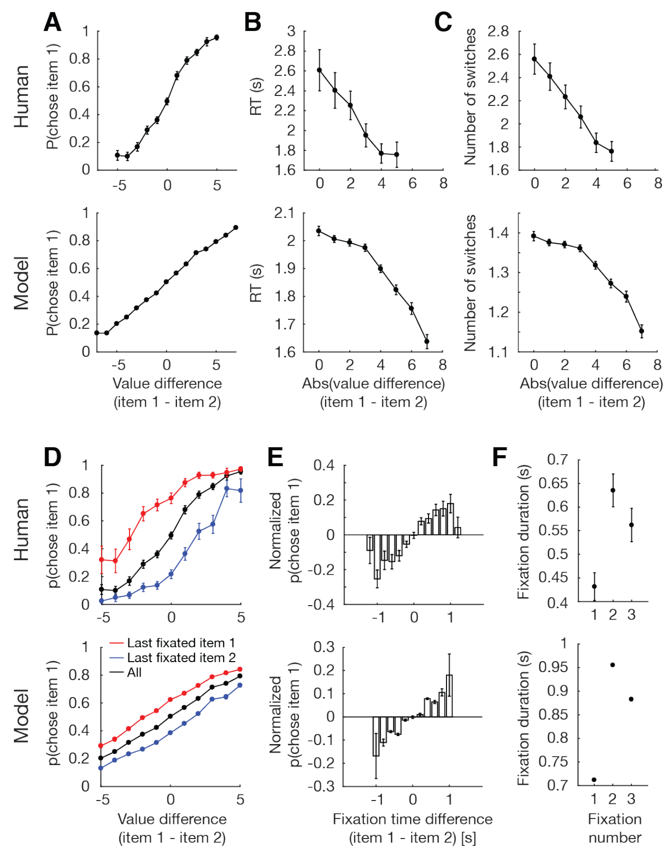

FIGURE 2.4 Replication of human behavior by simulated optimal model behavior [2]. (A) Monotonic increase in probability of choosing item 1 as a function of the difference in value between item 1 and 2 ($t(38) = 105.7, p < 0.001$). (B) Monotonic decrease in response time (RT) as a function of trial difficulty ($t(38) = -11.1, p < 0.001$). RT increases with increasing difficulty. (C) Decrease in the number of attention switches as a function of trial difficulty. More switches are made for harder trials ($t(38) = -8.10, p < 0.001$). (D) Effect of last fixation location on item preference. The item that was fixated on immediately prior to the decision was more likely to be chosen. (E) Attention's biasing effect on item preference. The item was more likely to be chosen if it was attended for a longer period of time ($t(38) = 5.32, p < 0.001$). Since the probability of choosing item 1 depends on the degree of value difference between the two items, we normalized the p(choose item 1) by subtracting the average probability of choosing item 1 for each difference in item value. (F) Replication of fixation pattern during decision making. Both model and human data showed a fixation pattern where a short initial fixation was followed by a long, then medium-length fixation. Error bars indicate SEM across both human and simulated participants ($N = 39$ for both). See Figure 2.9 for an analogous figure for the perceptual decision task. See Figure 2.5 for psychometric curves when evidence accumulation is less noisy, and exploration of switch rate, rather than switch number.

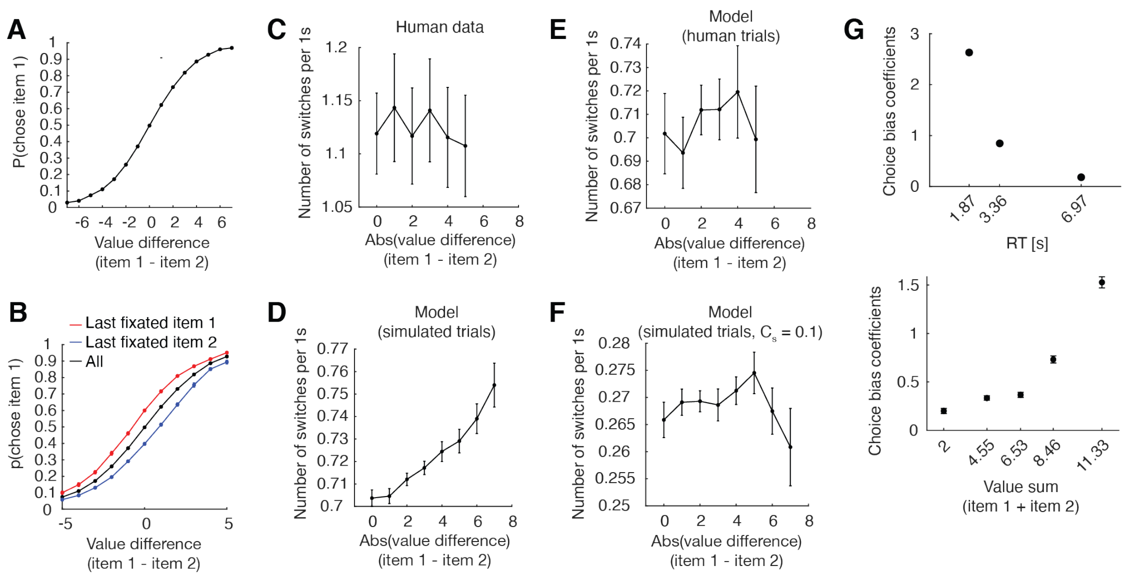

FIGURE 2.5 Psychometric curves with varying parameter values and exploration of switch rate rather than switch number for the optimal model, and choice bias predictions for the aDDM [2]. (A,B) Choice curves after decreasing the evidence noise term (σ^2) from 27 to 5. In Figure 3A,D there seemed to be qualitative difference in the choice curves between human and model behavior, in which model behavior exhibited more linear rather than sigmoid curves. We show that this is a result of the difficulty of the task set by the evidence noise term (σ^2), and not a generalizable property of the model. If we set (σ^2) to a lower value, the model will exhibit sigmoid choice curves because the decision becomes easier at extreme value differences. Consistent with this, the choice curves in A&B show sigmoid curves. (C) Switch rate (number of switches divided by time) as a function of trial difficulty in human data showed no significant relationship (t(38) = -0.32, p = 0.75). (D) The switch rate in the optimal model significantly increases with a decrease in task difficulty (t(38) = 2.96, p = 0.0052). (E) This relationship ceases to be apparent once we reduce the number of simulated trials to that of the human data (t(38) = 1.02, p = 0.31), suggesting the human data may be underpowered to show such a relationship. (F) The relationship between switch rate and trial difficulty is not a general property of the optimal model, as a significant increase in the switch cost (adjusting C_s from 0.018 to 0.1) removes the effect seen in D (t(38) = -0.50, p = 0.62), even with a large number of simulated trials. (G) Effect of RT and value sum on choice bias in the aDDM. The aDDM also replicated the same effects as predicted by the optimal model (RT: t(38) = -48.6, p < 0.001; value sum: t(38) = 14.7, p < 0.001). For both plots, trial data were binned into equally sized bins based on the variable on the x-axis. The plots show the average curves across participants, where vertical error bars indicate SEM for the choice bias coefficient (see Methods for how to compute the choice bias coefficient), and horizontal error bars indicate the SEM of the bin means.

ever, under the assumption that decision-makers start out with a zero-mean prior, Bayesian belief updating with attention modulation turns out to effectively account for a biasing effect of fixation on the subjective value of items [38]. For instance, consider choosing between two items with equal underlying value. Without an attention-modulated process, the model will accumulate evidence from both items simultaneously, and thus have no preference for one item over the other. However, once attention is introduced and the model attends to item 1 longer than item 2, it will have acquired more evidence about item 1's value. This will cause item 1 to have a sharper, more certain likelihood function compared to item 2 (Figure 2.6A). As posterior value estimates are formed by combining priors and likelihoods in proportion to their associated certainties, the posterior of item 1 will be less biased towards the prior than that of item 2. This leads to a higher subjective value of item 1 compared to that of item 2 even though their true underlying values are equal.

This insight leads to additional predictions for how attention-modulated choice bias should vary with certain trial parameters. For instance, the Bayesian account predicts that trials with longer response times should have a weaker choice bias than trials with shorter response times. This is because the difference in fixation times between the two items will decrease over time as the model has more opportunities to switch attention. Both the human and model behavior robustly showed this pattern (Figure 2.6B; human, $t(38) = -3.25, p = 0.0024$; model, $t(38) = -32.0, p < 0.001$). Similarly, choice bias should increase for trials with higher-valued items. In this case, since the evidence distribution is relatively far away from the prior distribution, the posterior distribution is "pulled away" from the prior distribution to a greater degree for the attended versus unattended item, leading to greater choice bias. Both human and model data confirmed this behavioral pattern (Figure 2.6C; human, $t(38) = 2.95, p = 0.0054$; model, $t(38) = 11.4, p < 0.001$). Since response time may be influenced by the sum of the two item values and vice versa, we repeated the above analyses using a regression model that includes both value sum and response time as independent variables (see Methods). The results were largely consistent for both model (effect of RT on choice bias: t(38) = -5.73, p < 0.001, effect of value sum: t(38) = 7.88, p < 0.001) and human (effect of RT: t(38) = -1.32, p = 0.20, effect of value sum: t(38) = 2.91, p = 0.006) behavior.

Next, we assessed how the behavioral predictions arising from the optimal model differed from those of the original attentional drift diffusion model (aDDM) proposed by [2]. Unlike our model, the aDDM follows from traditional diffusion models rather than Bayesian models. It assumes that inattention to an item diminishes its value magnitude rather than increasing the noisiness of evidence accumulation. Despite this difference, the aDDM produced qualitatively similar behavioral predictions as the optimal model (Figure 2.5G, Figure 2.7), although the optimal model was able to better reproduce some of the fixation patterns seen in human behavior Figure 2.8A,D). We also tested to which degree the optimal model yielded a higher mean reward than the aDDM, which, despite its simpler structure, could nonetheless collect competitive amounts of reward. To ensure a fair comparison, we adjusted the aDDM model parameters (i.e., attentional value discounting and the noise variance) so that the momentary evidence provided to the two models has equivalent

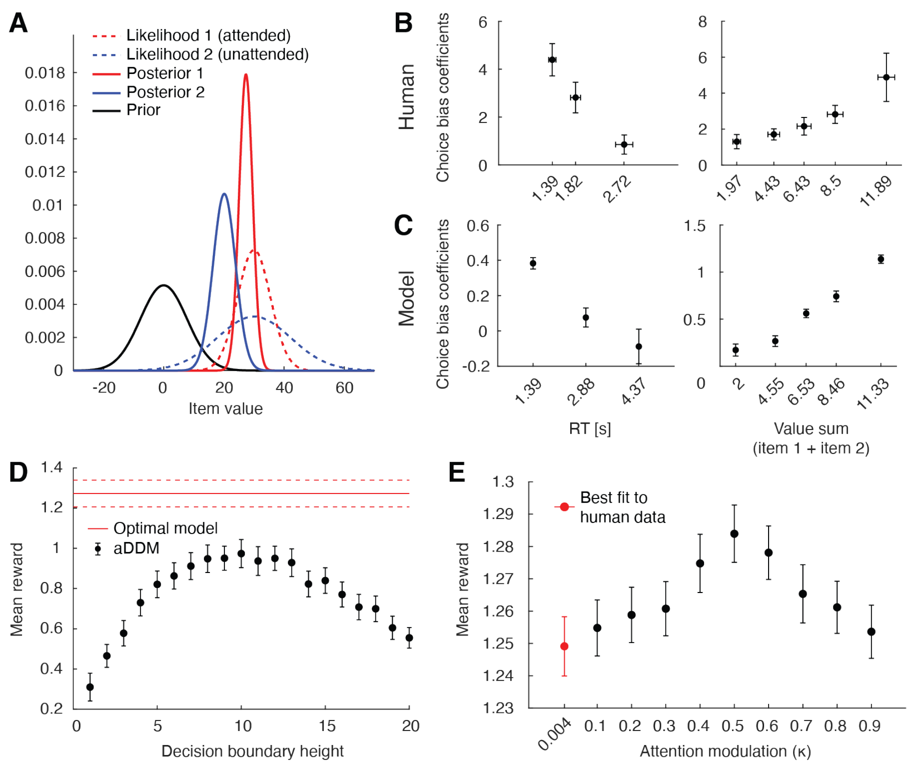

FIGURE 2.6 Behavioral predictions from Bayesian value estimation, and further properties of the optimal policy. (A) Bayesian explanation of attention-driven value preference. Attending to one of two equally-valued items for a longer time (red vs. blue) leads to a more certain (i.e., narrower) likelihood and weaker bias of its posterior towards the prior. This leads to a subjectively higher value for longer-attended item. (B) Effect of response time (RT; left panel; $t(38) = -3.25, p = 0.0024$) and sum of the two item values (value sum; right panel; $t(38) = 2.95, p = 0.0054$) on attention-driven choice bias in humans. This choice bias quantifies the extent to which fixations affect choices for the chosen subset of trials (see Methods) (C) Effect of response time (left panel; $t(38) = -32.0, p < 0.001$) and sum of the two item values (right panel; $t(38) = 11.4, p < 0.001$) on attention-driven choice bias in the optimal model. See Methods for details on how the choice bias coefficients were computed. For (B) and (C), for the left panels, the horizontal axis is binned according to the number of total fixations in a given trial. For the right panels, the horizontal axis is binned to contain the same number of trials per bin. Horizontal error bars indicate SEM across participants of the mean x-values within each bin. Vertical error bars indicate SEM across participants. (D) Comparing decision performance between the optimal policy and the original aDDM model. Performance of the aDDM was evaluated for different boundary heights (error bars = SEM across simulated participants). Even for the reward-maximizing aDDM boundary height, the optimal model significantly outperformed the aDDM ($t(76) = 3.01, p = 0.0027$). (E) Decision performance for different degrees of the attention bottleneck (κ) while leaving the overall input information unchanged (error bars = SEM across simulated participants). The performance peak at $\kappa = 0.5$ indicates that allocating similar amounts of attentional resource to both items is beneficial ($t(38) = -8.51, p < 0.001$).

signal-to-noise ratios (see Appendix). The original aDDM model fixed the decision boundaries at ± 1 and subsequently fit model parameters to match behavioral data. Since we were interested in comparing mean reward, we simulated model behavior using incrementally increasing decision barrier heights, looking for the height that yields the maximum mean reward (Figure 2.6D). We found that even for the best-performing decision barrier height, the aDDM model yielded a significantly lower mean reward compared to that of the optimal model ($t(76) = 3.01, p = 0.0027$).

Recent advances in artificial intelligence used attentional bottlenecks to regulate information flow with significant associated performance gains [39–43]. Analogously, attentional bottlenecks might also be beneficial for value-based decision-making. To test this, we asked if paying relatively full attention on a single item at a time confers any advantages over the ability to pay relatively less reliable, but equal attention to multiple options in parallel. To do so, we varied the amount of momentary evidence provided about both the attended and unattended items while keeping the overall amount of evidence, as controlled by σ_x^2, fixed. This was accomplished by varying the κ term. The effect of κ on the optimal policy was symmetric around $\kappa = 0.5$, such that information gained from attended item at $\kappa = 0.2$ is equal to that of the unattended item at $\kappa = 0.8$. Setting $\kappa = 0.5$ resulted in equal momentary evidence about both items, such that switching attention had no effect on the evidence collected about either item. When tuning model parameters to best match human behavior, we found a low $\kappa \approx 0.004$, suggesting that humans tend to allocate the majority of their presumably fixed cognitive resources to the currently attended item. This allows for reliable evidence accumulation for the attended item, but is more likely to necessitate frequent switching of attention.

To investigate whether widening this attention bottleneck leads to changes in decision performance, we simulated model behavior for different values of κ (0.1 to 0.9, in 0.1 increments). Interestingly, we found that mean reward from the simulated trials is greatest at $\kappa = 0.5$ and decreases for more extreme values of κ, suggesting that a more even distribution of attentional resources between the two items is beneficial for maximizing reward (Figure 2.6E;$t(38) = -8.51, p < 0.001$).

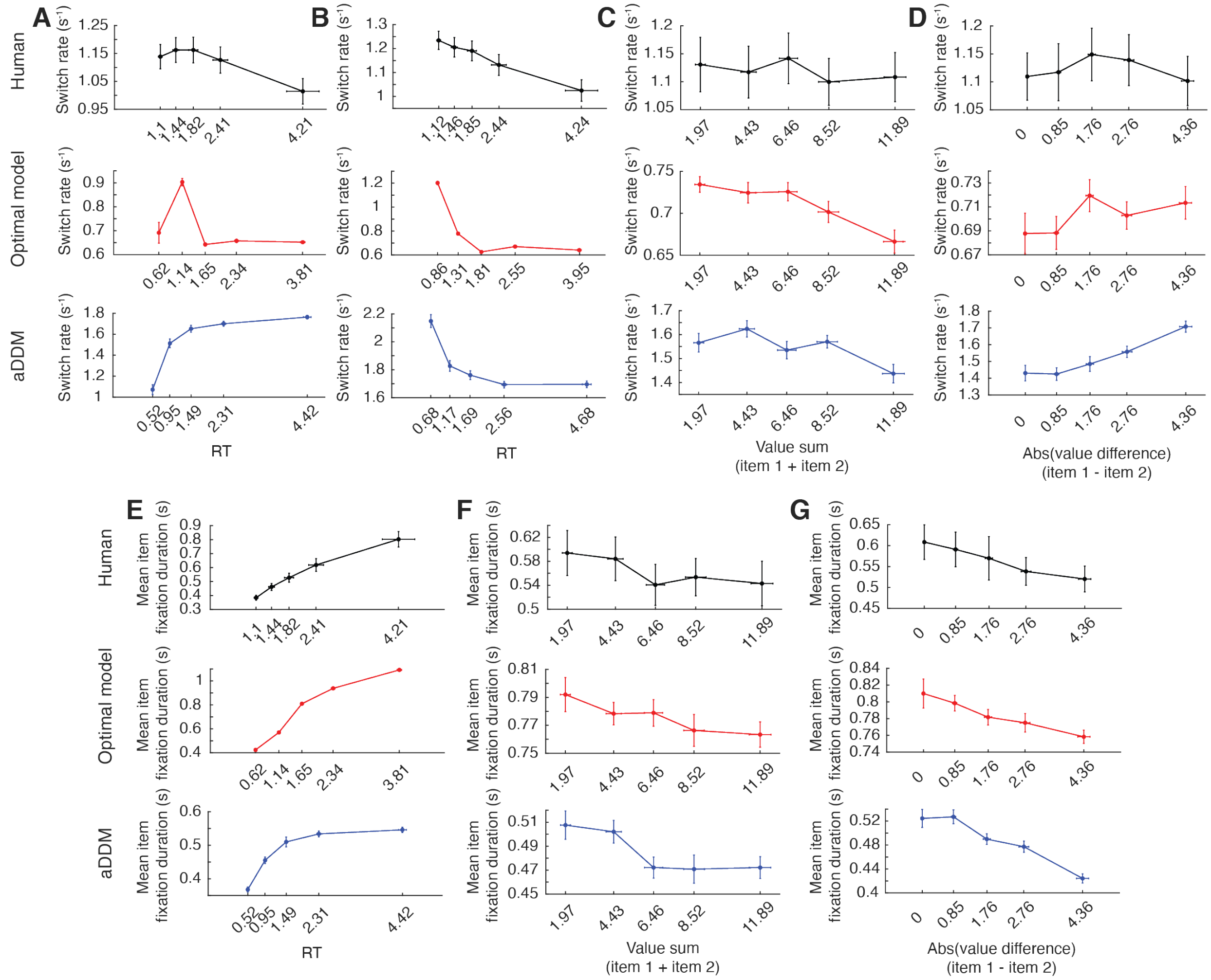

FIGURE 2.7 Comparison of attention switching and fixation behavior between human data, optimal model, and aDDM. For all plots, the trial data were binned into five equally sized bins based on the variable on the x-axis. The plots show the average curves across participants, where vertical error bars indicate SEM for the relevant y-variable (e.g., switch rate), and horizontal error bars indicate the SEM of the bin means. (A) Switch rate as a function of RT. In human data, the probability of switching decreases as a function of time ($t(38) = -4.49$, $p < 0.001$), while this relationship is neither apparent in the optimal model nor the aDDM. (B) When only including trials where at least one switch occurred, both models predicted a decrease in switch rate over time, consistent with human data (optimal model: $t(38) = -29.6$, $p < 0.001$, aDDM: $t(38) = -7.70$, $p < 0.001$). This suggests that in both models, single fixation trials significantly affect the switch rate. (C) Human data showed no significant relationship between switch rate and value sum ($t(38) = -0.84$, $p = 0.40$). However, both the optimal model and the aDDM showed a negative association, such that switch rate decreased as the value sum increased, suggesting that the model is less likely to switch attention within the same time-frame for trials where higher value items are being compared (optimal model, $t(38) = -4.11$, $p < 0.001$; aDDM, $t(38) = -2.09$, $p = 0.044$). (D) Human data again showed no significant relationship between switch rate and absolute value difference (i.e., trial difficulty; $t(38) = -0.67$, $p = 0.51$). The optimal model also showed no significant relationship between switch rate and value difference ($t(38) = -0.41$, $p = 0.68$). However, the aDDM showed a positive association, suggesting that more switches occurred within the same time-frame for easier trials ($t(38) = 4.62$, $p < 0.001$). (E) Human data and both models showed a positive association between mean fixation duration and RT (human: $t(38) = 9.28$, $p < 0.001$; optimal model: $t(38) = 85.6$, $p < 0.001$; aDDM: $t(38) = 13.65$, $p < 0.001$). (F) Human data and both models showed a negative association between mean fixation duration and value sum (human: $t(38) = -2.81$, $p = 0.0078$; optimal model: $t(38) = -4.19$, $p < 0.001$; aDDM: $t(38) = -3.32$, $p = 0.002$). (G) Human data and both models showed a negative association between mean fixation duration and absolute value difference (human: $t(38) = -5.46$, $p < 0.001$; optimal model: $t(38) = -3.60$, $p < 0.001$; aDDM: $t(38) = -6.44$, $p < 0.001$). Of note, for the aDDM simulations in A-G, we used the same parameter setup used in the original paper by [2] rather than the signal-to-noise-matched version we used to compare the mean reward between the optimal model and aDDM.

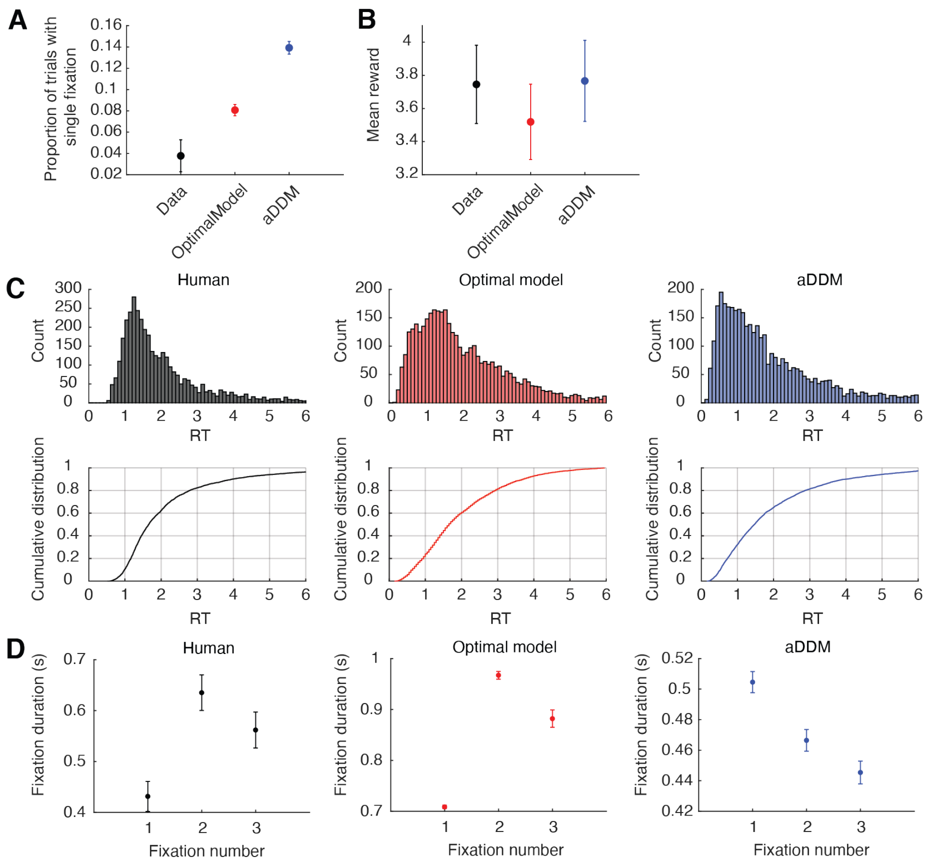

FIGURE 2.8 Additional analyses of fixation behavior, performance, and choice bias between human data, optimal model, and aDDM. (A) Proportion of all trials that ended after a single fixation. While both the optimal model and aDDM featured more single fixation trials than human data, the aDDM predicted significantly more than the optimal model ($t(76) = 5.84$, $p < 0.001$). (B) Comparing the mean reward received by humans versus the two models. There was no significant difference between the mean rewards of humans versus the optimal model ($t(76) = 0.69$, $p = 0.49$) and humans versus the aDDM ($t(76) = -0.062$, $p = 0.95$). Of note, the mean reward of the optimal model was not larger than the aDDM in this scenario because we used the original parameter setup rather than the signal-to-noise matched setup. To calculate mean reward, we used the same cost per unit time used for the optimal model ($c = 0.23$). (C) Distribution of response times. Both models predict a RT distribution that seem to include more <1s RT trials. (D) Mean fixation duration of the first three fixations across participants for humans and both models. The aDDM did not predict the same fixation pattern as the data and optimal model. This fixation pattern in the optimal model is well-preserved across different model parameter values. For the aDDM simulations in A-D, we used the same parameter setup used in the original paper by [2]

.

2.4 Optimal attention-modulated policy for perceptual decisions

The impact of attention is not unique to value-based decisions. In fact, recent work showed that fixation can bias choices in a perceptual decision-making paradigm [44]. In their task, participants were first shown a target line with a certain orientation, then shown two lines with slightly different orientations. The goal was to choose the line with the closest orientation to the previously shown target. Consistent with results in the value-based decision task, the authors demonstrated that the longer-fixated option was more likely to be chosen.

We modified our attention-based optimal policy to perform in such perceptual decisions, in which the goal was to choose the option that is the closest in some quantity to the target, rather than choosing the higher-valued option. Therefore, our model can be generalized to any task that requires a binary decision based on some perceptual quality, whether that involves finding the brighter dot between two dots on a screen, or identifying which of the two lines on the screen is longer. Similar to our value-based case, the optimal policy for perceptual decisions was successful at reproducing the attention-driven biases seen in humans in [44] (Figure 2.9).

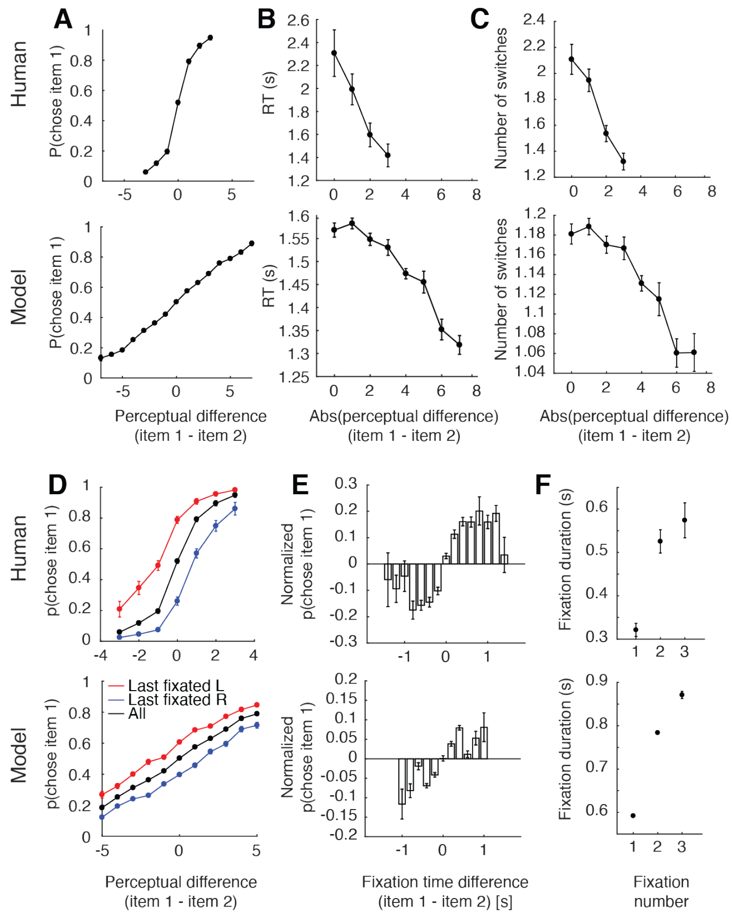

FIGURE 2.9 Replication of human behavior by simulated optimal model in a perceptual decision-making task. This task involves choosing the item with a greater degree of a certain a perceptual quality (e.g., brightness of a dot, angle of a line). Therefore, the decision maker is interested in the difference in the perceptual quality between the two items, rather than their difference in value. (A) Monotonic increase in probability of choosing item 1 as a function of the perceptual difference between item 1 and 2. (B) Decrease in response time (RT) as a function of trial difficulty. (C) Decrease in the number of switches as a function of trial difficulty. (D) Effect of last fixation location on item preference. The item that was fixated on immediately prior to the decision was more likely to be chosen. (E) Attention's biasing effect on item choice. The item was more likely to be chosen if it was attended to for a longer period of time. (F) Replication of fixation pattern during decision making. In the perceptual decision-making task, both model and human data showed increased duration for every subsequent fixation, a notable difference compared to fixation behavior in the value-based task. For (A)-(D), the behavioral data has a smaller range of perceptual differences due to insufficient trials with such large perceptual difference. Error bars indicate SEM across participants for both human and simulated data.

CHAPTER 3

Discussion

In this work, we derive a novel normative decision-making model with an attentional bottleneck, and show that it is able to reproduce the choice and fixation patterns of human decision-makers. Our model significantly extends prior attempts to incorporate attention into perceptual and value-based decision-making in several ways. First, we provide a unified framework in which fixations are endogenously generated as a core component of the normative decision-making strategy. In previous work, fixation patterns were assumed to be either independent of the decision-making strategy [2, 14] or generated by heuristics that relied on features such as the salience or value estimates of the choice options [17, 19]. Other models generated fixations under the assumption that fixation time to different information sources should depend on the expected utility or informativeness of the choice items [18, 37, 45]. For example, [18] assumed that the informativeness of each item differed, which means the model should attend to the less informative item longer in general. Furthermore, since their decision task involved a fixed-duration, attention switches also occurred at fixed times rather than being dynamically adjusted across time, as in our case with a free-response paradigm. A recent normative model supported a continuous change of attention across choice items, and so couldn't relate attention to the observed discrete fixation changes [46]. Our work significantly builds on these prior models by identifying the exact optimal policy using dynamic programming, demonstrating that fixation patterns could reflect active information gathering through controlling an attentional bottleneck. This interpretation extends previous work on visual attention to the realm of value-based and perceptual decision-making [1, 20–23].

Second, our model posits that attention enhances the reliability of information about the attended item [33]. In contrast, previous models accounted for attention by down-weighting the value of the unattended item [2, 14, 37], where one would a priori assume fixations to bias choices. Our approach was inspired by neurophysiological findings demonstrating that visual attention selectively increases the firing rate of neurons tuned to task-relevant stimuli [47], decreases the mean-normalized variance of individual neurons [26,48], and reduces the correlated variability of neurons at the population level [3,24,25]. In essence, selective attention appears to boost the signal-to-noise ratio, or the reliability of information encoded by neuronal signals rather than alter the magnitude of the value encoded by these signals.

Under this framework, we show that the optimal policy can be implemented as a 4-dimensional accumulation-to-bound model where the particle drifts and diffuses according to the fixation duration to either item, the currently attended item, and the difference in items' value estimates. This policy space is significantly more complex compared to previous attention-free normative models, which can be implemented in a 2-dimensional space. Nevertheless, the attention-modulated optimal policy still featured a collapsing boundary in time consistent with the attention-free case [9,10].

When designing our model, we took the simplest possible approach to introduce an attentional bottleneck into normative models of decision-making. When doing so, our aim was to provide a precise (i.e., without approximations), normative explanation for how fixation changes qualitatively interact with human decisions rather than quantitatively capture all details of human behavior, which is likely driven by additional heuristics and features beyond the scope of our model [49, 50]. For instance, it has been suggested that normative allocation of attention should also depend on the item values to eliminate non-contenders, which we did not incorporate as a part of our model [17, 19]. As such, we expect other models using approximations to have a better quantitative fit to human data [2, 51]. Instead, a normative understanding can provide a basis for understanding limitations and biases that emerge in human behavior. Consistent with this goal, we were able to qualitatively capture a wide range of previously observed features of human decisions (Figure 2.4, 2.5 and 2.7), suggest a computational rationale for fixation-based choice biases (Figure 2.6A), and confirm new predictions arising from our theory (Figure 2.6B-C).

Due to the optimal policy's complexity (Figure 2.2), we expect the nervous system to implement it only approximately (e.g., similar to [52] for multi-alternative decisions). Such an approximation has been recently suggested by [51], where they proposed a model of N-alternative choice using approaches from rational inattention to approximate optimal decision-making in the presence of an attentional bottleneck. Unlike our work, they assumed that the unattended item is completely ignored, and therefore could not investigate the effect of graded shifts of attentional resources between items (Figure 2.6E). In addition, their model did not predict a choice bias in binary choices due to a different assumption about the Bayesian prior.

In our model, we assumed the decision maker's prior belief about the item values is centered at zero. In contrast, [51] chose a prior distribution based on the choice set, centered on the average value of only the tested items. While this is also a reasonable assumption [53], it likely contributed to their inability to demonstrate the choice bias for binary decisions. Under the assumption of our zero-mean prior, formulating the choice process through Bayesian inference revealed a simple and intuitive explanation for choice biases (Figure 2.6A) (see also [54]). This explanation required the decision maker to a-priori believe the items' values to be lower than they actually are when choosing between appetitive options, consistent with evidence that item valuations vary inversely with the average value of recently observed items [55]. The zero-mean prior also predicts an opposite effect of the choice bias when deciding between aversive items, such that less-fixated items should become the preferred choice. This is exactly what has been observed in human decision-makers [56]. We justified using a zero-mean bias because participants in the decision task were allowed to rate items as having both positive or negative valence (negative-valence items were excluded from the binary decision task). However, there is some evidence that humans also exhibit choice biases when only choosing between appetitive items [15, 57, 58]. Although our setup suggests a zero-mean prior is required to reproduce the choice bias, the exact features and role of the Bayesian prior in human decisions still remains an open question for future work.

We show that narrowing the attentional bottleneck by setting κ to values closer to 0 or 1 does not boost performance of our decision-making model (Figure 2.6E). Instead, spreading a fixed cognitive reserve evenly between the attended and unattended items maximized performance. This is consistent with prior work that showed that a modified drift diffusion model with a continuously varying attention would perform optimally when attention is always equally divided [59]. However, this does not necessarily imply that equally divided attention always constitutes the normative behavior. If the decision maker has already paid more attention to one item over the other within a decision, it may be optimal to switch attention and gain more information about the unattended item rather than to proceed with equally divided attention.

Parameters fit to human behavior reveal that humans tend to allocate a large proportion of their cognitive resource toward the attended item, suggesting that the benefits of an attentional bottleneck might lie in other cognitive processes. Indeed, machine learning applied to text translation [39,40], object recognition [41,42], and video-game playing [43] benefits from attentional bottlenecks that allow the algorithm to focus resources on specific task subcomponents. For instance, image classification algorithms that extract only the relevant features of an image for high-resolution processing demonstrated improved performance and reduced computational cost compared to those without such attentional features [41]. Similarly, attentional bottlenecks that appear to limit human decision-making performance might have beneficial effects on cognitive domains outside the scope of binary value-based decisions. This is consistent with the idea that the evolutionary advantage of selective attention involves the ability to rapidly fixate on salient features in a cluttered environment, thereby limiting the amount of information that reaches upstream processing and reducing the overall computational burden [60].

An open question is whether our findings can be generalized to multi-alternative choice paradigms [17, 19, 45, 52]. While implementing the optimal policy for such choices may be analytically intractable, we can reasonably infer that a choice bias driven by a zero-mean prior would generalize to decisions involving more than two options. However, in a multialternative choice paradigm where heuristics involving value and salience of items may influence attention allocation, it is less clear whether an equally divided attention among all options would still maximize reward. We hope this will motivate future studies that investigate the role of attention in more realistic decision scenarios.

CHAPTER 4

Methods

Here, we provide an outline of the framework and its results. Detailed derivations are provided in the Appendix.

Attention-modulated decision-making model

Before each trial, z_1 and z_2 are drawn from $z_j \sim (\bar{z}, \sigma_z^2)$. z_1 and z_2 correspond to the value of each item. In each time-step $n > 0$ of duration δt, the decision-maker observes noisy samples of each z_j. This momentary evidence is drawn from $\delta x_{j,n} | z_j \sim \left(z_j \delta t, \frac{1}{1-\kappa} \sigma_x^2 \delta t \right)$ for the attended item $j = y_n$, and $\delta x_{k,n} | z_k \sim \left(z_k \delta t, \frac{1}{\kappa} \sigma_x^2 \delta t \right)$ for the unattended item $k \neq y_n$. We measure how informative a single momentary evidence sample is about the associated true value by computing the Fisher information it provides about this value. This Fisher information sums across independent pieces of information. This makes it an adequate measure for assessing the informativeness of momentary evidence, which we assume to be independent across time and items. Computing the Fisher information results in $(1-\kappa)\sigma_x^{-2}\delta t$ in $\delta x_{j,n}$ about z_j for the attended item, and in $\kappa \sigma_x^{-2} \delta t$ in $\delta x_{k,n}$ about z_k for the unattended item. Therefore, setting $\kappa \leq \frac{1}{2}$ boosts the information of the attended, and reduces the information of the unattended item, while keeping the total information about both items at a constant $(1-\kappa)\sigma_x^{-2}\delta t + \kappa \sigma_x^{-2} \delta t = \sigma_x^{-2} \delta t$. The posterior z_j for $j \in \{1, 2\}$ after $t = N\delta t$ seconds is found by Bayes' rule, $p\left(z_j | \delta x_{j,1:N}, y_{1:N}\right) \propto p(z_j) \prod_{n=1}^{N} p\left(\delta x_{j,n} | z_j, y_n\right)$, which results in Eq. (2.2). If $y_n \in \{1, 2\}$ identifies the attended item in each time-step, the attention times in this posterior are given by $t_1 = \delta t \sum_{n=1}^{N} (2 - y_n)$ and $t_2 = \delta t \sum_{n=1}^{N} (y_n - 1)$. The attention-weighted accumulated evidence is $X_1(t) = \sum_{n=1}^{N} \left(\frac{1-\kappa}{\kappa} \right)^{y_n - 1} \delta x_{1,n}$ and $X_2(t) = \sum_{n=1}^{N} \left(\frac{1-\kappa}{\kappa} \right)^{2 - y_n} \delta x_{2,n}$, down-weighting the momentary evidence for periods when the item is unattended. Fixing $\kappa = 1/2$ recovers the attention-free case of [10], and the associated posterior, Eq. (2.1).

We found the optimal policy by dynamic programming [9, 34], which, at each point in time, chooses the action that promises the larges expected return, including all rewards and costs from that point into the future. Its central component is the value function that specifies this expected return for each value of the sufficient statistics of the task. In our task, the sufficient statistics are the two posterior means, $\langle z_j | X_j(t), t_1, t_2 \rangle$ for $j \in \{1, 2\}$, the two accumulation times, t_1 and t_2, and the currently attended item y_n. The decision maker can choose between four actions at any point in time. The first two are to choose one of the two items, which is expected to yield the corresponding reward, after which the trial ends. The third action is to accumulate evidence for some more time δt, which comes at cost $c\delta t$, and results in more momentary evidence and a corresponding updated posterior. The fourth is to switch attention to the other item $3 - y_n$, which comes at cost $c_s > 0$. As the optimal action is the one that maximizes the expected return, the value for each sufficient statistic is the maximum over the expected returns associated with each action. This leads to the recursive Bellman's equation that relates values with different sufficient statistics (see Appendix for details) and reveals the optimal action for each of these sufficient statistics. Due to symmetries in our task, it turns out these optimal actions only depend on the difference in posterior means Δ, rather than each of the individual means (see Appendix). This allowed us to compute the value function and associated optimal policy in the lower-dimensional (Δ, t_1, t_2, y)-space, an example of which is shown in (Figure 2.2).

The optimal policy was found numerically by backwards induction [10,61], which assumes that at a large enough $t = t_1 + t_2$, a decision is guaranteed and the expected return equals Δ. We set this time point as $t = 6s$ based on empirical observations. From this point, we move backwards in small time steps of $0.05s$ and traverse different values of Δ which was also discretized into steps of 0.05. Upon completing this exercise, we are left with a 3-dimensional grid with the axes corresponding to t_1, t_2 and Δ, where the value assigned to each point in space indicates the optimal decision to take for the given set of sufficient statistics. The boundaries between different optimal actions can be visualized as 3-dimensional manifolds (Figure 2.2).

Model simulations

Using the optimal policy, we simulated decisions in a task analogous to the one humans performed in Krajbich et al., 2010 [2]. On each simulated trial, two items with values z_1 and z_2 are presented. The model attends to one item randomly ($y \in [1, 2]$), then starts accumulating noisy evidence and adjusts its behavior across time according to the optimal policy. Since the human data had a total of 39 participants, we simulated the same number of participants ($N = 39$) for the model, but with a larger number of trials. For each simulated participant, trials consisted of all pairwise combinations of values between 0 and 7, iterated 20 times. This yielded a total of 1280 trials per simulated participant.

When computing the optimal policy, there were several free parameters that determined the shape of the decision boundaries. Those parameters included the evidence noise term (σ_x^2), spread of the prior distribution (σ_z^2), cost of accumulating evidence ($c[s^{-1}]$), cost of switching attention (c_s), and the relative information gain for the attended vs. unattended items (κ). In order to find a set of parameters that best mimics human behavior, we performed a random search over a large parameter space and simulated behavior using the randomly selected set of parameters [62]. We iterated this process for 2,000,000 sets of parameters and compared the generated behavior to that of humans (see Appendix 1). After this search process, the parameter set that best replicated human behavior consisted of $c_s = 0.0065$, $c = 0.23$, $\sigma_x^2 = 27$, $\sigma_z^2 = 18$, $\kappa = 0.004$.

Statistical analysis

The relationship between task variables (e.g., difference in item value) and behavioral measurements (e.g., response time) were assessed by estimating the slope of the relationship for each participant. For instance, to investigate the association between response times and absolute value difference (Figure 2.4B), we fit a linear regression within each participant using the absolute value difference and response time for every trial. Statistical testing was performed using one-sample t-tests on the regression coefficients across participants. This procedure was used for statistical testing involving Figure 2.4B,C,E, and 2.6B,C. To test for the effect of RT and value sum on choice bias after accounting for the other variable, we used a similar approach and used both RT and value sum as independent variables in the regression model and the choice bias coefficient as the dependent variable. To test for a significant peak effect for Figure 2.6E, we used the same procedure after subtracting 0.5 from the original κ values. To compare performance between the optimal model and the aDDM (Figure 2.6D), we first selected the best-performing aDDM model, then performed an independent-samples t-test between the mean rewards from simulated participants from both models.

To quantify the degree of choice bias (Figure 2.6B,C), we computed a choice bias coefficient. For a given group of trials, we performed a logistic regression with fixation time difference ($t_1 - t_2$) as the independent variable and a binary dependent variable indicating whether item 1 was chosen on each trial. After performing this regression within each participant's data, we performed a t-test of the regression coefficients against zero. The the resulting t-statistic was used as the choice bias coefficient, as it quantified the extent to which fixations affected choice in the given subset of trials.

CHAPTER 5

Summary

Every day, we make decisions based on our subjective preferences, such as choosing which dish to order at a restaurant. When comparing between different options, we tend to shift our gaze between different items, evaluating each option before moving on to the next, rather than assess multiple options simultaneously. We may also fixate on certain options for longer periods of time if we are uncertain or less familiar with them (e.g., a special menu). This allows us to reliably dedicate attention to each individual option until we feel sufficiently informed to make a decision. While this is a seemingly effortless, automatic endeavor, it is unknown what computational principles underlie the brain's ability to gather information about multiple items in a statistically optimal manner.

Our work is aimed at developing a computational model of such a decision-making process. To do so, we developed a normative model that incorporates visual fixations as a part of the decision strategy. Importantly, the model accumulated information about a fixated option with higher fidelity compared to the unattended option, consistent with evidence that attention can sharpen our perception about the environment. Therefore, our model endogenously controls its own visual fixations (i.e., gathers information) in a way that allows for the best decisions in the shortest amount of time, similar to decision processes in humans.

Using an optimization method called dynamic programming, we derived the complex optimal decision strategy that allows the model to decide, across time, whether to 1) make a choice, 2) accumulate more evidence, or 3) switch attention to a different item. We found that the model successfully incorporates attentional shifts as a part of the optimal decision strategy, and that this strategy featured the same choice bias as observed in human decision makers: it switched attention more frequently when deciding between items with similar values, and was biased towards choosing items that were attended last, and attended longer. The model also provided insight into why choice biases may in fact be rational, using a Bayesian explanation. This insight also led to novel behavioral predictions which we confirmed in human data, suggesting that decision-making in our model and the human brain may involve similar computational principles.

CHAPTER 6

Appendix

Here we describe in more detail the derivations of our results, and specifics of the simulations presented in the main text. Of note, we sometimes use $x|y \sim p(y)$ to specify the conditional density $p(x|y)$. Furthermore, $\mathcal{N}(\mu, \sigma^2)$ denotes a Gaussian with mean μ and variance σ^2.

6.1 Task setup

6.1.1 Latent state prior

We assume two latent states z_j, $j \in \{1, 2\}$, (here, the true item values) that are before each choice trial drawn from their Gaussian prior, $z_j \sim \mathcal{N}(\bar{z}_j \sigma_z^2)$, with mean $\bar{z}_j$ and variance σ_z^2. Throughout the text, we will assume $\bar{z}_1 = \bar{z}_2$, to indicate that there is no a-priori preference of one item over the other.

6.1.2 Likelihood function of momentary evidence

The decision maker doesn't observe the latent states, but instead, in each time step of size δt, observes noisy evidence about both z_j's. Let us assume that, in the nth such time step, the decision maker attends to item $y_n \in \{1, 2\}$. Then, they simultaneously observe δx_1 and δx_2, distributed as

$$\delta x_{j,n} | y_n, z_j \sim \mathcal{N}\left(z_j \delta t, \left(\frac{1-\kappa}{\kappa} \right)^{|j-y_n|} \frac{\sigma_x^2}{1-\kappa} \delta t \right), \tag{6.1}$$

where we have defined the attention modulation parameter κ, bounded by $0 \leq \kappa \leq 1$ (we will usually assume $\kappa \leq \frac{1}{2}$), and the overall likelihood variance σ_x^2. For the attended item $j = y_n$, we have $|j - y_n| = 0$, such that the the variance of the momentary evidence for this item is $\sigma_x^2 \delta t / (1 - \kappa)$. For the unattended item, for which $|j - y_n| = 1$, this variance is instead $\sigma_x^2 \delta t / \kappa$. As long as $\kappa < \frac{1}{2}$ this leads to a larger variance for the unattended item than the attended item, making the momentary evidence more informative for the attended item. In particular if we quantify this information by the Fisher information in the momentary evidence $\delta x_{j,n}$ about z_j, then we find this information to be $(1 - \kappa)\sigma_x^{-2} \delta t$ for the attended, and $\kappa \sigma_x^{-2} \delta t$ for the unattended item. The total Fisher information across both items is thus $\sigma_x^{-2}\delta t$, independent of κ. This shows that σ_x^2 controls the total information that the momentary evidence provides about the latent states, whereas κ controls how much of this information is provided for the attended vs. the unattended item.

6.1.3 An alternative form for the likelihood

While the above form of the likelihood has a nice, intuitive parametrization, it is notationally cumbersome. Therefore, we will here introduce an alternative variance parametrization of this likelihood that simplifies the notation in the derivations that follow. We will use this parametrization for the rest of this Appendix.

This alternative parametrization assumes the variance of the momentary evidence of the attended item to be given by $\sigma^2\delta t = \sigma_x^2/(1-\kappa)$, while that of the unattended item is given by $\gamma^{-1}\sigma^2\delta t = \sigma_x^2\delta t/\kappa$, where the new attention modulation parameter γ is assumed bounded by $0 \leq \gamma \leq 1$. Thus, the previous parameter pair $\{\sigma_x^2, \kappa\}$ is replaced by the new pair $\{\sigma^2, \gamma\}$. A $\gamma < 1$ results in an increased variance for the unattended item, resulting in less information about the value of the unattended item. Overall, the momentary evidence likelihood is given with the alternative parametrization by

$$\delta x_{j,n}|y_n, z_j \sim \mathcal{N}\left(z_j\delta t, \frac{1}{\gamma^{|j-y_n|}}\sigma^2\delta t\right), \tag{6.2}$$

This is the likelihood function that we will use for the rest of this Appendix. Any of the results can easily be mapped back to the original parametrization (as used in the main text) by

$$\sigma^2 = \frac{\sigma_x^2}{1-\kappa}, \qquad \sigma_x^2 = (1-\kappa)\sigma^2, \tag{6.3}$$

$$\gamma = \frac{\kappa}{1-\kappa}, \qquad \kappa = \frac{\gamma}{\gamma+1}. \tag{6.4}$$

Note that the alternative parametrization does not preserve the separation between total information and balancing the information between the attended and unattended item. In particular, the total Fisher information is now given by $(\gamma+1)\sigma^{-2}\delta t$, which depends on both γ and σ^2.

Below we will derive the posterior z_j's, given the stream of momentary evidences $[\delta x_{1,1}, \delta x_{2,1}], [\delta x_{1,2}, \delta x_{2,2}], \dots$, and the attention sequence $y_1, y_2, \dots$. The mean and variance of the posterior distributions represent the decision maker's belief of the items' true values given all available evidence.

6.1.4 Costs, rewards, and the decision-maker's overall aim

While the posterior estimates provide information about value, it does not tell the decision maker when to stop accumulating information, or when to switch their attention. To address these questions, we need to specify the costs and rewards associated with these behaviors. For value-based decisions, we assume that the reward for choosing item j is the latent state z_j (i.e., the true value) associated with the item. Furthermore, we assume that accumulating evidence comes at cost c per second, or cost $c\delta t$ per time step. The decision maker can only ever attend to one item, and switching attention to the other item comes at cost c_s which may be composed of a pure attention switch cost, as well as a loss of time that might introduce an additional cost. As each attention switch introduces both costs, we only consider them in combination without loss of generality.

The overall aim of the decision maker is to maximize the total expected return, which consists of the expected value of the chosen item minus the total cost of accumulating evidence and attention switches. We address this maximization problem by finding the optimal policy that, based on the observed evidence, determines when to switch attention, when to accumulate more evidence, and when to commit to a choice. We initially focus on maximizing the expected return in a single, isolated choice, and will later show that this yields qualitatively similar policies as when embedding this choice into a longer sequence of comparable choices.

6.2 Bayes-optimal evidence accumulation

6.2.1 Deriving the posterior z_1 and z_2

To find the posterior over z_1 after having accumulated evidence $x_{1,1:N} \equiv x_{1,1}, \ldots, x_{1,N}$ for some fixed amount of time $t = N\delta t$ while paying attention to items $y_{1:N} \equiv y_1, \ldots y_N$, we employ Bayes' rule,

$$\begin{aligned} p\left(z_1 | \delta x_{1,1:N}, y_{1:N}\right) &\propto_{z_1} p(z_1) \prod_{n=1}^{N} p\left(\delta x_{1,n} | z_1, y_n\right) \\ &= \mathcal{N}\left(\bar{z}_1, \sigma_z^2\right) \prod_{n=1}^{N} \mathcal{N}\left(z\delta t, \frac{\sigma^2}{\gamma^{|1-y_n|}} \delta t\right) \\ &\propto_{z_1} \mathcal{N}\left(\frac{\bar{z}_1 \sigma^2 \sigma_z^{-2} + X_1(t)}{\sigma^2 \sigma_z^{-2} + t_1 + \gamma t_2}, \frac{\sigma^2}{\sigma^2 \sigma_z^{-2} + t_1 + \gamma t_2}\right), \end{aligned} \tag{6.5}$$

where we have defined $X_1(t) = \sum_{n=1}^{N} \gamma^{|1-y_n|} \delta x_{1,n}$ as the sum of all attention-weighted momentary evidence up to time t, and $t_j = t - \delta t \sum_{n=1}^{N} |j - y_n|$ as the total time that item j has been attended. Note that, for time periods in which item 2 is attended to, (i.e., when $y_n = 2$), the momentary evidence is down-weighted by γ. With $\delta t \to 0$, the process becomes continuous in time, such that $X_1(t)$ becomes the integrated momentary evidence, but the above posterior still holds.

Following a similar derivation, the posterior belief about z_2 results in

$$p\left(z_2 | \delta x_{2,1:N}, y_{1:N}\right) = \mathcal{N}\left(\frac{\bar{z}_2 \sigma^2 \sigma_z^{-2} + X_2(t)}{\sigma^2 \sigma_z^{-2} + \gamma t_1 + t_2}, \frac{\sigma^2}{\sigma^2 \sigma_z^{-2} + \gamma t_1 + t_2}\right) \tag{6.6}$$

where $X_2(t) = \sum_{n=1}^{N} \gamma^{|2-y_n|} \delta x_{2,n}$. As the decision maker acquires momentary evidence independently for both items, the two posteriors are independent of each other, that is $p(z_1, z_2 | \delta x_{1,1:N}, \delta x_{2,1:N}, y_{1:N}) = p(z_1 | \delta x_{1,1:N}, y_{1:N})\, p(z_2 | \delta x_{2,1:N}, y_{1:N})$.

6.2.2 The expected reward process

At each point in time, the decision maker must decide whether it's worth accumulating more evidence versus choosing an item. To do so, they need to predict how the mean estimated reward for each option might evolve if they accumulated more evidence. In this section we derive the stochastic process that describes this evolution for item 1. The same principles will apply for item 2.

Assume that having accumulated evidence until time $t = N\delta t$, the current expected reward for item 1 is given by $\hat{r}_1(t)$, where $\hat{r}_1(t) = \langle z_1 | \delta x_{1,1:N}, y_{1:N} \rangle$ is the mean of the above posterior, Eq. (6.5). The decision-maker's prediction of how the expected reward might evolve after accumulating additional evidence for δt is found by the marginalization,

$$
\begin{aligned}
&p(\hat{r}_1(t+\delta t) | \hat{r}_1(t), t_1, t_2, y_{N+1}) \\
&= \iint p(\hat{r}_1(t+\delta t) | \hat{r}_1(t), \delta x_{1,N+1}, t_1, t_2, y_{N+1})\, p(\delta x_{1,N+1} | z_1, y_{N+1})\, p(z_1 | \hat{r}_1(t), t_1, t_2)\, \mathrm{d}\delta x_{1,N+1} \mathrm{d}z_1.
\end{aligned}
\tag{6.7}
$$

As the last term in the above integral shows, $\hat{r}(t)$, t_1 and t_2 fully determine the posterior z_1 at time t. We can use this posterior to predict the value of the next momentary evidence $\delta x_{1,N+1} | z_1$. This, in turn, allows us to predict $\hat{r}_1(t+\delta t)$. As all involved densities are either deterministic or Gaussian, the resulting posterior will be Gaussian as well. Thus, rather than performing the integrals explicitly, we will find the final posterior by tracking the involved means and variances, which in turn completely determine the posterior parameters.

We first marginalize over $\delta x_{1,N+1}$, by expressing $\hat{r}_1(t+\delta t)$ in terms of $\hat{r}(t)$ and $\delta x_{1,N+1}$. To do so, we use Eq. (6.5) to express $\hat{r}_1(t+\delta t)$ by

$$
\hat{r}_1(t+\delta t) = \frac{\bar{z}_1 \sigma^2 \sigma_z^{-2} + X_1(t) + \gamma^{|y_{N+1}-1|} \delta x_{1,N+1}}{\sigma^2 \sigma_z^{-2} + t_1 + \gamma t_2 + \gamma^{|1-y_{N+1}|} \delta t},
\tag{6.8}
$$

where we have used $X_1(t+\delta t) = X_1(t) + \gamma^{|y_{N+1}-1|} \delta x_{1,N+1}$.

Note that, for a given $\delta x_{1,N+1}$, $\hat{r}(t+\delta t)$ is uniquely determined by $\hat{r}(t)$. $\hat{r}(t+\delta t)$ becomes a random variable once we acknowledge that, for any z_1, $\delta x_{1,N+1}$ is given by Eq. (6.2), which we can write as $\delta x_{1,N+1} = z_1 \delta t + \sqrt{\sigma^2 \gamma^{-|1-y_{N+1}|} \delta t}\, \eta_x$, where $\eta_x \sim \mathcal{N}(0,1)$. Substituting this expression into $\hat{r}_1(t+\delta t)$, and using Eq. (6.5) to re-express $X_1(t)$ as

$X_1(t) = \hat{r}_1(t)\left(\sigma^2\sigma_z^{-2} + t_1 + \gamma t_2\right) - \bar{z}_1\sigma^2\sigma_z^{-2}$, results in

$$\hat{r}_1(t+\delta t) = \frac{\hat{r}_1(t)\left(\sigma^2\sigma_z^{-2} + t_1 + \gamma t_2\right) + \gamma^{|1-y_{N+1}|} z_1 \delta t + \sqrt{\sigma^2 \gamma^{|1-y_{N+1}|}\delta t}\,\eta_x}{\sigma^2\sigma_z^{-2} + t_1 + \gamma t_2 + \gamma^{|1-y_{N+1}|}\delta t}. \tag{6.9}$$

The second marginalization over z_1 is found by noting the distribution of z_1 is given by Eq. (6.5), which can be written as

$$z_1 = \hat{r}_1(t) + \sqrt{\frac{\sigma^2}{\sigma^2\sigma_z^{-2} + t_1 + \gamma t_2}}\,\eta_z, \tag{6.10}$$

with $\eta_z \sim \mathcal{N}(0,1)$. Substituting this z_1 into the above expression for $\hat{r}(t+\delta t)$ results in

$$\hat{r}_1(t+\delta t) = \hat{r}_1(t) + \frac{\sqrt{\sigma^2\gamma^{|1-y_{N+1}|}\delta t}}{\sigma^2\sigma_z^{-2} + t_1 + \gamma t_2 + \gamma^{|1-y_{n+1}|}\delta t}\eta_x, \tag{6.11}$$

where we have dropped the η_z-dependent term which had a δt pre-factor, and thus vanishes with $\delta t \to 0$. Therefore, $\hat{r}_1(t)$ evolves as a martingale,

$$\hat{r}_1(t+\delta t)|\hat{r}_1(t), t_1, t_2, y_{N+1} \sim \mathcal{N}\left(\hat{r}_1(t), \frac{\sigma^2\gamma^{|1-y_{n+1}|}}{(\sigma^2\sigma_z^{-2} + t_1 + \gamma t_2 + \gamma^{|1-y_{N+1}|}\delta t)^2}\delta t\right). \tag{6.12}$$

Using the same approach, the expected future reward for item 2 is given by

$$\hat{r}_2(t+\delta t)|\hat{r}_2(t), t_1, t_2, y_{N+1} \sim \mathcal{N}\left(\hat{r}_2(t), \frac{\sigma^2\gamma^{|2-y_{N+1}|}}{(\sigma^2\sigma_z^{-2} + \gamma t_1 + t_2 + \gamma^{|2-y_{n+1}|}\delta t)^2}\delta t\right). \tag{6.13}$$

6.2.3 The expected reward difference process

In a later section, we will reduce the dimensionality of the optimal policy space by using the expected reward difference rather than each of the of the expected rewards separately. To do so, we define this difference by

$$\Delta(t) = \frac{\hat{r}_1(t) - \hat{r}_2(t)}{2}. \tag{6.14}$$

As for $\hat{r}_1(t)$ and $\hat{r}_2(t)$, we are interested in how $\Delta(t)$ evolves over time.

To find $\Delta(t+\delta t)|\Delta(t), t_1, t_2, y_{N+1}$ we can use

$$\begin{aligned} &p\left(\Delta(t+\delta t)|\Delta(t), t_1, t_2, y_{N+1}\right) \\ &\qquad = p\left(\Delta(t+\delta t) = \frac{\hat{r}_1(t+\delta t) - \hat{r}_2(t+\delta t)}{2} | \Delta(t) = \frac{\hat{r}_1(t) - \hat{r}_2(t)}{2}, t_1, t_2, y_{N+1}\right). \end{aligned} \tag{6.15}$$

As the decision-maker receives independent momentary evidence for each item, $\hat{r}_1(t)$ and $\hat{r}_2(t)$ are independent when conditioned on t_1, t_2 and $y_{1:N}$. Thus, so are their time-evolutions,

$\hat{r}_1(t+\delta t)|\hat{r}_1(t),\dots$ and $\hat{r}_2(t+\delta t)|\hat{r}_2(t),\dots$. With this, we can show that

$$\begin{aligned}&\Delta(t+\delta t)|\Delta(t),t_1,t_2,y_{N+1}\\&\sim\mathcal{N}\left(\Delta(t),\frac{\sigma^2\delta t}{4}\left(\frac{\gamma^{|1-y_{N+1}|}}{\left(\sigma^2\sigma_z^{-2}+t_1+\gamma t_2+\gamma^{|1-y_{N+1}|}\delta t\right)^2}+\frac{\gamma^{|2-y_{N+1}|}}{\left(\sigma^2\sigma_z^{-2}+\gamma t_1+t_2+\gamma^{|2-y_{N+1}|}\delta t\right)^2}\right)\right).\end{aligned} \tag{6.16}$$

Unsurprisingly, $\Delta(t)$ is again a martingale.

6.3 Optimal decision policy

6.3.1 Optimal policy for value-based decisions

We find the optimal decision policy by dynamic programming [34, 35]. A central concept in dynamic programming is the *value function* $V(\cdot)$, which, at any point in time during a decision, returns the *expected return*, which encompasses all expected rewards and costs from that point onwards into the future when following the optimal decision policy. Bellman's equation links value functions across consecutive times, and allows finding this optimal decision policy recursively. In what follows, we first focus on Bellman's equation for single, isolated choices.

For a single, isolated choice, accumulating evidence comes at cost c per second. Switching attention comes at cost c_s. The expected reward for choosing item j is $\hat{r}_j(t)$, and is given by the mean of Eqs. (6.5) and (6.6) for $j=1$ and $j=2$, respectively.

To find the value function, let us assume that we have accumulated evidence for some time $t=t_1+t_2$, expect rewards $\hat{r}_1(t)$ and $\hat{r}_2(t)$, and are paying attention to item $y\in\{1,2\}$. These statistics fully describe the evidence accumulation state, and thus fully parameterize the value function $V_y(\hat{r}_1,\hat{r}_2,t_1,t_2)$. Here we use y as a subscript rather than an argument to $V(\cdot)$ to indicate that y can only take one of two values, $y\in\{1,2\}$. At this point, we can choose among four actions. We can either immediately choose item 1, immediately choose item 2, accumulate more evidence without switching attention, or switch attention to the other item, $3-y$. The expected return for choosing immediately is either $\hat{r}_1(t)$ or $\hat{r}_2(t)$, depending on the choice. Accumulating more evidence for some time δt results in cost $c\delta t$, and changes in the expected rewards according to $\hat{r}_j(t+\delta t)|\hat{r}_j(t),t_1,t_2,y$, as given by Eqs. (6.12) and (6.13). Therefore, the expected return for accumulating more evidence is given by

$$-c\delta t+\left\langle V_y\left(\hat{r}_1(t+\delta t),\hat{r}_2(t+\delta t),t_1+|2-y|\delta t,t_2+|1-y|\delta t\right)|\hat{r}_1,\hat{r}_2,t_1,t_2,y\right\rangle, \tag{6.17}$$

where the expectation is over the time-evolution of $\hat{r}_1$ and $\hat{r}_2$, and $t_1+|2-y|\delta t$ and $t_2+|1-y|\delta t$ ensures that only the t_y associated with the currently attended item is increased by δt. Lastly, switching attention comes at cost c_s, but does not otherwise impact reward expectations, such that the expected return associated with this action is

$$-c_s + V_{3-y}\left(\hat{r}_1, \hat{r}_2, t_1, t_2\right), \tag{6.18}$$

where the use of $V_{3-y}(\cdot)$ implements that, after an attention switch, item $3-y$ will be the attended item.

By the Bellman optimality principle [34], the best action at any point in time is the one that maximizes the expected return. Combining the expected returns associated with each possible action results in Bellman's equation

$$\begin{aligned} &V_y\left(\hat{r}_1, \hat{r}_2, t_1, t_2\right) \\ &= \max \left\{ \begin{array}{c} \hat{r}_1, \hat{r}_2, \\ \left\langle V_y\left(\hat{r}_1(t+\delta t), \hat{r}_2(t+\delta t), t_1+|2-y|\delta t, t_2+|1-y|\delta t\right) \middle| \hat{r}_1, \hat{r}_2, t_1, t_2, y \right\rangle - c\delta t, \\ V_{3-y}\left(\hat{r}_1, \hat{r}_2, t_1, t_2\right) - c_s \end{array} \right\}. \end{aligned} \tag{6.19}$$

Solving this equation yields the optimal policy for any combination of $\hat{r}_1$, $\hat{r}_2$, t_1, t_2 and y by picking the action that maximizes the associated expected return, that is, the term that maximizes the left-hand side of the above equation. The optimal decision boundaries that separate the $(\hat{r}_1, \hat{r}_2, t_1, t_2, y)$-space into regions where different actions are optimal lie at manifolds in which two actions yield the same expected return. For example, the decision boundary at which it becomes best to choose item 1 after having accumulated more evidence is the manifold at which

$$\begin{aligned} &V_y\left(\hat{r}_1, \hat{r}_2, t_1, t_2\right) = \hat{r}_1 \\ &\quad = \left\langle V_y\left(\hat{r}_1(t+\delta t), \hat{r}_2(t+\delta t), t_1+|2-y|\delta t, t_2+|1-y|\delta t\right) \middle| \hat{r}_1, \hat{r}_2, t_1, t_2, y \right\rangle - c\delta t. \end{aligned} \tag{6.20}$$

In Section 6.4 we describe how we found these boundaries numerically.

Formulated so far, the value function is five-dimensional, with four continuous ($\hat{r}_1$, $\hat{r}_2$, t_1, and t_2) and one discrete (y) dimension. It turns out that it is possible to remove one of the dimensions without changing the associated policy by focusing on the expected reward difference $\Delta(t)$, Eq. (6.14), rather than the individual expected rewards. To show this, we use the value function property $V_y\left(\hat{r}_1, \hat{r}_2, t_1, t_2\right) + C = V_y\left(\hat{r}_1 + C, \hat{r}_2 + C, t_1, t_2\right)$ for any scalar C (see [63] for derivation leading to this property). Next, we define the value function on expected reward differences by

$$\bar{V}_y(\Delta, t_1, t_2) = V_y(\hat{r}_1, \hat{r}_2, t_1, t_2) - \frac{\hat{r}_1 + \hat{r}_2}{2} = V_y(\Delta, -\Delta, t_1, t_2). \tag{6.21}$$

Applying this mapping to Eq. (6.19) leads to Bellman's equation

$$\bar{V}_y(\Delta, t_1, t_2) = \max \left\{ \begin{array}{c} \Delta, -\Delta, \\ \left\langle \bar{V}_y(\Delta(t+\delta t), t_1 + |2-y|\delta t, t_2 + |1-y|\delta t) \,|\Delta, t_1, t_2, y \right\rangle - c\delta t, \\ \bar{V}_{3-y}(\Delta, t_1, t_2) - c_s \end{array} \right\}, \tag{6.22}$$

which is now defined over a four-dimensional rather than a five-dimensional space while yielding the same optimal policy. This also confirms that optimal decision-making doesn't require tracking individual expected rewards, but only their difference.

6.3.2 Optimal policy for perceptual decisions

To apply the same principles to perceptual decision-making, we need to re-visit the interpretation of the latent states, z_1 and z_2. Those could, for example, be the brightness of two dots on a screen, and the decision-maker needs to identify the brighter dot. Anternatively, they might reflect the length of two lines, and the decision maker needs to identify which of the two lines is longer. Either way, the reward is a function of z_1, z_2, and the decision maker's choice. Therefore, the expected reward for choosing either option can be computed from the posterior z's, Eqs. (6.5) and (6.6). Furthermore, these posteriors are fully determined by their means, $\hat{r}_1$, $\hat{r}_2$, and the attention times, t_1 and t_2. As a consequence, we can formulate the expected reward for choosing item j by the expected reward function $R_j(\hat{r}_1, \hat{r}_2, t_1, t_2)$.

What are the consequences for this change in expected reward for the optimal policy? If we assume the attention-modulated evidence accumulation process to remain unchanged, the only change is that the expected return for choosing item j changes from $\hat{r}_j$ to $R_j(\hat{r}_1, \hat{r}_2, t_1, t_2)$. Therefore, Bellman's equations changes to

$$\begin{aligned} &V_y(\hat{r}_1, \hat{r}_2, t_1, t_2) \\ &= \max \left\{ \begin{array}{c} R_1(\hat{r}_1, \hat{r}_2, t_1, t_2), R_2(\hat{r}_1, \hat{r}_2, t_1, t_2), \\ \langle V_y(\hat{r}_1(t+\delta t), \hat{r}_2(t+\delta t), t_1 + |2-y|\delta t, t_2 + |1-y|\delta t) \,|\hat{r}_1, \hat{r}_2, t_1, t_2, y \rangle - c\delta t, \\ V_{3-y}(\hat{r}_1, \hat{r}_2, t_1, t_2) - c_s \end{array} \right\}. \end{aligned} \tag{6.23}$$

The optimal policy follows from Bellman's equation as before.

The above value function can only be turned into one over expected reward differences under certain regularities of R_1 and R_2, which we will not discuss further at this point. Furthermore, for the above example, we have assumed two sources of perceptual evidence that need to be compared. Alternative tasks (e.g., the random dot motion task) might provide a single source of evidence that needs to be categorized. In this case, the formulation changes slightly (see, for example, [9]), but the principles remain unchanged.

6.4 Simulation details

6.4.1 Computing the optimal policy

In Section 6.3, we described the Bellman equation (Eq. (6.22)) which outputs the expected return given these four parameters: currently attended item (y), reward difference (Δ), expected return for accumulating more evidence, and expected return for switching attention. Note that the symmetry of the value function (see [63] for the derivation establishing symmetry) allows us to drop $-\Delta$ from the original Eq. (6.22). Solving this Bellman equation provides us with a 4-dimensional ”policy space” which assigns the optimal action to take at any point in this space defined by the four parameters above.

The solution to the optimal policy can be found numerically by backwards induction [10]. To do so, first we assume some large $t = t_1 + t_2$, where a decision is guaranteed. In this case, $V_y(\Delta, t_1, t_2) = \max\{-\Delta, \Delta\} = |\Delta|$ for both $y = 1$ and $y = 2$. We call this the base case. From this base case, we can move one time step backwards in t_1 ($y = 1$):

$$\bar{V}_1\left(\Delta, t_1 - \delta t, t_2\right) = \max \left\{ \begin{array}{c} \Delta, \\ \left\langle \bar{V}_1\left(\Delta, t_1, t_2\right) | \Delta, t_1, t_2 \right\rangle - c\delta t, \\ \bar{V}_2\left(\Delta, t_1 - \delta t, t_2\right) - c_s \end{array} \right\}, \tag{6.24}$$

The second expression in the maximum can be evaluated, since we assume a decision is made at time t. But $\bar{V}_2\left(\Delta, t_1 - \delta t, t_2\right) - c_s$, which is the value function for switching attention, is unknown. This unknown value function is given by

$$\bar{V}_2\left(\Delta, t_1 - \delta t, t_2\right) = \max \left\{ \begin{array}{c} \Delta, \\ \left\langle \bar{V}_2\left(\Delta, t_1 - \delta t, t_2 + \delta t\right) | \Delta, t_1, t_2 \right\rangle - c\delta t, \\ \bar{V}_1\left(\Delta, t_1 - \delta t, t_2\right) - c_s \end{array} \right\}, \tag{6.25}$$

In this expression, the second term can again be found, but $\bar{V}_1\left(\Delta, t_1 - \delta t, t_2\right) - c_s$ is unknown. Looking at the two expressions above, we see that under the parameters $(\Delta, t_1 - \delta t, t_2)$, $V_1 \geq V_2 - c_s$, and $V_2 \geq V_1 - c_s$, which cannot both be true. Therefore, we first assume that V_1 is not determined by $V_2 - c_s$, removing the $V_2 - c_s$ term from the maximum. This allows us to find $\bar{V}_1\left(\Delta, t_1 - \delta t, t_2\right)$ in Eq. (6.24). Then, we compute Eq. (6.25) including the $V_1 - c_s$ term. If we find that $V_2 = V_1 - c_s$, then $V_1 \neq V_2 - c_s$, which means the $V_2 - c_s$ term could not have mattered in Eq. (6.24), and we are done. If not, we re-compute V_1 with the $V_2 - c_s$ term included, and we are done. Therefore, we were able to compute V_1 and V_2 under the parameters $(\Delta, t_1 - \delta t, t_2)$ using information about $\bar{V}_1(\Delta, t_1, t_2)$ and $\bar{V}_2(\Delta, t_1 - \delta t, t_2 + \delta t)$.

Using the same approach, we can find $V_{1,2}(\Delta, t_1, t_2 - \delta t)$ based on $\bar{V}_1(\Delta, t_1 - \delta t, t_2 + \delta t)$ and $\bar{V}_2(\Delta, t_1, t_2)$. Thus, given that we know $V_y(\Delta, t_1, t_2)$ above a certain $t = t_1 + t_2$, we can move backwards to compute V_1 and V_2 for $(\Delta, t_1 - \delta t, t_2)$, then $(\Delta, t_1 - 2\delta t, t_2)$, and so on, until $(\Delta, 0, t_2)$ for all relevant values of Δ. Subsequently, we can do the same moving backwards in t_2, solving for $V_y(\Delta, t_1, t_2 - \delta t)$, $V_y(\Delta, t_1, t_2 - 2\delta t)$, ..., $V_y(\Delta, t_1, 0)$. Following this, we can continue with the same procedure from $V_y(\Delta, t_1 - \delta t, t_2 - \delta t)$, until we have found $V_{1,2}$ for all combinations of t_1 and t_2.

In practice, the parameters of the optimal policy space were discretized to allow for tractable computation. We set the large time at which decisions are guaranteed at $t = 6s$, which we determined empirically. Time was discretized into steps of $\delta t = 0.05s$. The item values, and their difference (Δ) were also discretized into steps of 0.05.

Upon completing this exercise, we now have two 3-dimensional optimal policy spaces. The decision-maker's location in this policy space is determined by t_1, t_2, and Δ. Each point in this space is assigned an optimal action to take (choose item, accumulate more evidence, switch attention) based on which expression was largest in the maximum of the respective Bellman equation. The decision-maker moves between the two policy spaces depending on which item they are attending to ($y \in [1, 2]$).

In order to find the 3-dimensional boundaries that signify a change in optimal action to take, we took slices of the optimal policy space in planes of constant Δ's. We found the boundary between different optimal policies within each of these slices. We in turn approximated the 3-dimensional contour of the optimal policy boundaries by collating them along the different Δ's.

6.4.2 Finding task parameters that best match human behavior

In computing the optimal policy, there were several free parameters that determined the shape of the policy boundaries, thereby affecting the behavior of the optimal model. These parameters included σ^2, σ_z^2, c, c_s, and γ. Our goal was to find a set of parameters that qualitatively mimic human behavior as best as possible. To do so, we performed a random search over the following parameter values: $c_s \in [0.001, 0.05]$ (steps size 0.001), $c \in [0.01, 0.4]$ (steps size 0.01), $\sigma^2 \in [1, 100]$ (step size 1), $\sigma_z^2 \in [1, 100]$ (step size 1), $\gamma \in [0.001, 0.01]$ (step size 0.001) [62].

To find the best qualitative fit, we simulated behavior from a randomly selected set of parameter values (see next section for simulation procedure). From this simulated behavior, we evaluated the match between human and model behavior by applying the same procedure to each of Figure 2.4B,C,E. For each bin for each plot, we subtracted the mean values between the model and human data, then divided this difference by the standard deviation of the human data corresponding to that bin, essentially computing the effect size of the difference in means. We computed the sum of these effect sizes for every bin, which served as a metric for how qualitatively similar the curves were between the model and human data. We performed the same procedure for all three figures, and ranked the sum of the effect sizes for all simulations. We performed simulations for over 2,000,000 random sets of parameter values. The set of parameters for which our model best replicated human behavior according to the above criteria was $c_s = 0.0065$, $c = 0.23$, $\sigma^2 = 27$, $\sigma_z^2 = 18$, $\gamma = 0.004$.

6.4.3 Simulating decisions with the optimal policy

The optimal policy allowed us to simulate decision making in a task analogous to the one humans performed in [2]. For a given set of parameters, we first computed the optimal policy. In a simulated trial, two items with values z_1 and z_2 are presented. At trial onset, the model attends to an item randomly ($y \in [1, 2]$), and starts accumulating noisy evidence centered around the true values. At every time step ($\delta t = 0.05$), the model evaluates Δ using the mean of the posteriors between the two items (see Eqs. (6.5) and (6.6)). Then, the model performs the optimal action associated with its location in the optimal policy space. If the model makes a decision, then the trial is over. If the model instead accumulates more evidence, then the above procedure is repeated for the next time step. If the model switches attention, it does not obtain further information about either item, but switches attention to the other item. Switching attention allows for more reliable evidence from the now-attended item, and also switches the optimal policy space to the appropriate one (see Figure 2.2).

To allow for a relatively fair comparison between the model and human data, we simulated the same number of subjects ($N = 39$) for the model, but with a larger number of trials. For each simulated subject, trials were created such that all pairwise combinations of values between 0 and 7 were included, and this was iterated 20 times. This yielded a total of 1280 trials per subject.

6.4.4 Attention diffusion model

In order compare the decision performance of the optimal model to that of the original attentional drift diffusion model (aDDM) proposed by Krajbich et al. (2010) [2], we needed to ensure that neither model had an advantage by receiving more information. We did so by making sure that the signal-to-noise ratios of evidence accumulation of both models were identical. In aDDM, the evidence accumulation evolved according to the following process, in steps of 0.05s (assuming y = 1):

$$v_t = v_{t-1} + d(z_1 - \gamma_k z_2) + \eta_t, \tag{6.26}$$

where v_t is the relative decision value that represents the subjective value difference between the two items at time t, d is a constant that controls the speed of integration (in ms^{-1}), γ_k controls the biasing effect of attention, and $\eta_t \sim \mathcal{N}(0, \sigma^2)$ is a normally distributed random variable zero mean and variance σ^2. Written differently, the difference in the attention-weighted momentary evidence between item 1 and item 2 can be expressed as

$$\begin{aligned}\delta\Delta = d\left(z_1 - \gamma_k z_2\right) + \eta_t &\sim \mathcal{N}\left(d(z_1 - \gamma_k z_2), \sigma^2\right) \\ &\sim \mathcal{N}\left(k(z_1 - \gamma_k z_2)\delta t, \sigma_k^2 \delta t\right),\end{aligned} \tag{6.27}$$

where d and σ^2 were replaced by $k\delta t$, and $\sigma_k^2\delta t$, respectively. Here, the variance term $\sigma_k^2\delta t$ can be split into two parts, such that the $\delta\Delta$ term can be expressed as

$$\delta\Delta \sim \mathcal{N}\left(z_1 k\delta t, \frac{1}{2}\sigma_k^2\delta t\right) - \mathcal{N}\left(\gamma_k z_2 k\delta t, \frac{1}{2}\sigma_k^2\delta t\right). \tag{6.28}$$

The signal-to-noise ratios (i.e., the ratio of mean over standard deviation) of the two terms in the above equation are $\frac{z_1 k\delta t}{\sqrt{\frac{\delta t}{2}}\sigma_k}$ and $\frac{z_2 k\delta t}{\frac{1}{\gamma_k}\sigma_k\sqrt{\frac{\delta t}{2}}}$, respectively.

Continuing to assume $y = 1$, in the Bayes-optimal model, evidence accumulation evolves according to

$$\begin{aligned}\delta x_1 &\sim \mathcal{N}\left(z_1\delta t, \sigma_b^2\delta t\right), \\ \delta x_2 &\sim \mathcal{N}\left(z_2\delta t, \gamma_b^{-1}\sigma_b^2\delta t\right).\end{aligned} \tag{6.29}$$

Therefore, the difference in the attention-weighted momentary evidence between item 1 and item 2 can be expressed as:

$$\begin{aligned}\delta\Delta &\sim \mathcal{N}\left(z_1\delta t, \sigma_b^2\delta t\right) - \gamma_b\mathcal{N}\left(z_2\delta t, \gamma_b^{-1}\sigma_b^2\delta t\right) \\ &\sim \mathcal{N}\left(z_1\delta t, \sigma_b^2\delta t\right) - \mathcal{N}\left(\gamma_b z_2\delta t, \gamma_b\sigma_b^2\delta t\right).\end{aligned} \tag{6.30}$$

The signal-to-noise ratios of the two terms in the above equation are $\frac{z_1\delta t}{\sqrt{\delta t}\sigma_b}$ and $\frac{z_2\delta t}{\frac{1}{\sqrt{\gamma_b}}\sigma_b\sqrt{\delta t}}$, respectively.

In order to match the signal-to-noise ratios of the two models, we set equal their corresponding expressions, to find the following relationship between the parameters of the two models:

$$\begin{aligned}k &= 1, \\ \sigma_k^2 &= 2\sigma_b^2, \\ \gamma_k &= \sqrt{\gamma_b}.\end{aligned} \tag{6.31}$$

Therefore, we simulated the aDDM with model parameters $\gamma_k = \sqrt{\gamma_b}$ and $\sigma_k^2 = 2\sigma_b^2$.

In the original aDDM model, the model parameters were estimated by fitting the model behavior to human behavior after setting a decision threshold at ± 1. Since we adjusted some of the aDDM parameters, we instead iterated through different decision thresholds (1 through 10, in increments of 1) and found the value that maximizes model performance. To keep it consistent with behavioral data, we generated 39 simulated participants that each completed 200 trials where the two item values were drawn from the prior distribution of the optimal policy model, $z_j \sim \mathcal{N}(\bar{z}, \sigma_z^2)$ using both the optimal model and the aDDM model.

6.4.5 Adjusting the attention bottleneck

We investigated whether changing the relative amount of attentional resource dedicated to the attended versus unattended item would influence decision-making performance. To do so, we varied the amount of momentary evidence provided about the attended and unattended items while keeping the overall evidence constant. We found the overall evidence from the base model by computing the Fisher information (I_{base}) it provides about the respective true item values. This Fisher information is computed as the sum of the reciprocal of the variance from the attended and unattended items, resulting in

$$I_{base} = \frac{1}{\sigma^2} + \frac{1}{\gamma^{-1}\sigma^2} = \frac{1+\gamma}{\sigma^2}. \tag{6.32}$$

Our goal is to use κ ($0 \leq \kappa \leq 1$) to control the relative attentional resource allocated to the attended versus unattended item, analogous to the γ term used in the base model. To do so, we set the variance of the two items as $\sigma_{tot}^2/(1-\kappa)$ and σ_{tot}^2/κ for the attended and unattended items, respectively, where $\sigma_{tot}^2 = \frac{1}{I_{base}}$ represents the total variance associated with evidence accumulation of both items. This satisfies our requirement of flexibly changing attention allocation while maintaining the Fisher information of the base model,

$$\frac{1-\kappa}{\sigma_{tot}^2} + \frac{\kappa}{\sigma_{tot}^2} = \frac{1}{\sigma_{tot}^2} = I_{base}. \tag{6.33}$$

To implement this adjusted model, for each value of κ, we found the associated σ_κ^2 and γ_κ to replace the σ^2 and γ terms in the base model. To do so, we set the variance of the attended item above equal to that from the base model,

$$\frac{1-\kappa}{\sigma_{tot}^2} = \frac{1}{\sigma_\kappa^2}. \tag{6.34}$$

Since $\sigma_{tot}^2 = \frac{1}{I_{base}} = \frac{\sigma^2}{1+\gamma}$, we can rearrange the above and solve for σ_k^2 and γ_κ to get,

$$\begin{aligned} \sigma_\kappa^2 &= \frac{\sigma_{tot}^2}{1-\kappa}, \\ \gamma_\kappa &= \frac{\kappa}{1-\kappa}. \end{aligned} \tag{6.35}$$

Using the above σ_κ^2 and γ_κ, we computed the optimal policy and simulated behavior using the same approach as for the base model.

Bibliography

[1] S. C. H. Yang, M. Lengyel, and D. M. Wolpert, "Active sensing in the categorization of visual patterns," *eLife*, 2016.

[2] I. Krajbich, C. Armel, and A. Rangel, "Visual fixations and the computation and comparison of value in simple choice," *Nature Neuroscience*, 2010.

[3] M. R. Cohen and J. H. Maunsell, "Attention improves performance primarily by reducing interneuronal correlations," *Nature Neuroscience*, 2009.

[4] A. Rangel and T. Hare, "Neural computations associated with goal-directed choice," *Current Opinion in Neurobiology*, 2010.

[5] M. N. N. Shadlen and D. Shohamy, "Decision Making and Sequential Sampling from Memory," *Neuron*, 2016.

[6] R. Ratcliff and G. McKoon, "The diffusion decision model: Theory and data for two-choice decision tasks," 2008.

[7] M. Milosavljevic, J. Malmaud, A. Huth, C. Koch, and A. Rangel, "The Drift Diffusion Model can account for the accuracy and reaction time of value-based choices under high and low time pressure," *Judgment and Decision Making*, 2010.

[8] R. Bogacz, E. Brown, J. Moehlis, P. Holmes, and J. D. Cohen, "The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks," *Psychological Review*, 2006.

[9] J. Drugowitsch, R. N. Moreno-Bote, A. K. Churchland, M. N. Shadlen, and A. Pouget, "The cost of accumulating evidence in perceptual decision making," *Journal of Neuroscience*, 2012.

[10] S. Tajima, J. Drugowitsch, and A. Pouget, "Optimal policy for value-based decision-making," *Nature Communications*, 2016.

[11] A. A. Kustov and D. L. Robinson, “Shared neural control of attentional shifts and eye movements,” *Nature*, 1996.

[12] C. W. Mohler and R. H. Wurtz, “Organization of monkey superior colliculus. Intermediate layer cells discharging before eye movements,” *Journal of Neurophysiology*, 1976.

[13] S. Shimojo, C. Simion, E. Shimojo, and C. Scheier, “Gaze bias both reflects and influences preference,” *Nature Neuroscience*, 2003.

[14] I. Krajbich and A. Rangel, “Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions,” *Proceedings of the National Academy of Sciences of the United States of America*, 2011.

[15] J. F. Cavanagh, T. V. Wiecki, A. Kochar, and M. J. Frank, “Eye tracking and pupillometry are indicators of dissociable latent decision processes,” *Journal of Experimental Psychology: General*, 2014.

[16] A. J. Yu, P. Dayan, and J. D. Cohen, “Dynamics of Attentional Selection Under Conflict: Toward a Rational Bayesian Account,” *Journal of Experimental Psychology: Human Perception and Performance*, 2009.

[17] R. B. Towal, M. Mormann, and C. Koch, “Simultaneous modeling of visual saliency and value computation improves predictions of economic choice,” *Proceedings of the National Academy of Sciences of the United States of America*, 2013.

[18] T. C. Cassey, D. R. Evens, R. Bogacz, J. A. Marshall, and C. J. Ludwig, “Adaptive sampling of information in perceptual decision-making,” *PLoS ONE*, 2013.

[19] S. Gluth, N. Kern, M. Kortmann, and C. L. Vitali, “Value-based attention but not divisive normalization influences decisions with multiple alternatives,” *Nature Human Behaviour*, 2020.

[20] D. Hoppe and C. A. Rothkopf, “Learning rational temporal eye movement strategies,” *Proceedings of the National Academy of Sciences of the United States of America*, 2016.

[21] M. Hayhoe and D. Ballard, “Eye movements in natural behavior,” 2005.

[22] L. Chukoskie, J. Snider, M. C. Mozer, R. J. Krauzlis, and T. J. Sejnowski, “Learning where to look for a hidden target,” *Proceedings of the National Academy of Sciences of the United States of America*, 2013.

[23] M. Corbetta and G. L. Shulman, “Control of goal-directed and stimulus-driven attention in the brain,” *Nature Reviews Neuroscience*, 2002.

[24] B. B. Averbeck, P. E. Latham, and A. Pouget, “Neural correlations, population coding and computation,” 2006.

[25] J. F. Mitchell, K. A. Sundberg, and J. H. Reynolds, “Spatial Attention Decorrelates Intrinsic Activity Fluctuations in Macaque Area V4,” *Neuron*, 2009.

[26] J. H. Wittig, A. I. Jang, J. B. Cocjin, S. K. Inati, and K. A. Zaghloul, “Attention improves memory by suppressing spiking-neuron activity in the human anterior temporal lobe,” *Nature Neuroscience*, 2018.

[27] M. I. Posner, “Orienting of attention.” *The Quarterly journal of experimental psychology*, 1980.

[28] W. S. Geisler and L. K. Cormack, “Models of overt attention,” in *The Oxford Handbook of Eye Movements*, 2012.

[29] M. R. Cohen and J. H. Maunsell, “A neuronal population measure of attention predicts behavioral performance on individual trials,” *Journal of Neuroscience*, 2010.

[30] L. Wang and R. J. Krauzlis, “Visual Selective Attention in Mice,” *Current Biology*, 2018.

[31] A. M. Ni, D. A. Ruff, J. J. Alberts, J. Symmonds, and M. R. Cohen, “Learning and attention reveal a general relationship between population activity and behavior,” *Science*, 2018.

[32] D. A. Ruff, A. M. Ni, and M. R. Cohen, “Cognition as a Window into Neuronal Population Space,” *Annual Review of Neuroscience*, 2018.

[33] J. Drugowitsch, R. Moreno-Bote, and A. Pouget, “Optimal decision-making with time-varying evidence reliability,” in *Advances in Neural Information Processing Systems*, 2014.

[34] R. Bellman, “On the Theory of Dynamic Programming,” *Proceedings of the National Academy of Sciences*, 1952.

[35] D. P. Bertsekas, *Dynamic programming and optimal control*, 4th ed. Athena scientific, 1995, vol. 1.

[36] R. H. Wurtz, “Neuronal mechanisms of visual stability,” *Vision Research*, 2008.

[37] M. Song, X. Wang, H. Zhang, and J. Li, “Proactive information sampling in value-based decision-making: Deciding when and where to saccade,” *Frontiers in Human Neuroscience*, 2019.

[38] S. Z. Li and W. J. Ma, “Valuation as inference: A New Model for the Effects of Fixation on Choice,” in *Conference on Cognitive Computational Neuroscience*, 2019.

[39] D. Bahdanau, K. H. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015.

[40] J. Gehring, M. Auli, D. Grangier, D. Yarats, and Y. N. Dauphin, "Convolutional sequence to sequence learning," in *34th International Conference on Machine Learning, ICML 2017*, 2017.

[41] V. Mnih, N. Heess, A. Graves, and K. Kavukcuoglu, "Recurrent models of visual attention," in *Advances in Neural Information Processing Systems*, 2014.

[42] J. L. Ba, V. Mnih, and K. Kavukcuoglu, "Multiple object recognition with visual attention," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015.

[43] I. Sorokin, A. Seleznev, M. Pavlov, A. Fedorov, and A. Ignateva, "Deep attention recurrent Q-network," *arXiv preprint arXiv:1512.01693*, 2015.

[44] G. Tavares, P. Perona, and A. Rangel, "The attentional Drift Diffusion Model of simple perceptual decision-making," *Frontiers in Neuroscience*, 2017.

[45] T. T. Ke, Z. J. M. Shen, and J. M. Villas-Boas, "Search for information on multiple products," *Management Science*, 2016.

[46] B. Hébert and M. Woodford, "Rational Inattention when decisions take time," *Nber Working Paper Series*, 2019.

[47] J. H. Reynolds and L. Chelazzi, "ATTENTIONAL MODULATION OF VISUAL PROCESSING," *Annual Review of Neuroscience*, 2004.

[48] J. F. Mitchell, K. A. Sundberg, and J. H. Reynolds, "Differential Attention-Dependent Response Modulation across Cell Classes in Macaque Visual Area V4," *Neuron*, 2007.

[49] L. Acerbi, S. Vijayakumar, and D. M. Wolpert, "On the Origins of Suboptimality in Human Probabilistic Inference," *PLoS Computational Biology*, 2014.

[50] J. Drugowitsch, V. Wyart, A. D. Devauchelle, and E. Koechlin, "Computational Precision of Mental Inference as Critical Source of Human Choice Suboptimality," *Neuron*, 2016.

[51] F. Callaway, A. Rangel, and T. L. Griffiths, "Fixation patterns in simple choice are consistent with optimal use of cognitive resources," *PsyArXiv*, 2020.

[52] S. Tajima, J. Drugowitsch, N. Patel, and A. Pouget, "Optimal policy for multi-alternative decisions," *Nature Neuroscience*, 2019.

[53] A. Shenhav, C. K. Dean Wolf, and U. R. Karmarkar, "The evil of banality: When choosing between the mundane feels like choosing between the worst," *Journal of Experimental Psychology: General*, 2018.

[54] Z. Li and W.-J. Ma, "An uncertainty-based model of the effects of fixation on choice," *PsyArXiv*, 2020.

[55] M. W. Khaw, P. W. Glimcher, and K. Louie, "Normalized value coding explains dynamic adaptation in the human valuation process," *Proceedings of the National Academy of Sciences of the United States of America*, 2017.

[56] K. C. Armel, A. Beaumel, and A. Rangel, "Biasing simple choices by manipulating relative visual attention," *Judgment and Decision Making*, 2008.

[57] S. M. Smith and I. Krajbich, "Attention and choice across domains," *Journal of Experimental Psychology: General*, 2018.

[58] ——, "Gaze Amplifies Value in Decision Making," *Psychological Science*, 2019.

[59] D. Fudenberg, P. Strack, and T. Strzalecki, "Speed, accuracy, and the optimal timing of choices," 2018.

[60] L. Itti and C. Koch, "Computational modelling of visual attention," *Nature Reviews Neuroscience*, 2001.

[61] A. E. Brockwell and J. B. Kadane, "A Gridding Method for Bayesian Sequential Decision Problems," *Journal of Computational and Graphical Statistics*, 2003.

[62] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *Journal of Machine Learning Research*, 2012.

[63] A. I. Jang, R. Sharma, and J. Drugowitsch, "Optimal policy for attention-modulated decisions explains human fixation behavior," *bioRxiv*, 2020.