Confidence as an Arbiter of Attentional Allocation During Learning in Multidimensional Environments

Julie Elizabeth Newman

A senior thesis submitted in partial fulfillment
of the requirements for the degree of
Bachelor of Arts
in the Department of Neuroscience at
Princeton University

Adviser: Professor Yael Niv

7 May 2018

Abstract [1]

Reinforcement learning algorithms are notoriously inefficient in high-dimensional environments, and yet people manage to solve these complex problems with ease. One way in which our brains are thought to make such high-dimensional problems tractable is by using selective attention to reduce the dimensionality to only those features that are relevant for the task. Prior work has demonstrated that in reward-based learning, there is a bi-directional relationship between learning and attention, but how the brain decides where to employ attention over the course of learning is debated. Drawing on ideas from theoretical and experimental work, we propose that internal confidence computations may arbitrate between different attention strategies. To test this hypothesis, we used a high-dimensional reinforcement learning task in which efficient learning and concomitant maximization of reward requires narrowing of subjects' attention to the relevant dimension. Our results demonstrate a clear link between confidence and the breadth of attention during learning. We also incorporated our hypotheses into two novel computational models that predict trial-by-trial attention during the task. While neither of our models performed better than the value-based model to which we compared them, our results do not disprove the idea that confidence modulates the distribution of attention during learning. Rather, they suggest that models which allocate attention purely based on value miss an important component of how subjects actually distribute their attention, and that more thought needs to be given to the role of confidence, as well as perseverance and hypothesis-testing.

[1] This section contains text that is based closely on, or identical to, text found in my junior paper (2017).

Acknowledgements

I would like to thank...

Professor Niv, for your guidance about thesis, college, and life after Princeton.

Angela Radulescu, for being there every step of the way through this thesis. Thank you for all of your comments and suggestions, for your insight, and for helping me make sense of a confusing whirlwind of ideas.

My friends, for making these last four years at Princeton such a wonderful experience. Thanks for all the laughter and smiles, the late meal runs, the silent dance parties, the late night study sessions, the wine and movie nights, the countless tea breaks, and the amazing conversations. Thanks for dancing in the rain, playing in the snow, and relaxing in the sunshine, even when there was work we should have been doing.

My family, for your constant support, encouragement, and unconditional love. Kathleen and Anne, you are the two best sisters anyone could ask for, and I'm so grateful that I was able to share part of my Princeton experience with each of you. Mom and Dad, you have always been there for me and always believed in me. I couldn't have made it here without you. Thank you for everything.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Learning to Learn[1]

In challenging situations, maximizing learning is often as much about determining which pieces of information are relevant as it is about cramming more information into the brain. But how do we figure out what is important to learn about and what is not? We live in a multi-dimensional, highly complex world in which we are constantly inundated with sights, sounds, and other sensations. While some of these stimuli are important, others must be filtered out so as not to overwhelm us. This is especially true in learning and decision-making situations, as we attempt to determine which stimuli should influence our future choices. Relevant (reward) signals often co-occur with irrelevant stimuli, making learning more challenging. While it might be optimal to learn about every feature of every stimulus, and make decisions based on that wealth of information, we do not have the neural capacity to do so (Feng et al., 2014; reviewed in Desimone and Duncan, 1995). Thus, it is important to understand what strategies our brains use to learn efficiently in a complex world.

[1]This section and the following sections in this chapter contain text that is based closely on, or identical to, text found in my junior paper (2017).

1.2 Reinforcement Learning

One of the most popular theories of neural learning is reinforcement learning (RL), a trial-and-error based model adapted from computer science. Reinforcement learning models of decision-making aim to maximize long-term reward and minimize punishment, using previous experience as a guide to gauge future outcomes (Schultz et al., 1997; Sutton and Barto, 1998, Niv and Schoenbaum, 2008). In RL paradigms, stimulus values (corresponding to future expected reward) are learned through making choices and receiving feedback. If there is a difference between the expected value of a choice and the actual reward/punishment received (a quantity known as reward prediction error), values are updated to reflect that information.

Reinforcement learning is particularly popular as a model because many studies have shown a neural substrate for the type of error-driven learning that characterizes RL (Montague et al., 1996, Schultz et al., 1993; Schultz et al., 1997). Reward pathways in the midbrain-basal ganglia circuit are thought to encode prediction errors through dopamine release: positive prediction results in greater bursts of activity from dopamine neurons while a negative prediction results in decreased activity (Schultz et al., 1993).

Despite its attraction as a neurobiologically plausible model, RL fails to account for learning in complex, multidimensional environments (Sutton and Barto, 1998; Bellman, 1957). As the dimensionality of the problem increases, RL models break down, becoming less and less efficient – a trend referred to as the curse of dimensionality (Sutton and Barto, 1998). Recent work in representational learning has indicated that the brain may solve this curse of dimensionality by selecting a smaller subset of dimensions that are relevant for learning, thereby reducing the complexity of neural computations and decisions (Gershman and Niv, 2010; Jones and Canas, 2010; Niv et al., 2015; Leong et al., 2017)

1.3 Dimensionality Reduction and Use of Selective Attention

Determining which (and how many) dimensions of the task are relevant is key for good performance. If too few dimensions are chosen, the learner misses out on information that could be relevant to reward. If too many dimensions are chosen, learning is inefficient and the learner wastes resources trying to pay attention to and remember irrelevant stimuli. In the computer-science field of RL, this dimension-reduction problem is formalized as learning state-spaces, which correspond to the internal representation of the current task, and selective perception is used to reduce the state-space (McCallum, 1996).

Exactly how the brain learns to do dimensionality-reduction is still not fully understood, but a key factor is thought to be selective attention (Dayan et al., 2000; Nosofsky, 1994; Rehder and Hoffman, 2005; Roelfsema and van Ooyen, 2005). Selective allocation of attention reflects the narrowing of dimensions that the brain is learning about, and might allow the brain to successfully employ simpler computations like RL, which is intractable in large state-spaces (reviewed in Dayan and Niv, 2008; O'Doherty, 2012). Various groups have proposed a mechanism whereby people learn over time what features of a stimulus to attend to, and that attention in turn regulates future learning and biases learning towards those relevant dimensions (Jones and Canas, 2010; Niv et al., 2015). But in a situation where there is uncertainty about the value of different features, how should attention be allocated to maximize learning and reward?

Recent work suggests that when humans learn in an uncertain environment, the most optimal learning strategies are too computationally demanding, and human performance is better reflected by suboptimal strategies based on selective attention (Wilson and Niv, 2011). Work in the animal learning literature has led to two

different theories about how attention should impact learning. In one view, attention should be directed to stimuli about which least is known (Pearce and Hall, 1980). In the other view, attention should be directed to those features most predictive of reward (Mackintosh, 1975). Both strategies have found experimental support, and many recent studies have focused on finding a way to reconcile the two strategies into one comprehensive theory of selective attention during learning (Rehder and Hoffman, 2005; Haselgrove et al., 2010; Dopson et al., 2010, Leong et al., 2017).

1.4 Role of Confidence

Some insight into this problem has come from the field of machine learning (ML), where a similar tradeoff between learning strategies exists. In ML, various algorithms are used to determine how to best solve the tradeoff between exploration (collecting more information) and exploitation (making the best decision given the current information) (Tokic, 2010). In humans, attention can be thought of as a mechanism that is used to solve the explore/exploit problem for high-dimensional state-spaces. One way in which this tradeoff has been dealt with in ML is by using confidence bounds to establish criteria on when to explore and when to exploit (Auer, 2003). Computations of statistical confidence bounds reflect the reliability of and the uncertainty about the current knowledge of the environment (Auer, 2003).

When allocating attention during learning, it is unknown how the brain chooses which strategy will be most beneficial for maximizing long-term reward. However, this type of statistical computation, incorporating both data values and the reliability of those values, could be an ideal way for the brain to do so. It has been hypothesized that this statistical information is precisely what the brain uses when it calculates confidence (Kepecs et al., 2008; De Martino et al., 2013). While confidence is often thought of as a subjective, metacognitive feeling, Kepecs et al. have shown that

self-reported confidence is highly consistent with models of statistical confidence, and thus might reflect the brain's underlying computations (2008).

1.5 Hypothesis

We propose that the neural computations which manifest as subjectively felt confidence might be a way in which the brain chooses which strategy to utilize when distributing attention. How attention changes as people become more confident in their knowledge of a task is so far an unexplored area. Will they continue to seek out as much information as possible, even once they have more information about which features are most rewarding? Or, as they become more confident, will they begin to more narrowly confine their attention to the features that they believe are most relevant? Our research seeks to explore these questions, and to understand how confidence affects attention during learning. The answers will contribute to understanding how our brains solve high-dimensional learning tasks that cannot be accounted for by simple reinforcement learning models.

Chapter 2

Task Design

2.1 Overview

In order to investigate the relationship between learning, attention, and confidence, we utilized a high-dimensional, trial-and-error learning task. Subjects played a game where efficient learning, and thus maximization of reward, required selective attention to certain features and dimensions. Throughout the game, a self-report probe was used to assess subjects' confidence, and, as a measure of attention, an eye-tracker was used to measure where subjects were looking on the display screen.

2.2 Dimensions Task [1]

The "Dimensions Task" is a multi-armed bandit task previously developed by the Niv lab to study reinforcement learning paradigms (Niv et al., 2015; Leong et al., 2017). In the game, nine different images (3 faces, 3 landmarks, and 3 common tools) are arranged on the screen into a grid of compound stimuli (Figure 2.1). Each stimulus is

[1] This section contains text that is based closely on, or identical to, text found in my junior paper (2017).

a column composed of one image from each image category. The same nine features appear in every trial, arranged in different orders.

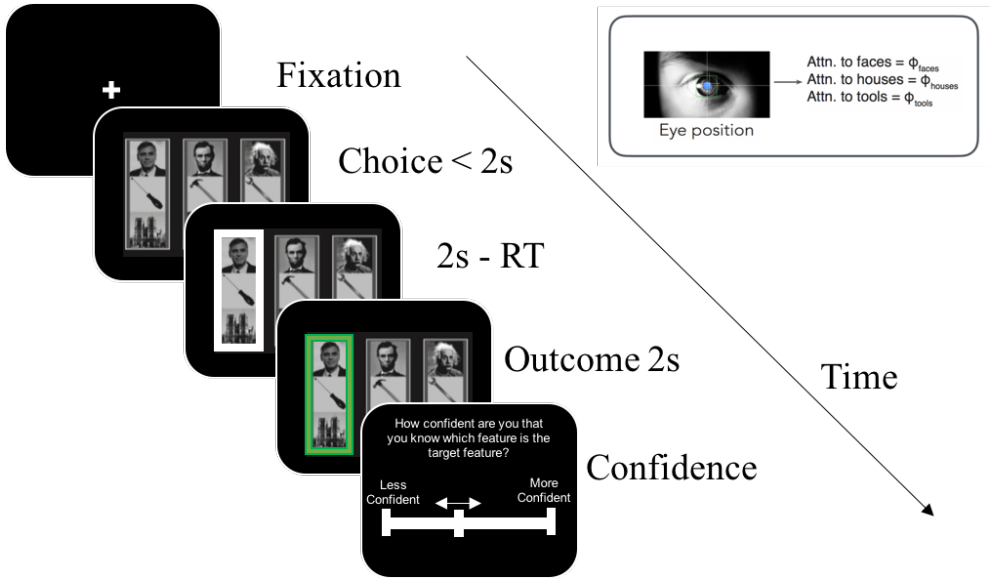

Figure 2.1: **Schematic of the "Dimensions Task".** On each trial, the subject was presented with three compound stimuli, each composed of a face, a landmark, and a tool. The participant chose one of the stimuli, received feedback on their choice (green rectangle indicating reward, red rectangle indicating no reward), and continued to the next trial. On some trials the subject was presented with a confidence probe that asked them to rate on a sliding scale how confident they were in their knowledge of the target feature. Each trial was initiated after a brief period of fixation on a white cross in the middle of the screen, included to stabilize eye tracking.

On each trial, subjects choose one of the compound stimuli and receive feedback on their choice in the form of a point reward (indicated by either a green or red outline around the chosen stimulus). In any given round of the game, only one of the three "dimensions" (faces, landmarks, or tools) determines reward. Within the relevant dimension, one "target" feature is more highly associated with reward ($p = .75$), while the other features in that dimension are associated with a lower probability of reward ($p = .25$). When a stimulus (one of the three columns) is chosen, reward is based solely on the feature of the relevant dimension. This probabilistic reward scheme means that even if a subject selects the stimulus containing the target feature, there is a chance the choice will not be rewarded. Similarly, even if the subject chooses

a stimulus that does not contain the target feature, there is still a possibility that the feedback will be positive.

Rewards are maximized when subjects learn the target feature and then choose the stimulus which contains that feature. Subjects play the game multiple times (6 rounds of three games, each game consisting of 20 trials), and the relevant dimension and target feature change at the end of each game.

2.3 Confidence Measure [2]

We wished to look at confidence as an internal computation that could direct shifts of attention allocation. Many different strategies have been used to study confidence, with the method choice depending highly on the experimental model being used. A study by Kepecs et al. found that subjective reports of confidence, measured on a scale from 1 to 5, were both highly correlated with statistical predictions of confidence, and were a strong predictor of choice accuracy (2008). As a statistical computation of confidence is precisely what we propose may underlie the shifts in attention, and as subjectively reported confidence was found to be a good analogue of the underlying computations the brain may perform, we chose to measure confidence using a similar self-reporting model.

In our task, subjects are asked to rate how confident they are that they know which feature is most predictive of reward (which feature is the target feature). The confidence probe appears only once every three trials, at the end of the trial. Confidence is measured on a sliding scale, with the low (left) end labeled less confident and the high (right) end labeled more confident (See Figure 2.1). The sliding confidence bar begins at the low end, as subjects – lacking any information about the task features – should begin with little confidence in their knowledge of the

[2]This section contains text that is based closely on, or identical to, text found in my junior paper (2017).

target feature. Subjects are able to quickly adjust their confidence rating (or leave it at the same level) using dedicated key strokes that move the reported confidence level either higher or lower on the bar compared to their previous trial.

We adjusted the design to a sliding scale bar rather than a 1-to-5 report as in Kepecs et al., because using dedicated motor responses that simply move the indicator position on the bar left or right is less likely to distract subjects from their main task and allows us to measure a greater distribution of confidence (continuous rating rather than a small number of discrete positions). Positions along the bar correspond to points along the continuous interval 0 to 1.

In the Dimensions Task, we ask subjects to rate their confidence immediately following the trial. We chose to place the measure of confidence here and focus on confidence after the outcome, rather than during reward anticipation as others have done, for multiple reasons. First, by placing the measure of confidence at the end of the trial, confidence judgments are less likely to be influenced by which stimulus the subject just chose. There is also a logistical impediment to placing the measure of confidence in between choice and feedback – the intrusion might affect how attention is distributed during feedback, both because the stimulus display will be interrupted (subjects won't have continuous viewing of the stimuli) and because subjects are made more aware of their own cognition about confidence. Importantly, we are asking subjects to report confidence in their knowledge of which feature is the target feature, rather than their confidence that the choice they have made is correct. While measuring confidence in this way is less likely to give us a measure that relates confidence to choice accuracy, it is more likely to give us a good measure of the subjects' overall confidence that they know which feature is the target feature.

However, there is still some danger that asking subjects to report their confidence between trials will either 1) distract from the main experiment, disrupting their attention and interfering with their memory for past trials, or 2) affect their

performance by causing them to think more about their own judgments instead of concentrating on learning. To control for this, we chose to probe for confidence only once every three trials. We also ran a pilot study that incorporated confidence probes but differed in no other way from a previous version of the Dimensions Task. No significant differences were observed in performance of subjects in the games with and without the confidence reports.

2.4 Tracking Eye Gaze

Through visual attention, humans are able to selectively send only the most relevant information in a scene to higher-level cortical areas for further processing (Yi & Chun, 2005; Reviewed in Baluch & Itti, 2011; Desimone & Duncan, 1995). Selective attention has been highly implicated in both learning and memory processes (Jones & Cañas, 2010; Uncapher & Rugg, 2009; Reviewed in Chun & Turk-Browne, 2007). Eye movements, which provide a highly direct and continuous measure of visual activity, are a good proxy for how much subjects are attending to various elements in their environment (Duc et al., 2008; Borji et al., 2013). In this task, eye tracking allows us to measure– with good spatial and temporal precision– which features/dimensions subjects are attending to as they play the Dimensions Task.

We used an EyeLink 1000 Plus system (SR Research) to track the eye gaze of subjects as they performed the Dimensions Task. Subjects were seated approximately 60cm away from the screen. A chin and forehead rest was used to keep subjects' heads at a stable position throughout the duration of the task, aiding in the accuracy of the eye-tracker. The system had a sampling rate of 500 Hz.

Eye-tracking data was pre-processed using an in-house MATLAB code that extracted the proportion of time fixation was directed towards each feature (ϕ_F) or dimension (ϕ_D) of the task. Features were defined by a rectangular area of

interest (AOI) corresponding to the space that feature occupied on the visual display. Similarly, dimensional attention was defined by a rectangular AOI that encompassed all features within that dimension.

Chapter 3

Behavioral Results

3.1 Participants

24 subjects participated in the principal study. Subjects were recruited from among the Princeton University community. Three subjects were not included in the analysis – either due to technical malfunctions during the experiment (2 subjects) or failure to follow task instructions (1 subject) – leaving a total of 21 participants. An additional 6 subjects participated in a pilot version of the study. All subjects were compensated for their time. Study procedures were approved by the Princeton University Institutional Review Board.

3.2 Behavioral Results

Evaluating Learning

Subjects each played 18 total games of the Dimensions Task, divided into six runs of three games each. Games were twenty trials long. The beginning and end of each game were clearly indicated to the subject through on-screen messages. Subjects were informed that the relevant dimension and feature changed in between each game.

Through trial and error, subjects were able to learn which feature was the "target" feature (Figure 3.1). By the final trial of each game, subjects chose the stimulus containing the target feature on average 72% (SE 2.31) of the time. Each game was followed by a feedback screen where subjects could indicate which feature they believed was the target feature. If a subject correctly chose the target feature, the game was marked as "learned". If they chose the wrong feature or indicated that they did not know, the game was marked as "unlearned". On average, subjects learned 60.32% (12 out of 20) games. Feedback reports from participants were generally consistent with end of game performance (Figure 3.1D).

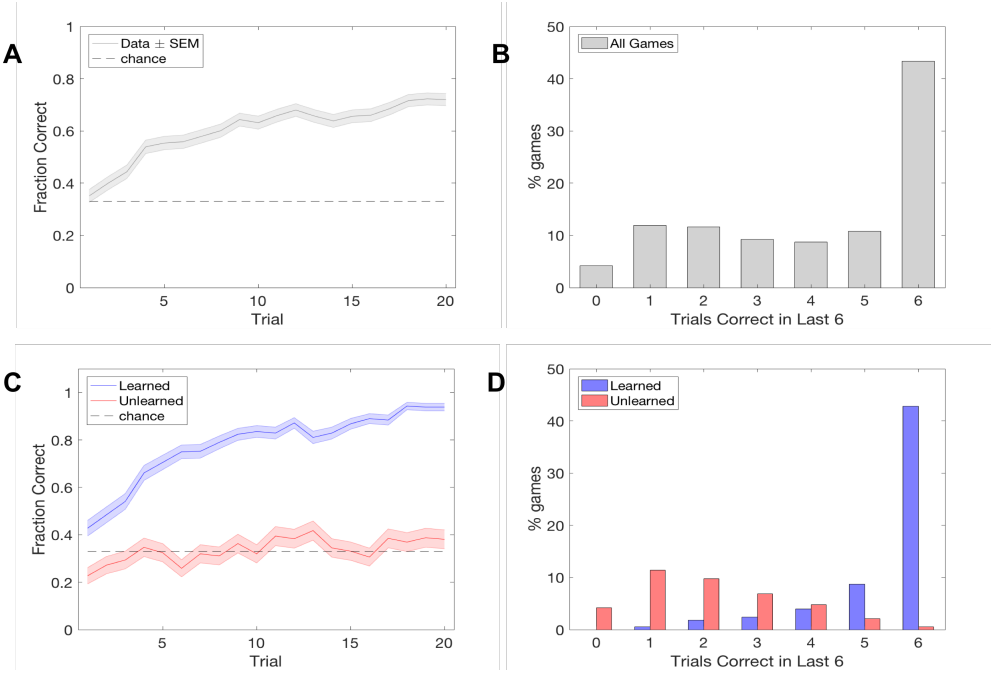

Figure 3.1: **Behavioral Results. A.** Average learning over the course of the game. Plotted is the average percent of correct choices made on that trial, across subjects and games. A correct trial was one in which the participant chose the stimulus containing the target feature. Dashed line shows chance performance. **B.** Percentage of games in which subjects chose the correct stimulus on 0 – 6 of the last 6 trials of each game **C.** Same as A, but performance separated by whether the game was "learned" or "unlearned". **D.** Same as B, but separated by whether the game was "learned" or "unlearned". In learned games, subjects consistently chose the correct stimulus, with perfect performance (6 out of 6 trials correct) 71.05% of the times. In unlearned games, performance on the last six trials was near chance (correct choice on 30.3% of trials).

Trial-By-Trial Analysis

The Dimensions Task is a very rich game in terms of the complexity of the task itself and in terms of the information that can be gathered/inferred from the way subjects play the game. On each trial, subjects choose one of three stimuli, each composed of three different features. Shown below are twelve trials from the beginning of one game of the Dimensions Task, highlighting the stimulus chosen on each game and whether the choice was rewarded (Figure 3.2A). This example game demonstrates the difficulty of ascertaining the subject's thoughts about the identity of the target feature solely on the basis of choice. In the first three trials, when Bill Gates, the Taj Mahal, and the wrench all appeared in the chosen stimulus multiple times, it is difficult to know whether the subject was testing three different hypotheses or concentrating on only one. Following the first unrewarded trial, in which the only feature consistent with the past three rewarded trials was Bill Gates, the subject appears to switch hypotheses, from Bill Gates to the Taj Mahal. The subject continues to choose the stimulus containing the Taj Mahal for several trials, despite being consistently unrewarded, making it uncertain whether a hypothesis was being tested or whether that landmark appeared in the chosen stimulus by chance. From this complexity, it is evident that chosen stimuli are not a clear measure of which features a subject is most interested in or is learning most about.

Below the subject's choices are plots of several other variables of interest over the same twelve trials of the game (Figure 3.2 B,C). These variables, including confidence, value, attention, and information, have all been proposed to have an effect on how subjects learn during the game and will be examined in more detail below.

Evaluating Confidence

In the dimensions task, subjects rated their confidence every three trials on a continuous scale from 0-1. In order to obtain a trial-by-trial comparison of confidence

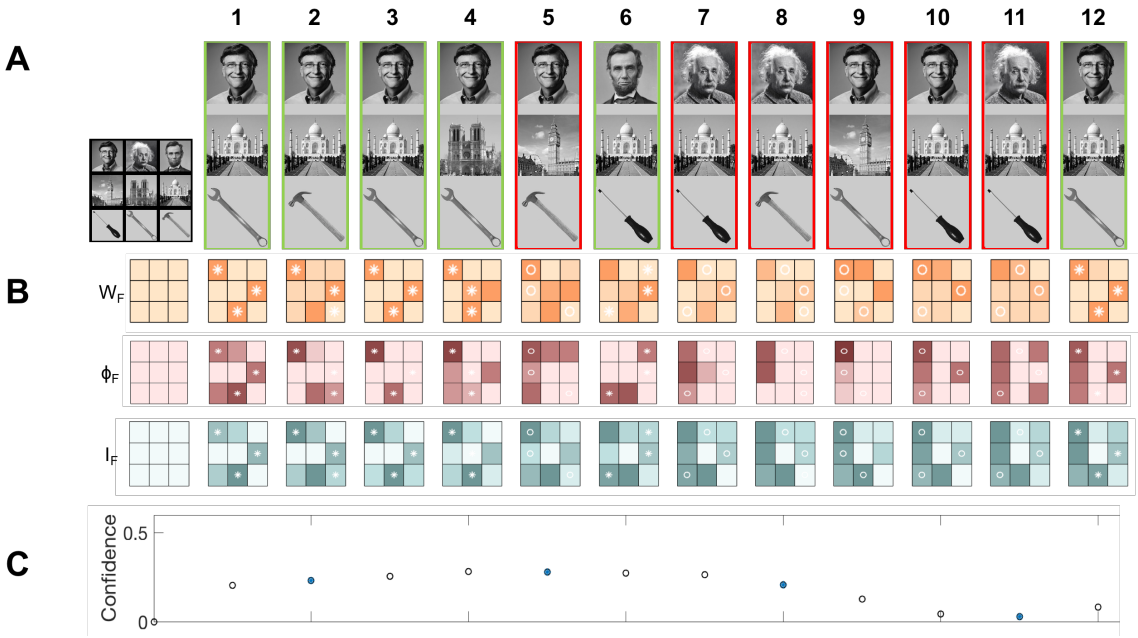

Figure 3.2: **Example Sequence of Dimensions Task.** Subject behavior and latent variables for the first twelve trials of one game of the dimensions task. **A.** Stimuli chosen by one subject on the first twelve trials of the dimensions task. Shading around choices indicates whether that trial was rewarded (green) or not rewarded (red). (*Inset-left*) Reference locations of each feature in the grid. **B.** Value (W_f), attention(ϕ_F), and information (I_F) for each feature (See section 3.3 for details). Each square on the grid corresponds to the feature in the same location in the reference grid (inset, above). Darker shading indicates a higher value/greater attention/more information for that feature. White markers indicate features in the stimulus chosen by the subject in that trial (filled = rewarded, open circles = unrewarded). Prior to first trial, all features have equal value, as shown in leftmost grid. (*Top*) Value of each feature according to the RL w/ Decay model. All features start at zero value. (*Middle*) Proportion of time on that trial spent looking at each feature. All features start out with $\phi_F = 1/9$. (*Bottom*) Information score for each feature. All features start with a score of zero. **C.** Subject's confidence ratings across the same twelve trials. Filled, blue markers indicate trials on which confidence was explicitly probed. Hollow circle markers are interpolated confidence scores.

to other metrics, we linearly interpolated confidence scores across trials within a game. Before the game started, subjects were assumed to have a confidence score of zero, as they had no knowledge of how any of the features related to reward. Setting initial confidence to zero allowed us to infer confidence over the first two trials of the game, when confidence had not yet been explicitly measured.

In the Dimensions Task, subjects did not all treat the confidence scale in the same way. For instance, the subject whose confidence scores are shown in Figure 3.3D never indicated being more than 80% confident, even in learned games when they consistently chose the correct stimulus. Other subjects quickly reported being 100% confident in their knowledge of the target feature, even when later in the game they identified a different feature as being the target feature. To control for this variability in use of the confidence scale, and to better compare confidence scores across subjects, we z-scored all of the confidence measures. Z-scores reflect how many standard deviations away from the mean a data point is, and thus serve as a more standardized measure for comparing across subjects who might have different general distributions of scores across the same interval, as is the case in our data. Within a subject, z-scoring the confidence reports did not affect the shape of the data or relative differences between scores (Figure 3.3 C,D).

On average, we found that confidence tended to increase over the course of the game, with subjects becoming more sure of their knowledge of the target feature as the game went on (Figure 3.3A). As expected, confidence was generally higher for learned games compared to games where the subjects did not learn the target feature (Figure 3.3B). Interestingly, even for unlearned games, where on average subjects were guessing at no more than chance levels in the last six trials, confidence was still significantly higher than at the beginning of the game.

This could be attributed to two different factors. First, it is possible that in a few games, subjects mistakenly believed they had correctly identified the target feature. This would lead to high confidence even though the game itself was unlearned. A second explanation would be that as subjects learn more about the different features in the game, and each feature's possible reward value, they feel more confident in their knowledge, despite not knowing exactly which feature is most rewarding. For instance, after fifteen trials, a subject may have tested and discarded several different

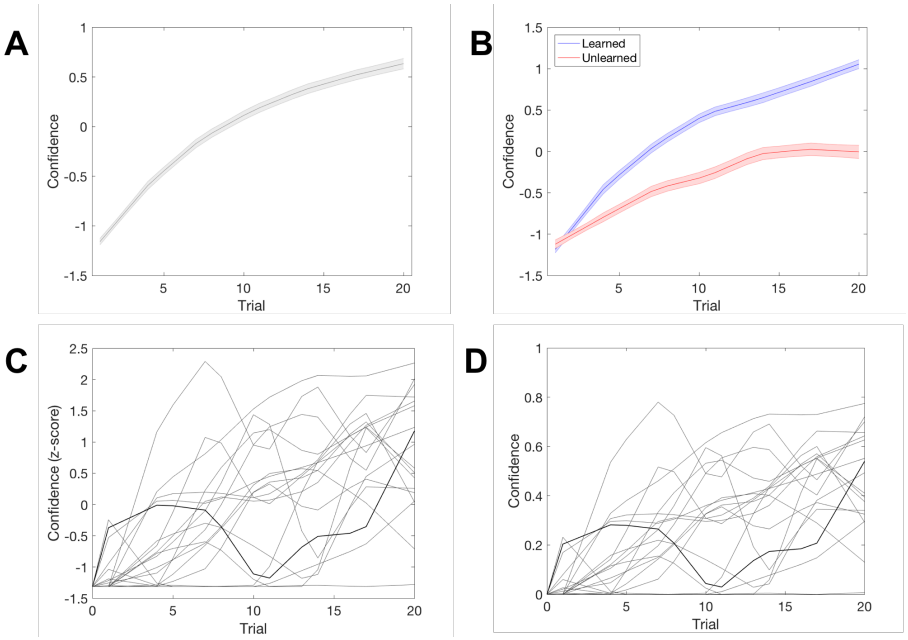

Figure 3.3: **Confidence increases over time in both learned and unlearned games**. **A.** Average confidence over the course of the game. Reported confidence values were interpolated and z-scored by subject, then averaged. **B.** Same as A, but separated by learned and unlearned games. **C.** All confidence reports for one subject. Each line represents one game. The darker line is the same game shown in Figure 3.2. Reports were highly variable across games. **D.** Same as C, but showing original confidence scores, prior to transform through z-scoring. Shape of the data and relative differences within subject are unaffected by z-scoring.

hypotheses. Although they have not identified the correct feature, and thus continue to perform at close to a chance level, their greater amount of information about the various features may give them a feeling of greater confidence.

Evaluating Attention

We are interested in studying how confidence affects the way in which we distribute our attention as we learn about features in our environment. In the Dimensions Task, we measured attention using an eye-tracker that captured where subjects were looking at any given moment, from the time the subject started playing to the end of the final game. To capture task-relevant information, we confined our analysis to attention only during trials (moments when the stimuli grid was on the screen) and

only to stimulus features and dimensions (ignoring moments when subjects looked at blank areas of the screen surrounding the grid or off the screen entirely). Within each trial, time was further broken up into a "choice" period and a "learning" period. The choice period extends from the time the stimulus display appears to the time they receive feedback about their choice. The learning period starts the instant they receive that feedback and lasts for the duration of the outcome period until the stimuli disappear from the screen and the inter-trial interval begins. Past analysis indicates that attention is distributed differently during choice and learning (Leong et al., 2017). This breakdown allowed us to examine if there was any differential effect of confidence on attention across the two periods.

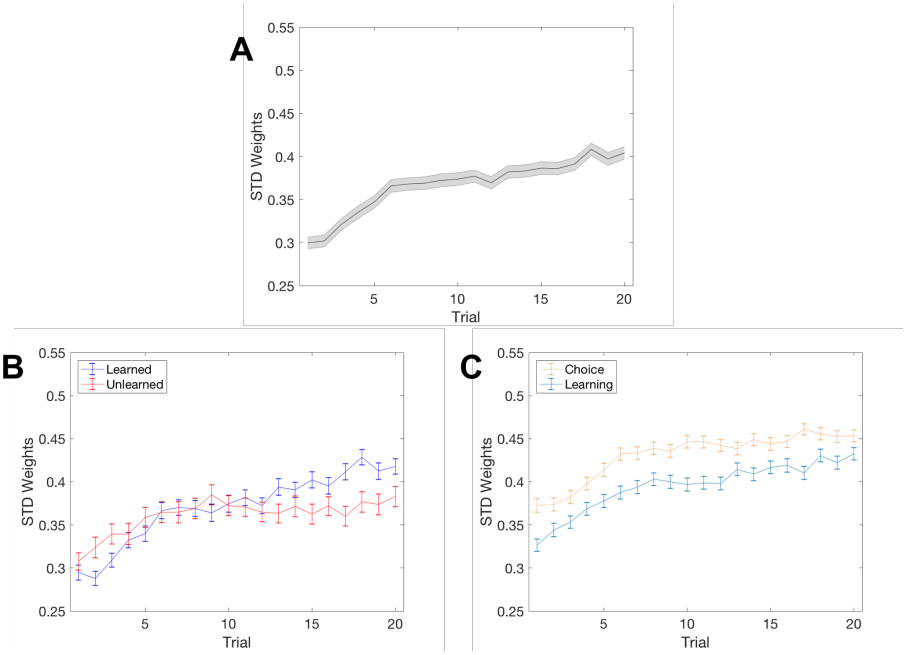

Figure 3.4: **Attention narrows over the course of games**. Attention bias was computed as the standard deviation of dimensional attention weights for that trial. Shaded region and error bars indicate SEM. **A.** Average attention bias across all subjects and games. **B.** Attention bias, separated by learned and unlearned games. **C.** Attention bias during choice and learning. Average attention during the choice period was focused more narrowly than during the learning period, across all trials of the game.

We measured subjects' attention bias (how strongly they attended to one dimension rather than the other two) by computing the standard deviation of ϕ_D – the vector of attention weights corresponding to the proportion of time subjects were looking at each of the three dimensions. A low standard deviation implies that attention was distributed relatively uniformly across all three dimensions, whereas a high standard deviation indicates attention was directed to a smaller set of dimensions. Replicating the results of Leong et al., we found that attention narrowed as the game progressed (Figure 3.4A), consistent with subjects identifying and prioritizing attention to the relevant dimension.

Further supporting the idea that subjects narrow their attention as they identify the target feature, attention bias was stronger near the end of learned games compared to unlearned games (Figure 3.4B). Interestingly, in comparing learned and unlearned games, attention in the first few trials was narrower for unlearned games. This stronger early bias could indicate subjects mistakenly latching onto a feature early in the game, only to realize later they are incorrect, and subsequently broadening their attention as they explore other options. In contrast, if subjects took more time to accumulate information about the reward probabilities of different features, their attention would initially be broader, but they would subsequently be more likely to have identified the correct feature and thus to have narrow attention for that feature/dimension by the end of the game.

Interaction Between Confidence and Attention

The distance in attention bias between learned and unlearned games suggests that having better knowledge of the target feature is associated with narrower attention. But what causes the interaction between learned games and narrow attention? Or, in other words, what determines when a participant will narrow (or broaden) their attention? As a subject does not, while playing, know whether the game is correctly

learned or not, he must make judgments on the basis of some internal variable. In looking at both attention bias and confidence, there is a similarity in trends when comparing learned versus unlearned games over the course of the task. This suggests that confidence could be the basis on which decisions about attention distribution are made.

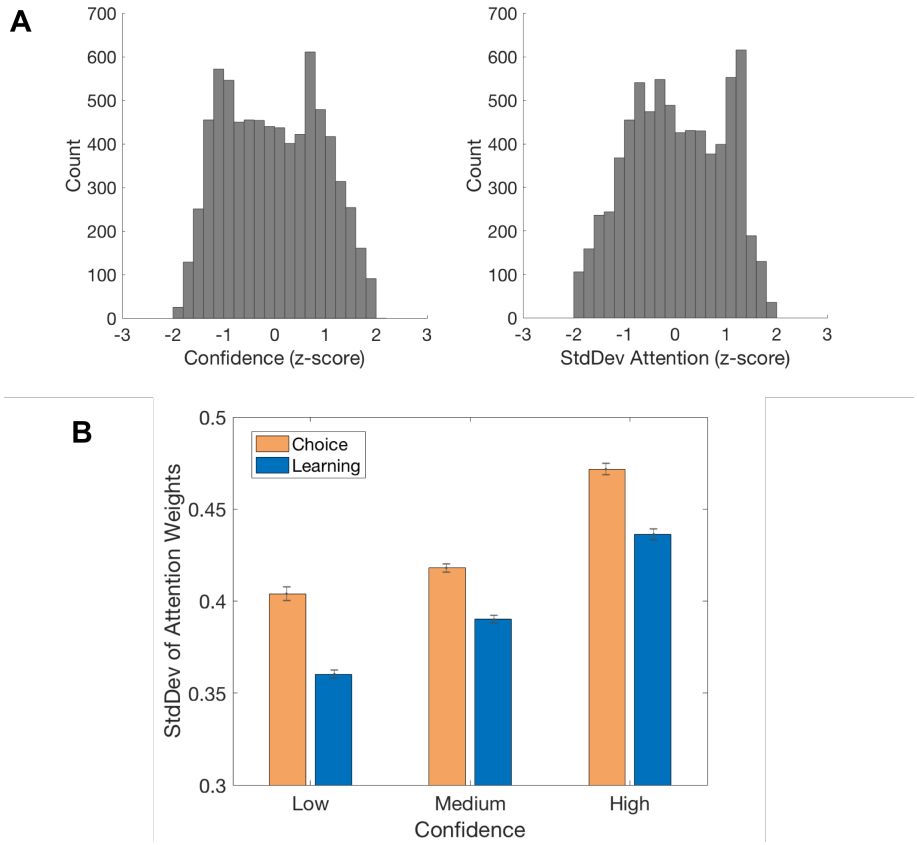

Figure 3.5: **Interaction Between Confidence and Attention**. **A.** Distribution of confidence scores (left) and attention bias (right), across all all trials, games, and subjects. Both confidence scores and the standard deviation of dimensional attention (ϕ_D) were standardized through z-scores in order to compare across subjects. Both distributions are roughly bimodal, with one cluster of lower peaks and one cluster of higher peaks. **B.** For every trial, attention bias (standard deviation of ϕ_D) was calculated for both the choice and learning periods, and then binned according to whether the subject had low, medium, or high confidence on that trial. Confidence bins were determined by subject, with the lowest 1/3 of that subject's confidence scores in the "low" bin, the highest 1/3 of their scores in the "high" bin, and the rest in the "medium" bin. Because confidence binning was by subject, there are no consistent cut-offs for which raw confidence scores correspond to which bin.

If a subject's confidence affects the way they distribute their attention during learning, we would expect to see similarities not just in their average trends over time, but also in the shape of the two distributions, as is indeed the case (Figure 3.5A).On a trial-by-trial basis, we would also expect to see a correlation between confidence and how distributed attention was on that trial. To look at this, we binned attention bias based on the confidence score for the corresponding trial (Figure 3.5B). We found that lower confidence is associated with a lower standard deviation of attention weights, corresponding to a wider distribution of attention. Similarly, high confidence is associated with more biased attention, indicating a narrower focus. These results hold for both attention during the choice period and attention during the learning period.

3.3 Latent Variables

Overview

Investigating the relationship between confidence and attention is particularly complicated given the number of other variables entangled in the analysis. In particular, the latent variables of value and information are likely to be implicated in how subjects determine their own confidence levels and how they allocate their attention during learning.

Computational Modeling of Value

Reinforcement learning has proved to be a powerful tool for understanding how people learn in low-dimensional environments. Through trial-and-error feedback, subjects learn to associate values (corresponding to future expected reward), with different features. Previous studies have shown that value is an important predictor of both choice and attention during the Dimensions Task, and that attention has

an effect on the learning of values (Niv et al., 2015; Leong et al., 2017). Niv et al., found that a computational model incorporating reinforcement learning with an added decay for unchosen stimuli best explained subjects' choice behavior, better even than Bayesian optimal models. Here, we adopt this fRL + decay model in order to examine the effect of value on the distribution of attention, and its relationship to confidence. The principal aspects of the model are summarized below (See Niv at al., 2015 for full details).

In the fRL + decay model, reinforcement learning is used to learn value weights for each of the nine features. The model assumes that subjects linearly combine the values of features within a stimulus to obtain a compound stimulus value:

$$V(S) = \sum_{f \in S} W(f). \tag{3.1}$$

Following choice, the weights of features within the chosen stimulus are updated according to a modified temporal difference learning algorithm:

$$W^{new}(f) = W^{old}(f) + \eta[R_t - V(S_{chosen})] \quad \forall f \in S_{chosen}, \tag{3.2}$$

whereas the weights of features not in the chosen stimulus are decayed to 0:

$$W^{new}(f) = (1-d)W^{old}(f) \quad \forall f \notin S_{chosen} \tag{3.3}$$

In these equations, η and d are subject-specific rate parameters, and R_t is the reward on that trial (either 0 or 1). On each trial,a softmax decision rule is used to determine which stimulus is most likely to be chosen, based on the compound weights of each stimulus.

$$p(chooseS_i) = \frac{e^{\beta V(S_i)}}{\sum_{j=1}^{3} e^{\beta V(S_j)}} \tag{3.4}$$

Here β is the softmax inverse temperature parameter. The inverse temperature parameter determines the noisiness of the choice, with a high value indicating a more deterministic decision and a low beta indicating a noisier, more random choice. The values of free parameters are fit for each subject using trial-by-trial choice behavior and asking to what extent the model explains participants' choices. The best fit parameters assigned to each subject can then be used to compute predicted feature, dimension, and stimulus values over the course of the Dimensions Task. Figure 3.2B shows an example of feature values computed over the course of the first twelve trials of one game. The values of features that were in a rewarded stimulus became higher (darker colors), while the values of unchosen features gradually decay. Unrewarded trials can lead to more rapid decreases in feature value, especially when the value of a feature in the chosen stimulus was high. All feature weights start out at zero.

Information Score

Another latent variable we were interested in looking at was information. The Pearce and Hall theory of attention suggests that attention should be directed to novel information, or to areas in the environment about which least is known. Evaluating this claim requires some metric of how much knowledge, or "information", subjects have about each feature.

One major takeaway from selective attention theory is that subjects do not gather information about every feature equally. Although different theories disagree about which elements of a scene or environment are most likely to be attended to, there is broad consensus around the claim that whichever features are attended to most during learning are subsequently better processed by the brain (Cowan & Wood, 1997; Reviewed in Desimone & Duncan, 1995; Buschman & Kastner, 2015). Thus, knowledge of their reward value is more likely to be stored by the subject. Accordingly, we computed an "Information" metric representing the cumulative time spent looking

at each feature over the course of the game so far. The measure should be a proxy for how much knowledge has been accumulated by the subject about each feature, and allows us to make distinctions about "low-information" versus "high-information" features. Figure 3.2B shows an example of information scores (I_f) for all nine features, computed over the first twelve trials of one game. Darker colors indicate higher scores, consistent with subjects having looked at that feature comparatively more often over the course of the game. While information is derived from trial-by-trial attention, the two measures are not identical, especially as the game progresses.

3.4 Regression Analysis

In order to quantitatively assess the contribution of confidence to the distribution of attention, we fit a generalized linear mixed-effects regression model to the data (Table 3.1). Reaction time (RT), confidence, maximum value, and trial number were tested as possible predictors of attention bias (the standard deviation of dimensional attention ϕ_D). Shuffled trial order, which logically should have no impact on attention bias, was included as a control. To help account for subject-specific variation, subject number was incorporated into the model as a random-effects variable.

With the exception of the shuffled trial order, all tested predictors were found to have a significant contribution ($p < .001$) to attention bias (Table 3.1, Figure 3.6). Reaction time had the single greatest effect, followed by confidence, and then maximum value, as can be seen by the regression coefficients of the GLME model (Figure 3.6). When reaction time (which might be expected to vary with confidence) was not included in the analysis, the effect of confidence was even greater (data not shown).

The results of this regression analysis demonstrate the contribution of confidence to the breadth of attention. Even when maximum value, reaction time, and trial

number were included in the regression as possible predictors, confidence still had a significant effect on attention bias. This, together with our earlier findings about the interaction between confidence and attention, substantiates our hypothesis that confidence is involved in allocating attention during learning. To explore exactly how confidence affects that process, and how it relates to theories about what governs attention decisions, we turn to computational modeling.

Predictors	Estimate	SE	tStat	pValue
(Intercept)	-0.1493	0.0334	-4.459	8.36e-06
RT	-0.1929	0.0114	-16.914	5.33e-63
Confidence	0.1308	0.0163	8.020	1.22e-15
Max Value	0.0598	0.0169	3.521	4.32e-4
Trial #	0.0113	0.0024	4.799	1.63e-06
Shuffled Trials	0.0026	0.0019	1.386	0.166

Table 3.1: **Linear Regression of Attention Bias** A generalized linear mixed-effects model was used to analyze the contribution of various possible predictors to attention bias (standard deviation of ϕ_D). Reaction time (RT), confidence, the maximum value of the nine features, trial number, and shuffled trial order (random permutation of trial number) were included as possible predictors. Subject number was included in the regression as a random-effects variable. In order to compare across subjects, all continuous predictors (RT, confidence, max value) and the standard deviation of attention were z-scored (within subject) prior to being included in the analysis. The table displays the output of the GLME model, including the regression coefficient ("Estimate") for each possible regressor, as well as the associated standard error, t-statistic, and p-value.

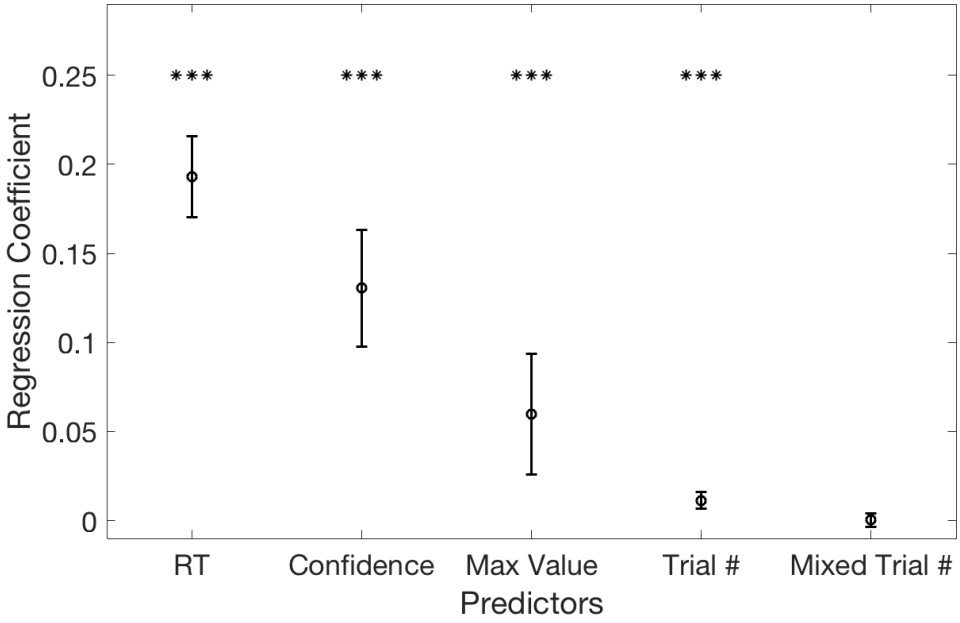

Figure 3.6: **Regression Coefficients of GLME Model.** Coefficients of a linear regression to predict attention bias. Reaction time, confidence, maximum value, trial number, and shuffled trial order were included as predictors. Reaction time, which had a negative effect, is shown here as its absolute value for the purposes of comparison with other predictors. Asterisks above the plot indicate significance ($p < .001$). Full statistics can be found in Table 3.1.

Chapter 4

Computational Modeling of Attention

4.1 Predicting Attention Across Dimensions

Previous studies indicate that there may be a bidirectional interaction between learning and attention. People not only use selective attention to restrict the dimensionality of learning, but they also learn over time which dimensions of a task are relevant for attention. How people decide where to allocate their attention, and how this changes over the course of learning, is an open question. To gain insight into this ambiguity, we tested two novel computational models, each of which makes different claims about how confidence modulates the way attention is distributed during learning.

Modeling Attention

Uniform Attention:

One of the simplest models of attention is a parameter free model that assumes uniform attention to all dimensions. Under the uniform attention model, attention is distributed equally across the three dimensions on every trial.

$$\phi_D = [1/3 \quad 1/3 \quad 1/3]. \tag{4.1}$$

This model, while useful as a baseline comparison, is unlikely to be a good reflection of subjects' attention. Indeed, as the model would suggest that subjects neither update their attention in response to task changes nor use attention to narrow the number of features/dimensions they are learning about, it stands in direct opposition to most claims about the utility of selective attention during learning (Wilson and Niv, 2011; O'Doherty, 2012; Dayan et al., 2005, Leong et al., 2017). However, the model still serves as a simple, zero-parameter standard against which to compare the performance of our more complex models that seek to explain why and how attention changes during learning.

Value-Based Attention:

This model, from Leong et al., 2017, predicts that attention will track feature values, with greater attention being given to dimensions with more highly valued features. All feature values were initialized at zero and then updated through a reinforcement learning paradigm. Following the reinforcement learning with decay model presented in Niv et al., 2015., on each trial, the values of features in the chosen stimulus were updated according to a reward prediction error while the values of unchosen features decayed toward zero. Both update and decay were scaled according to subject-specific rate parameters (η and d). To obtain the prediction of attention for the next trial, the maximum feature value in each dimension was passed through

a softmax function with inverse learning parameter β. The output of the softmax was a vector of three attention weights that sum to 1. The attention weights correspond to the percentage of time subjects are predicted to spend looking at each dimension. (See Section 3.3 for a more thorough summary of fRL + decay, and related equations.)

Biased Value and Information:

Building off the classic explore-exploit paradigm, this model formalizes the idea that there should be a trade-off in strategies of attention distribution, and that this trade-off should depend on a subject's confidence. According to Mackintosh's classic theory, attention should be directed to features that are most predictive of reward (1975). In this view, rewarding features are more likely to be attended to both because of their greater reward-linked saliency and because greater attention to, and thus greater learning for, those features is likely to result in greater future reward. This corresponds to the exploit portion of the explore-exploit paradigm, when organisms try to maximize future reward by choosing features that, according to their current knowledge, are most rewarding. Such a strategy, however, contains the potential pitfall that the feature currently thought to be most predictive of reward might not truly be the most rewarding in the long run. Thus, in order to maximize total reward, an organism should be sure to acquire enough information to maximize its chances of identifying the most rewarding feature, even if in the short-term that is likely to result in fewer rewards. Consistent with this exploratory pattern, the Pearce and Hall theory suggests attention should be directed to features about which least is known, allowing for the accumulation of new evidence (1980). Consequently, if a feature's relationship to reward is already reliably established, it should receive less attention because that attention is not likely to result in as much learning. In the framework of the Dimensions Task, these two opposing theories can be conceptualized as a trade-off between attending to highly-valued features and attending to features with low information.

What governs the balance between the two strategies is an open question. If attention to low information features does correspond with a more exploratory strategy, and attention to high value features does correspond to a more exploitative strategy, then what determines which strategy (or possibly, to what extent each strategy) is used during different phases of the learning process?

Theoretical findings indicate that confidence could be an ideal arbiter between the two, as confidence reflects both the reliability of and uncertainty about the current knowledge of the environment (Auer, 2003). Accordingly, low confidence, indicating a lack of certainty or reliability in the reward environment, would bias attention toward gathering more knowledge about features whose reward predictability is less well known. High confidence, indicating greater certainty, would bias attention towards exploiting the high-valued features already seen as reliably predicting reward.

Adding to this theoretical support, our regression results indicate that confidence has a significant effect on the distribution of attention during the Dimensions Task, suggesting it is involved in the process of attention allocation. Here, we propose a hybrid model in which confidence directly biases the balance between the two different selective attention approaches described above.

In our model, the distinction between attention to highly-valued features versus attention to low-information features is formalized by assigning to each feature a hybrid score X_f that includes a contribution from both the feature's value W_f and information I_f weights:

$$X_f = \alpha(1 - I_f) + (1 - \alpha)W_f. \tag{4.2}$$

The parameter α biases the relative contributions of value and information to the total score for each feature and is modulated by confidence according to the sigmoid:

$$\alpha = \frac{1}{1 + e^{\lambda(Confidence - \theta)}}. \tag{4.3}$$

The center θ and slope λ of the sigmoid are treated as free parameters that together indicate how early and how rapidly each subject tends to switch between favoring the low-information regime and the high-value regime as their confidence increases. The resulting hybrid score (X_f) for each feature is used in place of the value weights in computing predicted attention. As in the Value-Based Attention model, the score of the maximal feature in each dimension is used as input to the softmax, which then produces predicted probabilities for each of the three dimensions:

$$\phi_{D_i} = \frac{e^{\beta X(D_i)}}{\sum_{j=1}^{3} e^{\beta X(D_j)}}. \tag{4.4}$$

Feature values are extracted from the same reinforcement learning paradigm described in the Value-Based Attention model, giving the additional free parameters η for learning rate and d for decay.

When confidence is high (suggesting the presence of at least one feature that is reliably rewarding), our model predicts that subjects will be in a more exploitative phase and attention will be directed to features with a high value. When confidence is low (suggesting uncertainty about how features map to reward), our model predicts that subjects will be more inclined to explore, and thus attention will be directed to features with low information scores. In our initial approach, value for each feature was computed according to the fRL + decay model. Information for each feature was computed as a cumulative sum of looking time over the course of the game so far (See section 3.3).

Confidence-Modulated Gain:

Another possible way that confidence could affect the distribution of attention is by directly modulating the breadth of attention. This proposal reflects the similarity in trends we observed between confidence and attention bias over the course of learning. Our earlier results show confidence increasing and attention narrowing as

subjects learn, reflected both by increases across time and by significant differences between learned and unlearned games (Figure 3.3, Figure 3.4). This phenomenon supports the possibility that confidence directly impacts how biased attention is during learning.

In this conception, low confidence reflects a high degree of uncertainty about the reward probabilities associated with different feature values, whereas high confidence reflects a high degree of certainty. Instead of this uncertainty biasing what type of features are seen as important (low-information vs. high-value) as in the previous model, here uncertainty changes the scope of features being attended to, modulating between broad attention to a high number of features or narrow attention to a few features.

When subjects are more confident and more sure in their knowledge of which feature is most predictive of reward, there is less incentive to attend to other stimuli. Thus, their allocation of attention should reflect a more greedy algorithm, where the highest valued features are awarded the vast majority of attention, and lower valued features receive diminished attention. Conversely, when subjects are less confident, they will seek information about a wider array of choices. By attending more broadly, they will be more likely to gain knowledge about which features are linked to reward.

This hypothesis is similar to ideas formulated by Dayan et al. surrounding how the values of different features should be incorporated into decision-making. In their model of selective attention in classical conditioning paradigms, Dayan et al., highlight two important principles which affect the distribution of attention: uncertainty and unreliability (2000). Uncertainty is attached to the reward prediction associated with each stimulus and measures the amount of evidence associated with the stimulus. Unreliability concerns the relationship between the true value of the reward and the prediction error associated with each stimulus. These measures are formalized into statistical models governing how learning and responsibility for making predictions

should be competitively allocated among stimuli. While these models were not formulated with visual attention prediction in mind, aspects of them still bear on questions of how attention is likely to be distributed during learning. Dayan et al., use estimates of stimulus uncertainty and reliability to modulate, respectively, the rate at which an animal learns and how much different stimuli contribute when making predictions about reward. These modulations are similar to the way in which we believe confidence might modulate how widely attention is distributed on each trial.

Adjusting this idea of trial-by-trial modulation to our task, we created a model in which subjects' reported confidence affects how widely attention was allocated in response to differences in feature weights. As in the Value-Based Attention model, this model assumes that subjects will give greater attention to dimensions with high valued features. In our model, however, the softmax inverse temperature parameter β, which controls how strongly attention is biased toward the maximally-scored dimension, is modified by the confidence associated with that trial:

$$\phi_{D_i} = \frac{e^{B_t * W(D_i)}}{\sum_{j=1}^{3} e^{B_t * W(D_j)}}; \quad B_t = \beta * Confidence_t. \tag{4.5}$$

When subjects are less confident, the softmax temperature will be lower, leading to a noisier allocation of attention and a more equal distribution of attention. Conversely, when confidence is high, the inverse temperature will also be higher, leading to a more biased allocation of attention in favor of the dimension with the highest feature value, and a correspondingly narrower breadth of attention. If subjects modify how widely they distribute their attention based on confidence, we would expect this model to perform better in our comparison than the Value-Based Attention model, in which the breadth of attention is solely dependent upon the differences between the values of the features.

Model Optimization and Evaluation

Models of attention were evaluated according to how well they predicted attention to each of the three dimensions on every trial, using a leave-one-game-out cross-validation procedure. For each subject and game, optimal parameters were obtained by minimizing the model's prediction error, which was calculated as the root mean squared deviation (RMSD) between the predicted vector of attention weights (obtained from the model) and the actual vector of attention weights (obtained from eye-tracking analysis). The fitted parameters were then used in the model to predict attention on the left-out game. For each model, the mean RMSD per trial was calculated from the prediction errors in the left-out game.

4.2 Model Performance and Comparison

If confidence dynamically determines how attention is allocated during learning, we would expect that our models which incorporate confidence into attention predictions would do better than those that that use value alone, or that assume uniform attention.[1] We tested this approach by first comparing subject-by-subject RMSDs obtained using the best parameters from model fitting. A lower RMSD indicates a shorter distance between the actual and predicted attention vectors, and thus a better prediction.

We found that across subjects, the value and confidence models generally did better at predicting attention compared to the model that assumed uniform confidence (Figure 4.1). While for some subjects the differences between models were quite small (e.g. subjects 1, 4, 6), other subjects' attention was distinctly better represented by

[1]Throughout this section, we use the terms confidence-based models and value-based model to separate our three main models. While all three use value to inform the attention predictions, for the ease of discussion it is useful to distinguish between the novel hypotheses that incorporate confidence and the baseline value model which does not. Therefore, the original Value-Based Attention model is referred to as the value-based model while the two new models ("Biased Value and Information" and "Confidence-Modulated Gain") are together referred to as the confidence-based models.

one or a few of the models (e.g. subjects 9,12,18). To quantitatively compare the performance of each of the models, we averaged RMSDs across all subjects (Figure 4.1B). Contrary to our expectations, the Value-Based Attention model had the lowest average RMSD, doing slightly, although not significantly, better than either of our confidence-based models. This is a surprising result given that the confidence models, while based on value, incorporated additional metrics (i.e., confidence) that should have given the models extra information, leading to a better prediction. This is especially the case in the Biased Value and Information model, which uses both value and information to calculate scores for features.

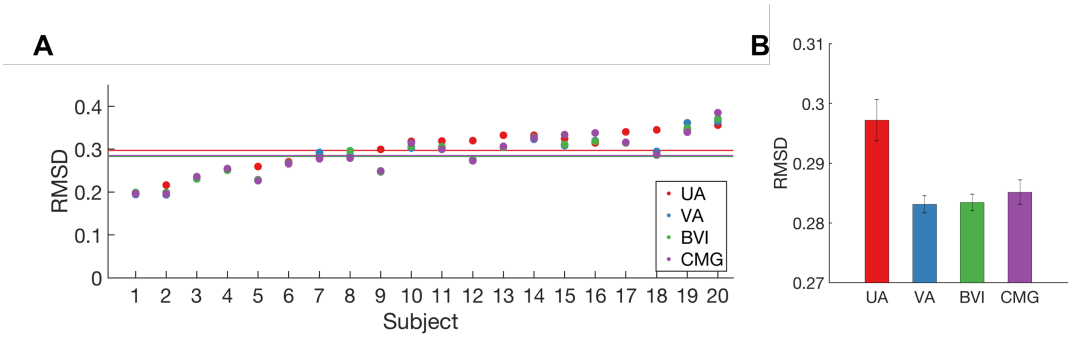

Figure 4.1: **Initial Model Performance. A.** Model performances for each subject, measured by average RMSD. Models: UA – Uniform Attention; VA – Value-Based Attention; BVI – Biased Value and Information; CMG – Confidence-Modulated Gain. RMSDs for each subject were obtained by averaging across all the per trial RMSDs generated from running the model using the best fit parameters for that subject. Scatter shows model results for each subject, horizontal line of the same color indicates the average RMSD across subjects for that model. Subjects ordered by mean performance of models. Model optimization for one subject was unable to converge, so that subject was excluded from the analysis. **B.** Average RMSD for each of the four initial models (abbreviations and colors as in A). Bars correspond to the horizontal lines in A. Error bars show across-subject SEM. Average RMSDs for VA, BVI, and CMG models were all significantly different from uniform attention model (VA-UA: p = .0054; BVI-UA: p = .0078; CMG-UA: p = .0294; paired sample t-test). Differences between the other three models were not significant.

Results and Discussion: Biased Value and Information

To understand these unexpected findings, we looked more closely at the best fit parameters for each of the models. If model optimization routinely hit bounds, or

was returning nonsensical parameter values, this would help to explain the poorer performance of these models. Model fits are summarized in Table 4.1.

The hypothesis underlying the Biased Value and Information model suggests that the balance of attention to low-information versus high-value features should exist on a curve determined by confidence. As confidence shifts, the shape of the curve dictates whether information or value will have greater input to the overall score assigned to that feature, which will directly impact how attention is distributed. We chose to model this curve as a sigmoid whose center θ and slope λ were free parameters determined by the model optimization process (Equation 4.3). According to our hypothesis, this fitting should have resulted in a parameterization that began with high α (biased toward low-information) and ended with low α (biased toward high-value) as confidence increased (See Equation 4.2). However, this is not what we see in the model fits. Instead, the average parameter values indicate that subjects tended to bias towards value for the entirety of the game. The shape and center of the sigmoid were, on average, so steep and so far to the left that for the available confidence range (0-1), the resulting sigmoid α value was always 0, meaning the model only used value to determine attention. Therefore, in the majority of cases (75% of subjects), the model acted identically to the Value-Based model, leading to a correspondingly similar performance in prediction. The slightly greater error overall could be due to the remainder of subjects, all but one of whom had free parameter values that hit against bounds and thus represent a poor model fit.

There are several possible explanations for why the model fitting was so contrary to our presumed results. First, the results could indicate that people's attention only ever operates in a value-maximizing, reward-seeking manner, according to the most stringent version of the Mackintosh theory.

Another possible explanation has to do with the structure of our task. Unlike many real world scenarios, when multiple aspects of the environment are likely to

Model	Parameter	Mean (SEM)	Range	Prior
Value-Based	η (learning rate)	.232 ±.045	0-1	None
	β (softmax inverse temperature)	16.3 ±11.2	1-500	Gamma (2,3)
	d (decay)	.299 ±.059	0-1	None
Biased Value	η (learning rate)	.243 ±.045	0-1	None
+ Information	β (softmax inverse temperature)	12.0 ±7.20	1-500	Gamma (2,3)
	d (decay)	.432 ±.087	0-1	None
	λ (sigmoid slope)	3.73 ±.763	0-500	None
	θ (sigmoid center)	-2.64 ±.734	-5-5	None
Confidence-	η (learning rate)	.265 ±.046	0-1	None
Modulated Gain	β (softmax inverse temperature)	17.5 ±8.54	1-500	Gamma (2,3)
	d (decay)	.363 ±.072	0-1	None

Table 4.1: **Free Parameters and Best Fits for Each Model**. Parameters were confined to lie within the specified range of values. Running up against boundaries would generally indicate a poor model fit and/or a model design that was a poor predictor of the data. To aid model fitting, softmax inverse temperature β was always regularized with a prior distribution that favored realistic values.

have varying levels of reward, the Dimensions Task contains only one target feature. Knowing only one feature is predictive of reward, subjects have less incentive to be exploratory and acquire knowledge about the remaining features. So long as subjects have any hypothesis about which feature is likely to be the target feature, they might prioritize value judgments, regardless of how confident they report feeling. It might only be when subjects have mistaken hypotheses, such as when previously high-valued features are proven less likely to be the target feature, that their lower confidence would lead them to attend more to low information features which as of yet have not been linked one way or the other with reward. This being the case, using confidence to bias the importance of value versus information would not be a good predictor of subjects' attention in this task, although that does not preclude the model from being a good description of people's attention strategies in environments with more varied risk and reward.

A third possible explanation for this result has more to do with the way in which we conceived of the separation between attention to value and information. Our model assumed that subjects would treat every feature equally with respect to the

balance between attention to low-information versus high-value features. However, a pitfall of this assumption is that high-value and high-information are difficult to separate in our task (features only have high values if they have been chosen and rewarded many times, which implies the subject has higher information about that feature). Therefore, any bias towards the low-information metric would tend to directly discount the highest valued features.

A hypothetical scenario demonstrates the problematic nature of this assumption: At the beginning of the game, the subject chooses a random stimulus and is rewarded for the choice. On the next trial, he picks the "face" feature from the previous rewarded stimulus, and is again rewarded. As the subject begins to learn, he is likely to attend at least slightly more to whichever feature currently has the highest value, in this case, the face. However, even though the subject is in a hypothesis-testing, exploitative regime, because he has not yet had time to substantively test his hypothesis, he may indicate that he is not very confident. Thus, a model which only looks at value would capture this situation better, whereas one using confidence to modulate the balance between information and value would say the subject should attend to low-information features even if those features are not associated with reward.

When confidence is mid-range, especially near the beginning of the game, including low information in the attention score might be particularly harmful, as it would discount the most looked-at feature, which is likely the hypothesis the subject is testing. Indeed, we found that attention on the previous trial was actually by far the best predictor of attention on the next trial, suggesting a significant perseverance of attention across trials (Figure 4.2). As our information metric is derived from trial-by-trial attention, the high performance of the "model" that uses previous attention to predict current attention suggests that any model which indiscriminately privileges low information features is likely to suffer from greater prediction errors.

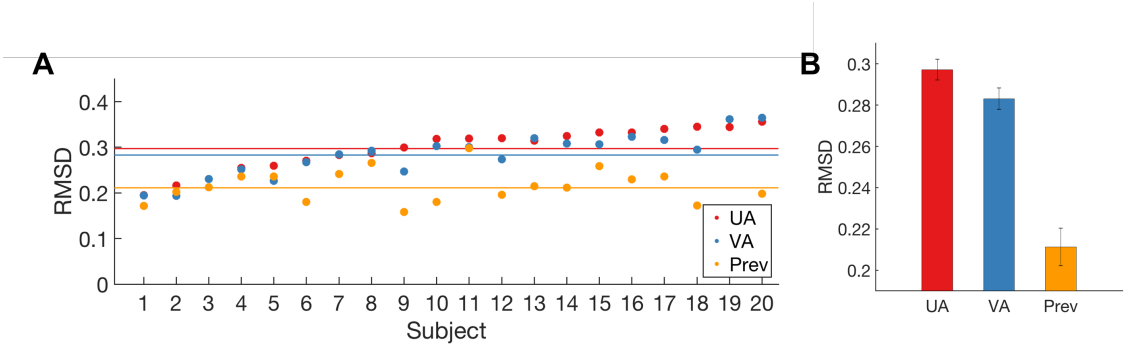

Figure 4.2: **Comparison to Previous Attention. A.** Model performances for each subject, measured by average RMSD. Models: UA – Uniform Attention; VA – Value-Based Attention; Prev – Previous Attention. The previous attention "model" has no parameters; it is simply a comparison between the vector of attention weights on the previous trial and the vector of attention weights on the current trial. In other words, it asks to what extent previous attention is a predictor of current attention. **B.** Average RMSD for each model in A. Bars correspond to the horizontal lines in A. Error bars show across-subject SEM. Differences between all models were significant (UA-VA: p = .0054; UA-Prev: p = 5.2e-6; VA-Prev: p = 5.6e-5; paired sample t-test)

Results and Discussion: Confidence-Modulated Gain The other hypothesis we were interested in testing was whether confidence modulates the breadth of attention during learning. Based on our regression results, confidence had a strong influence on the standard deviation of attention. Therefore, we would expect that this model, which uses confidence to either narrow or broaden the distribution of attention on each trial, should perform better than the value-based model, which has no such modulation. However, looking at the modeling results, this does not appear to be the case: The Confidence-Modulated Gain model does slightly worse, not better, than the Value-Based Attention model (Figure 4.1).

Comparing the results of the modeling by subject, we found that subjects were relatively evenly split between having their attention better explained by the model that incorporated confidence into the softmax (CMG) and the original value model where confidence did not bias the softmax (VA) (Figure 4.3). One possible explanation of this result is that different people have different methods for determining how narrowly to distribute attention, with some subjects incorporating

confidence into their decisions about how to allocate attention and others just using the spread of values. Based on the strength with which confidence explained the standard deviation of attention in our regression analysis, we find this simplistic explanation unsatisfying. However, that still leaves us with the difficult task of explaining why, if the regression result is so strong, the Confidence-Modulated Gain model did not show an improvement over the Value-Based Attention model.

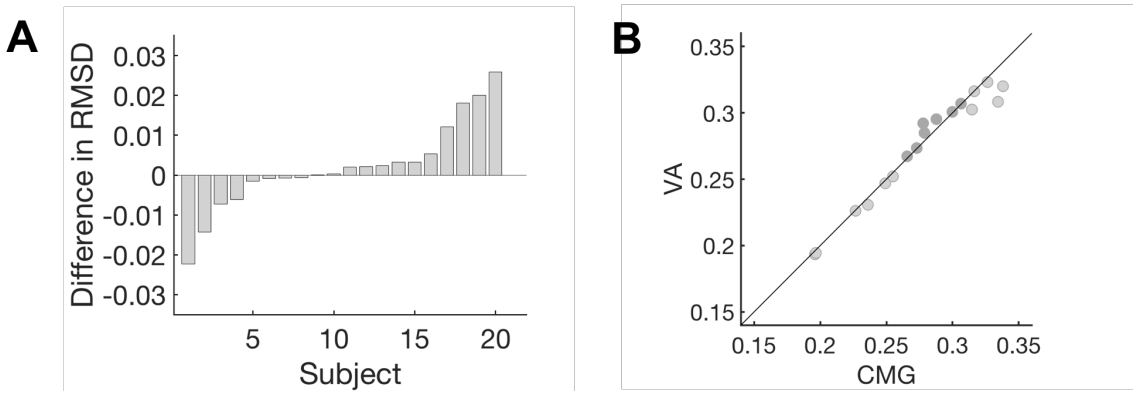

Figure 4.3: **Performance of Confidence-Modulated Gain Model versus Value-Based Attention Model. A.** Difference between subject-average RMSDs of the Confidence-Modulated Gain (CMG) model and the Value-Based Attention model (VA). For each subject, CMG model performance was subtracted from the VA model score. Since low RMSD scores indicate a better fit, a negative difference indicates that the VA model had a better prediction for that subject while a positive score indicates the CMG model did better. **B.** Scatter plot of the same information, with the RMSD scores plotted against each other. Each dot represents a single subject. Line represents equal performance. Dots that are below and to the right of the line (in light gray) have higher average RMSDs for the CMG model versus the the VA model (indicating value-based attention was a better model fit). Dots shown above the line, in dark grey, were subjects whose attention fit better to the Confidence-Modulated Gain model.

We suggest that the problem lies not in the confidence modulation, but in the feature values that are passed to the softmax function. Modulating the gain parameter of the softmax does not alter the order of the inputs in any way: whichever dimension had the highest value will still be apportioned the highest attention. Modulating the gain only affects how greedy that allocation is– how much it exaggerates the relative attention that should be given to the highest valued features. The general shape of the

prediction in the Confidence-Modulated Gain model is therefore still highly dependent upon how feature values are calculated. One side effect of this relationship is that if the value predictions are wrong– if the value model predicts the majority of attention will go to the wrong dimension– the confidence modulation will only exacerbate this effect, resulting in an even greater prediction error compared to the simpler value model. This would explain the difference in performance between the two models, as well as the difference between our expectation and the results. Consequently, the poor performance of the Confidence-Modulated Gain model suggests that the value metric is not necessarily capturing subjects' true evaluations of which features should be attended to. Instead, we might need to look at a model that incorporates other factors in order to determine which feature is most relevant for attention. If we could find that ideal combination, it is likely the confidence-modulated softmax gain would improve upon the performance of that model, surpassing both its predictions and the predictions of the value-based attention model.

4.3 Further Investigations

In our model fitting analysis, we concentrated on predicting attention to each of the dimensions across the entire length of each trial. We also chose to have as a baseline a value-based model that used fRL with decay to assign values to each feature in the game, and then used value to (at least partially) make predictions about the distribution of attention. In this section, we address some alternative modeling choices and ask how our main models perform with those measures.

Modeling Value Without Decay

As mentioned above, a Value-Based Attention model adopted from Leong et al., 2017 was used as the basis for all the models we tested (not including the uniform

attention control). The fRL w/ decay model has been shown to be a good predictor for subject's choices in the Dimensions Task, as well as in modeling their attention (Niv et al., 2015; Leong et al., 2017). However, compared to a naive reinforcement learning model, the fRL w/ decay model is potentially problematic for use in our study because of the added decay element. The neurological substrate of decay is not as well understood as other elements of the RL model, though it might represent forgetting of non-chosen stimuli. The more opaque nature of the decay could be a potential confound in our study, because it is possible that the decay absorbs attention effects that are actually attributable to confidence changes, which would provide a partial explanation for the failure of our confidence-based models to perform significantly better than the value-based attention model.

To investigate this possibility, we tested the same three models described earlier, but this time using feature RL without any decay to compute and update values. If the decay aspect is absorbing some of the effect of confidence, we would expect the confidence-based models to do comparatively better (versus value alone) in predicting attention when the models do not incorporate decay. We found that performance on all models declined when fRL without decay was used as the basis for value computations (Figure 4.4). Absolute RMSDs were higher for all the models, and model performances were also less different than the prediction of uniform attention. The decrease in performance is not unexpected given what we know about the ability of fRL w/ decay versus fRL without decay to predict subject's choices in the Dimensions Task. More pertinent to our question about the effect of decay was the finding that the confidence-based models did not have any relative advantage in this decay-less version. This suggests that the presence of decay in the RL model is not responsible for the poor performance of the models we tested.

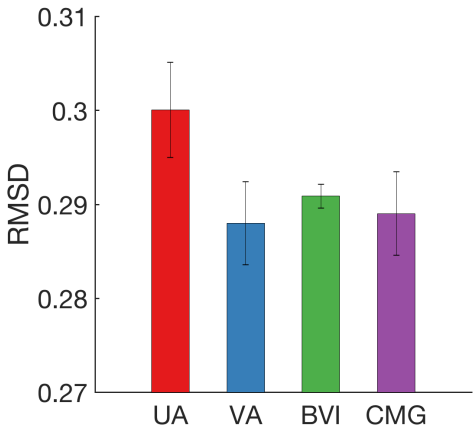

Figure 4.4: **Model Performance with No Decay** Average model performance across subjects. Model abbreviations as in Figure 4.1. The three latter models (VA, BVI, and CMG) all used an RL model that did not include a decay component for the unchosen stimuli. Otherwise, models are all as described in section 4.1. Compared to the original models, removing decay resulted in lower performance, marked by higher average RMSDs. Error bars show SEM. RMSDs for VA, BVI, and CMG models were all significantly different from uniform attention model (VA-UA: $p = .0084$; BVI-UA: $p = .0129$; CMG-UA: $p = .0235$; paired sample t-test). Differences between the other three models were not significant.

Modeling Attention Within Chosen Stimulus

In our main modeling study, we attempted to predict the proportion of time subjects would spend looking at each dimension on every trial. However, another interesting aspect of attention to model would be looking at attention just to features within the chosen stimulus on each trial. We modified our models to make predictions about the relative attention to features just within the chosen stimulus and then compared those results to the model performance when looking across the entire dimension (See Figure 4.5 for details of modifications). We found that, on average, our models were better at predicting dimensional attention than predicting attention to the features within the chosen stimulus, as measured by absolute RMSD and by comparison to the uniform attention model. Compared to dimensional attention, attention within the chosen stimulus appears to be slightly less uniform. This is

consistent with subjects predominantly attending to the two or three most highly-valued features. While the highest-valued feature is likely to be in the chosen stimulus, there may be other features of interest outside the chosen stimulus. If so, assuming the other features are in a different dimension, attention will be relatively more uniform when measuring across the dimension than when just measuring within the stimulus, where there is only one dimension with a highly valued feature. As a caveat to this explanation, we might expect attention within the chosen stimulus to be more uniform during the outcome portion of the task, when the subjects receive feedback on reward, than during the choice period. After feedback, all features in the stimulus have the potential to be associated with a reward prediction error, and it might make more sense to attend to all three features in order to maximize learning options.

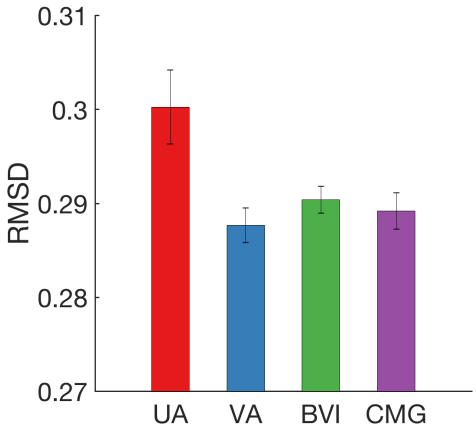

Figure 4.5: **Predicting Attention Within Chosen Stimulus.** Average model performance across subjects. Model abbreviations as in Figure 4.1. In contrast to the original models, which predicted dimensional attention by putting the maximal feature values/scores for each dimension into a softmax, these models took as input to the softmax just the values/scores of the features that were in the chosen stimulus. The resulting prediction was compared to the vector of attention weights corresponding to the proportion of time subjects spent looking at each feature within the chosen stimulus (time spent looking outside the chosen stimulus not included, attention weights summed to 1). Otherwise, models are all as described in section 4.1. Error bars show SEM. Significant difference between UA and VA models ($p = .0405$: paired sample t-test).

Attention during Choice and Outcome

This distinction between how attention might be distributed at choice versus learning is not captured in our original modeling paradigm. However, there is evidence from past studies that attention at choice and attention at learning are not equivalent, and that they have differential effects on learning (Leong et al., 2017). Our modeling results further validate this claim. One major finding was that attention at choice was much narrower than attention at learning, measured by the relative performance of the uniform attention model. The higher RMSD for attention at choice indicates that attention was less well predicted by the uniform attention model, implying that in general, attention was more narrowly concentrated on one or two of the dimensions. Across all models, RMSD measures for attention at choice and learning were higher than those for whole trial attention (Figure 4.1). Interestingly, the Confidence-Modulated Gain model did slightly better than the Value-Based attention model in predicting attention at choice, indicating that confidence may have a greater effect on attention during choice than during learning. Alternatively, confidence might be equally influential at learning, and the difference might be due to using confidence on the previous trial to predict attention during learning on the next trial. Since the learning period occurs after feedback has been given, it is likely that confidence has also changed in response to the feedback, and thus the distribution of attention might reflect the updated confidence. Further testing would be needed to make this distinction.

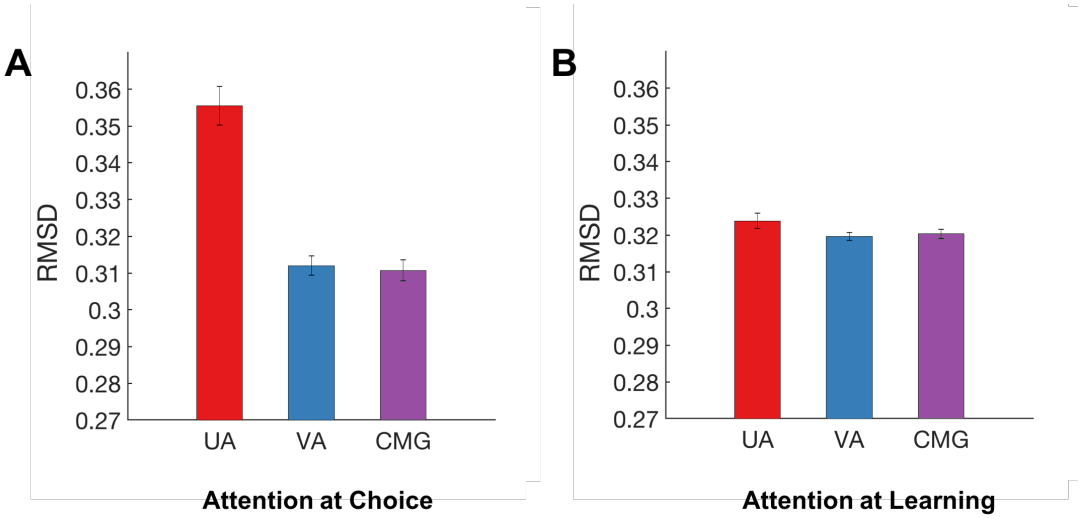

Figure 4.6: **Predicting Attention at Choice and at Learning.** Average model performance across subjects when predicting attention just during the the choice (**A**) or outcome (**B**) periods. Model predictions computed as for dimensional attention. Prediction error for choice was computed by taking the RMSD between the model prediction and the attention weights calculated from looking time just during the choice period of the trial (from beginning to just before feedback). Predictions for attention at learning computed similarly for the period after feedback until the end of the trial. Model abbreviations as in Figure 4.1. Error bars show SEM. At choice, the average RMSDs of both the VA and CMG models were significantly different from that of the UA model ($p < .0001$: paired sample t-tests). At learning, there were no significant differences between the performances of any models.

Chapter 5

Discussion

5.1 Evidence for An Interaction Between Confidence and Attention

The behavioral and modeling results presented here provide support for our hypothesis that there is an interaction between confidence and attention during learning, and more specifically that confidence modulates how attention is allocated. We found that the distributions of confidence and attention were similar throughout the course of the confidence task, suggesting at least a correlative relationship between the two measures. Lower confidence was associated with a lower standard deviation of attention, indicating that subjects distributed their attention more broadly, whereas higher confidence was associated with a higher standard deviation of attention, indicating a more focused distribution of attention. Our regression analysis built on this result, showing that confidence was a reliable predictor of the breadth of attention, even accounting for other task-related variables such as feature value, reaction time, and trial order.

We tested two hypotheses regarding how confidence might affect the distribution of attention. According to the first theory, which built off the Pearce and Hall and

Mackintosh models of attention, confidence should modulate the balance between attending to low-information versus high-value features. According to the second hypothesis, confidence acts as a more general modulator of the breadth of attention, adjusting the "greediness" of the allocation of attention. Neither of these models improved upon the value-based attention model on which they were based. In the case of the Biased Value and Information model, we suspect that the way we applied our information versus value criterion is responsible for the poor performance of the model, rather than anything explicitly related to confidence.

In the case of the Confidence-Modulated Gain model, the issue is more complex. The comparatively worse performance of the Confidence-Modulated Gain model is at odds with the regression results, which give a clear indication that confidence affects the breadth of attention. This discrepancy suggests that the original value-based model is failing to capture some important aspect of how subjects attend to different features. In particular, we make note of the finding that attention on the previous trial is a highly reliable predictor of attention on the current trial. This constancy of attention, which can also be inferred from our information metric, could be indicative of subjects using a hypothesis-testing strategy. If this is the case, it helps to explain the unexpectedly poor performance of both of our models and suggests several intriguing possibilities for future studies, as well as offering a new lens with which to examine the success of past models.

5.2 Future Directions

Our findings suggest several avenues of future exploration relating to how confidence impacts the distribution of attention at learning. To address the deficiencies of the Biased Value and Information model, we would be interested to go back and incorporate our better understanding of the role of information in attention allocation.

The coincidence of high-value and high-information features is problematic for our current model, which applies to all features the same heuristic for determining whether high value or low information should matter more in determining attention. At lower levels of confidence, this even application risks artificially dampening the contribution of the highest valued feature, which, not coincidentally, is the feature the subject is most likely to be testing as a hypothesis and as a result is most likely to be attending to.

This suggests two possible remedies for our model, either of which could be consistent with how subjects actually use confidence judgments to impact their attention allocation. First, it is possible that for the most highly valued features, subjects are not sensitive to confidence when considering how to weight the importance of that feature, and assign it a weight only on the basis of value. Thus, we should include the information score in the calculation only for low-valued features. This would presumably give a bump to only those features that are low-valued because their relationship to reward has not yet been explored, while having little impact on features whose link to reward has been proven unfruitful.

Alternatively, hearkening back to the idea of confidence bounds as a criteria of when to explore and when to exploit (Auer, 2003), it is possible subjects implement something more similar to a confidence threshold when determining the balance between biasing value and information. Below a certain confidence level, subjects may attend to features both according to their value and the level of information, helping them gain knowledge in a reward-sensitive manner. Above that confidence threshold, when subjects are more sure of their target and are trying to exploit their knowledge to maximize reward, they would attend only to value. Both of these alternative approaches could easily be tested through computational modeling.

Another direction to explore would be the role of hypothesis-testing in the allocation of attention during learning. The interaction between value and

perseverance of attention could have important implications for the learning process. An interesting related possibility is that confidence affects the breadth of attention by biasing the number of different hypotheses subjects attend to. Future work could incorporate these possibilities into models and test their ability to predict attention during the Dimensions Task.

The fact that our modeling results were so different for predicting attention across the three dimensions of the task versus predicting attention to features within the chosen stimulus is also a possible direction for further study. Based on these findings, it is worth exploring the impact attention to features outside of the chosen stimulus has on learning. While earlier models of choice behavior indicate that subjects employ reinforcement learning rather than Bayesian-optimal methods to the Dimensions Task (Niv et al., 2015), it would be interesting to test whether any counterfactual learning occurs specifically for those well-attended features that are outside of the chosen stimulus. This could be an extension of past findings that indicate that the amount of learning for different features within the chosen stimulus is biased by attention (Leong et al., 2017).

5.3 Limitations of Our Design

The greatest limitation in our ability to make claims about the way confidence impacts the modulation of attention was our method of acquiring confidence scores. Specifically, because we relied on a self-report taken once every three trials, we risked missing important shifts in confidence that occurred in between those trials. While the general pattern of confidence is likely to be captured by our data, we expect that the interpolation measure leads to smoother fluctuations in confidence than is actually the case. For instance, a sharp downturn in confidence could occur over the course of one trial, and then only gradually shift over the next two trials. As

we only measure confidence after the third trial, and then interpolate between that and the former confidence measure, the rate of descent we record would not be an accurate description of the true process. The discrepancies between the interpolated and actual confidence scores could be enough to affect the accuracy of any predictions the models make regarding the allocation of attention.

As such, it is worth considering other methods of recording confidence scores. The original motivation behind collecting confidence scores only once every three trials was that a more frequent report would disrupt subjects' ability to play the game by interrupting their attention and memory. Given the importance of having linked trial-by-trial measures of confidence and attention, it might be worth testing how much of an impact a more frequent confidence probe would have. Alternatively, we could consider measuring confidence not through self-report, but by analyzing changes in pupil size. Pupil size – which is already measured by our eye-tracking system– is generally linked to state of arousal; however, in a 2017 study, Urai et al. found that in perceptual decision making tasks, change in pupil diameter was related to multiple signatures of decision confidence. While this study evaluated perceptual confidence rather than confidence about value relationships, it is still possible that changes in pupil diameter could be used as a measure of confidence that does not alter subjects' performance in the dimensions task. Further studies would be needed to analyze the feasibility of this approach, as well as to confirm the relationship between confidence as analyzed by pupil diameter and confidence as reported by subjects.

Another limitation relates to our method of inferring attention from eye-tracking data. Our methods assume that subjects are attending to features at whichever location they are looking, and assign importance to that attention. However, especially when confidence is high, there is a possibility that eye gaze will not be as tightly correlated to top-down, deliberate attention processes and will instead reflect more random drifting and fixating. While subjects are learning, we expect that they

will attend to those features they consider most important for determining future reward. On the other hand, once subjects are highly confident in their knowledge of the target feature, attention is only necessary to find that feature during the pre-choice part of the trial. After selecting a stimulus, the subject can afford to allow their attention to wander, with no particular learning goal in mind. Assuming the subject is rewarded, it is likely attention will appear similarly undirected during the outcome section of the trial, as perfect prediction means attention isn't necessary to update values. This "Goldilocks effect" of attention has been demonstrated in past studies when a task is too easy or far too complex (Kidd et al., 2012). Because subjects do not perfectly learn the target feature on most games, we do not expect this effect to be significant, but it could be a potential confound.

5.4 Conclusion

In complex, high-dimensional environments, selective attention is necessary to narrow the scope of information to a manageable level. Selective attention helps to relieve cognitive processing constraints, in part by reducing the dimensionality of the task to levels where simpler algorithms can be used both efficiently and effectively. In order to best make use of selective attention to learn about the world, it is necessary to learn where in the world attention should be directed. Our results strongly suggest that there is an interaction between confidence and attention during learning, with confidence modulating how people allocate their attention across different features of a task. Future studies should be directed to uncovering the exact nature of that relationship, as well as elucidating the role of hypothesis-testing in the attention process.

Appendix A

Honor Code

This paper represents my own work in accordance with University regulations.

References

[1] Auer, P. (2003). Using confidence bounds for exploitation-exploration trade-offs. *J. Mach. Learn. Res.*, 3, 397-422.

[2] Baluch, F., & Itti, L. (2011). Mechanisms of top-down attention. *Trends Neurosci*, 34(4), 210-224.

[3] Bellman, R., & Rand Corporation. (1957). *Dynamic Programming*. Princeton, NJ: Princeton University Press.

[4] Borji, A., Sihite, DN., & Itti, L. (2013). What stands out in a scene? A study of human explicit saliency judgment. *Vision Res*, 91, 62-77.

[5] Buschman, T., & Kastner, S. (2015). From Behavior to Neural Dynamics: An Integrated Theory of Attention. *Neuron*, 88(1), 127-144.

[6] Chun, M., & Turk-Browne N. (2007). Interactions between attention and memory. *Curr Opin Neurobio*, 17(2), 177-184.

[7] Cowan, N., & Wood, N. (1997). Constraints on Awareness, Attention, Processing, and Memory: Some Recent Investigations with Ignored Speech. *Conscious Cogn*, 6(2-3), 182-203.

[8] Dayan, P., Kakade, S., & Montague, P. R. (2000). Learning and selective attention. *Nat Neurosci*, 3 Suppl, 1218-1223.

[9] Dayan, P., & Niv, Y. (2008). Reinforcement learning: the good, the bad and the ugly. *Curr Opin Neurobiol*, 18(2), 185-196.

[10] Duc, AH., Bays, P., & Husain, M. (2008). Eye movements as a probe of attention. *Prog Brain Res*, 171, 403-411.

[11] De Martino, B., Fleming, S.M., Garrett, N., & Dolan, R.J. (2013). Confidence in value-based choice. *Nat Neurosci*, 16(1), 105-110.

[12] Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu Rev Neurosci*, 18, 193-222.

[13] Dopson, J.C., Williams, N.A., Esber, G.R., & Pearce, J.M. (2010). Stimuli that signal the absence of reinforcement are paid more attention than are irrelevant stimuli. *Learn Behav*, 38(4), 337-347.

[14] Feng, S. F., Schwemmer, M., Gershman, S.J., & Cohen, J.D. (2014). Multitasking versus multiplexing: Toward a normative account of limitations in the simultaneous execution of control-demanding behaviors. *Cogn Affect Behav Neurosci*, 14(1), 129-146.

[15] Gershman, S.J., & Niv, Y. (2010). Learning latent structure: carving nature at its joints. *Curr Opin Neurobiol*, 20(2), 251-256.

[16] Haselgrove, M., Esber, G.R., Pearce, J.M., & Jones, P.M. (2010). Two kinds of attention in Pavlovian conditioning: evidence for a hybrid model of learning. *J Exp Psychol Anim Behav Process*, 36(4), 456-470.

[17] Jones, M., & Canas, F. (2010). Integrating reinforcement learning with models of representation learning. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society: Cognitive Science Society*.

[18] Kepecs, A., Uchida, N., Zariwala, H.A., & Mainen, Z.F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature*, 455(7210), 227-231.

[19] Kidd, C., Piantadosi, S. T., & Aslin, R.N. (2012). The Goldilocks effect: human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS One*, 7(5), e36399.

[20] Leong, Y.C., Radulescu, A., Daniel, R., DeWoskin, V., & Niv, Y. (2017). Dynamic Interaction between Reinforcement Learning and Attention in Multidimensional Environments. *Neuron*, 93(2), 451-463.

[21] Lewinsohn, S., & Mano, H. (1993). Multiattribute Choice and Affect - the Influence of Naturally-Occurring and Manipulated Moods on Choice Processes. *J Behav Decis Mak*, 6(1), 33-51.

[22] McCallum, A.K. (1996). *Reinforcement learning with selective perception and hidden state*. The University of Rochester.

[23] Mackintosh, N.J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychol Rev*, 82(4), 276-298.

[24] Montague, P.R., Dayan, P., & Sejnowski, T.J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J Neurosci*, 16(5), 1936-1947.

[25] Niv, Y., Daniel, R., Geana, A., Gershman, S.J., Leong, Y.C., Radulescu, A., & Wilson, R.C. (2015). Reinforcement Learning in Multidimensional Environments Relies on Attention Mechanisms. *J Neurosci*, 35(21), 8145.

[26] Niv, Y., & Schoenbaum, G. (2008). Dialogues on prediction errors. *Trends Cogn Sci*, 12(7), 265-272.

[27] Nosofsky, R.M., Gluck, M.A., Palmeri, T.J., McKinley, S.C., & Glauthier, P. (1994). Comparing models of rule-based classification learning: a replication and extension of Shepard, Hovland, and Jenkins (1961). *Mem Cognit*, 22(3), 352-369.

[28] O'Doherty, J.P. (2012). Beyond simple reinforcement learning: the computational neurobiology of reward-learning and valuation. *Eur J Neurosci*, 35(7), 987-990.

[29] Pearce, J.M., & Hall, G. (1980). A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol Rev*, 87(6), 532-552.

[30] Rehder, B., & Hoffman, A.B. (2005). Eyetracking and selective attention in category learning. *Cogn Psychol*, 51(1), 1-41.

[31] Roelfsema, P.R., & Ooyen, A.V. (2005). Attention-Gated Reinforcement Learning of Internal Representations for Classification. *Neural Comput*, 17(10), 2176-2214.

[32] Schultz, W., Apicella, P., & Ljungberg, T. (1993). Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *J Neurosci*, 13(3), 900-913.

[33] Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, 275(5306), 1593-1599.

[34] Sutton, R. S., Barto, A. G., & MITCogNet. (1998). *Reinforcement Learning: An Introduction* (pp. xviii, 322 p. ill. 324 cm.).

[35] Tokic, M. (2010). Adaptive epsilon-Greedy Exploration in Reinforcement Learning Based on Value Differences. Ki 2010: *Advances in Artificial Intelligence*, 6359, 203-210.

[36] Uncapher, M., & Rugg, M. (2009). Selecting for memory? The influence of selective attention on the mnemonic binding of contextual information. *J Neurosci*, 29(25), 8270-8279.

[37] Urai, A.E., Braun, A., & Donner, T. H. (2017). Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. *Nat Commun*, 8, 14637.

[38] Wilson, R.C., & Niv, Y. (2011). Inferring relevance in a changing world. *Front Hum Neurosci*, 5, 189.

[39] Yi, D., & Chun, M. (2005). Attentional Modulation of Learning-Related Repetition Attenuation Effects in Parahippocampal Cortex. *J Neurosci*, 25(14), 3593-3600.