The role of Prediction Error in Probabilistic Associative Learning

Jiří Čevora

September 2017

Corpus Christi College

This dissertation is submitted for the degree of Doctor of Philosophy

This thesis is dedicated to my parents

Tato disertace je věnována mým rodičům

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

This thesis does not to exceed 60,000 words excluding bibliography, figures and appendices.

Chapter 3 is based on paper "Cevora, J., & Henson, R. N. (2017). Reconsidering the imaging evidence used to implicate prediction error as the driving force behind learning. Frontiers in Psychology, 8 , 1380."

Acknowledgements

First and foremost I would like to thank my supervisor Rik Henson for his guidance and support throughout my PhD. I was very lucky to have a supervisor who gave me as much freedom as he did while at the same time was always happy to help. I'm indebted to Matt Lambon-Ralph for shaping my thinking about science and the support that led to the beginning of this PhD. Advice and guidance of Dr. Maté Lengyel was invaluable especially during the writing of Section 2.2.

More people than could be mentioned here made my experience in Cambridge extraordinary and in many respects truly transformative. Special places among these people belong to Alex, Charlie, Joe, Reneé and Saskia.

I must express my deepest gratitude to my parents, for their continued support and encouragement without which none of this would be possible.

Finally, I would like to thank to Medical Research Council for funding my PhD.

Contents

1 Introduction **10**

1.1 Computational level 12

1.2 Algorithmic level 15

1.3 Implementational level 18

1.4 Summary 19

1.5 Overview of the thesis 20

2 Stimulus associability effects **22**

2.1 Formalisation of associative memory 25

2.1.1 Notation 25

2.1.2 Formalised experimental procedure 26

2.2 Rational model 27

2.2.1 Discussion 30

2.3 Algorithmic models 31

2.3.1 Hebbian learning 32

2.3.2 Rescorla-Wagner 33

2.3.3 Factors scaling the learning 33

2.3.4 Normalisation 36

2.4 Discussion 37

3 Neuroimaging evidence for PE **39**

3.1 The nature of PE-correlated signal in fMRI 40

3.1.1 Analysis 41

3.1.2 Simulations 42

3.1.3 Discussion 43

4 Statistical inference of subjective probability distributions 46

4.0.1 Task design 47

4.0.2 Generative model 49

4.1 Methods 51

4.1.1 Estimation of decision-making parameters 51

4.1.2 Calculating likelihood of subjective distributions 54

4.1.3 Experimental paradigm 55

4.2 Discussion 59

5 Is probabilistic associative learning driven by PE? 61

5.1 Blocking 61

5.2 Theory of learning in subjective probability space 64

5.2.1 Learning as a flow 66

5.3 Compound learning 70

5.3.1 Generalised blocking 71

5.3.2 False blocking 72

5.4 Methods 74

5.5 Results 78

5.5.1 Post-hoc analysis 80

5.6 Discussion 81

6 Discussion 84

6.1 Summary 85

6.2 Future directions 89

6.3 Conclusion 92

References 93

Appendix 98

A Derivation of likelihood of γ 98

B Decision-making parameters optimisation 101

C High and low performing participants 102

Chapter 1

Introduction

This thesis concerns associative learning in probabilistic contexts. Associative learning is a particularly exciting area of research because it potentially subsumes a number of other areas of learning. The definition of associative learning relies on the concept of associating cues with outcomes, but this can be extended to some of the contexts traditionally considered non-associative, such as habituation (Rumelhart, McClelland, Group, et al., 1988), by generalising the associative context into a non-associative one. This thesis defines associative learning very generally: as discovering contingencies in the world (or task) that can be exploited to predict future events.

For example, simple habituation to a stimulus can be framed in an associative context as associations between a single cue and multiple outcomes, while dishabituation occurs when an unexpected outcome appears. In Chapter 3, I consider a *parameter estimation* paradigm, which can also be viewed as associating a single cue with multiple outcomes, except that the participant is required to explicitly estimate the value of the parameter.

Episodic memory can be seen as a large number of strong associations; however, the learning mechanisms that operate within episodic memory are certainly different to the ones that operate in associative memory, which is generally unable to encode a complex episode from a single exposure (Rumelhart & McClelland, 1982). An important distinction between associative and episodic memory for this thesis is that associative memory is considered *ahistoric*, i.e. the learning episode is forgotten after the associative model is updated by an observation. Episodic and associative memory often interact. It is certainly the case that associative instances might be encoded as episodic memories during an associative task and then used to make decision in associative contexts, but at that point the use of these episodes is seen as an associative memory mechanism for the purpose of this thesis.

Before formalizing probabilistic associative learning, it is useful to formalize deterministic associative learning. Associative learning is measured in tasks where cues are associated with outcomes. When those associations are deterministic, we can consider the task in terms of set theory, as finding an appropriate cue-outcome map $C \to O$. From the learner's perspective, this task can be solved simply by sampling all cues once and recording which outcomes are contingent on a particular cue. This can be expressed as a $N^C \times N^O$ binary matrix, where N^C and N^O are number of cues and outcomes. This set-theoretical conceptualisation is useful not only to contrast with probabilistic associative learning, but will also be used to motivate methods in Chapter 4, where it greatly simplifies analysis of learning problems.

Probabilistic associative learning can no longer be characterised by $C \to O$ maps. Instead, a cue c is associated with all outcomes o with probability $P(o|c)$. This task is more complicated than deterministic associative learning because the learner has to estimate N^C discrete probability distributions with N^O states. Now the $N^C \times N^O$ matrix contains real values that are suitably bounded by axioms of probability. Moreover, the convergence of the estimates with the real contingencies is only guaranteed for an infinite number of samples from C.

Often, posterior probabilities over outcomes are not represented explicitly by learning theories, but instead a *weight* matrix is updated after each learning trial. Apart from being monotonic, the relationship between weights and probabilities has many forms in the literature. Importantly, weights are not bound by the axioms of probability. The theories that specify weights generally concern themselves with a quantity $\Delta w_{ij\tau} = w_{ij(\tau+1)} - w_{ij\tau}$ which defines how an element $[i, j]$ of the weight matrix W changes with exposure to new data at time τ.

David Marr's seminal *levels of analysis* (Marr & Vision, 1982) provide us with a useful framework for how to approach associative probabilistic learning from multiple perspectives. Marr's highest, *computational level*, describes the goals of the system, and conforms to *rational* or *normative* theories of learning. These theories provide descriptions of the tasks to be solved, and operate on the assumption that evolutionary pressure has pushed the neural system to operate in a mode that is close to optimal computation (Anderson, 1990). In other words, rational theories describe what the system ought to do. Most of the learning mechanisms derived from rational theories assume that learning is driven by *prediction error* [PE]. PE represents the difference between the outcome on the current trial and the outcome(s) predicted from previous trials. PE-learning generally provides performance that is closer to optimal than non-PE learning. However, PE is not a necessary consequence of approximately optimal inference, as I will argue in this thesis.

Theories pitched at Marr's *algorithmic* level of analysis attempt to describe the rules by which a system operates, i.e., the specific algorithm (of many potential ones) that achieves the system's computational goals. Learning theories at this level of analysis are informed largely by behavioural evidence, such as the use of *blocking* experiments to infer the utilization of PE in learning, as expanded below. Lastly, theories pitched at the *implementational* level describe how a system such as the brain realizes learning algorithms, subject to the constraints of the biological substrate. However, as discussed in Chapter 3, the neural evidence for PE in learning is far from established, and consistent with other non-PE rules too.

1.1 Computational level

Rational theories of learning are a relatively new approach to the theory of associative learning, first introduced in Anderson's seminal monograph *The adaptive character of thought* (1990). These theories look at the problem of interest and find the statistically optimal solution to that problem. The rational theory is then simply the optimal statistical inference procedure. The rationale behind this approach is the evolutionary pressure on organisms to maximise fitness and that the optimal statistical inference achieves this.

The rational approach has seen tremendous success during the last two decades, demonstrating how learning behaviours are close to the optimal statistical inference across many different experimental paradigms and different species (e.g. Courville, Daw, & Touretzky, 2006). In one sense, this is hardly surprising because it is essential for all organisms to appropriately react to the environment and predict its changes (Bray, 2009), and optimal statistical inference is basically a characterisation of good predictions. However, the fact that organisms deal well with their environment does not mean that they implement optimal inference. Instead, natural selection means that it is likely that some problems are solved by rather arbitrary computations that arise by accidental means, or by adapting solutions to other problems that the organism faces. Moreover, when there are computational limits of a system, these can result in less than optimal inference. Finally, when learning is measured in laboratory behavioural tasks, the computational goals are not always clear, and different participants may make different assumptions about the learning task. These considerations mean that the rational approach to learning is not always appropriate.

One example of a failure of rational models of cognition are order effects, especially *primacy* and *recency* effects (Daw, Courville, & Dayan, 2008). These effects in associative learning are analogous to those in list learning: cue-outcome pairs presented at the beginning and end of an experiment have disproportionately greater influence on participants' learning, even though the order of trials is completely irrelevant.

While there are several algorithmic models that predict primacy and recency effects (e.g. Kruschke, 2006), it is not possible for a rational model to produce these effects. Daw and colleagues (2008) attempt to solve this problem by using semi-rational models. To do this, they bound the rationality of the rational model in two important ways, each of which reveals a fundamental way in which humans diverge from optimal statistical inference.

To illustrate this, consider the *generative model* for an associative task, in which each cue is associated with a probability distribution across outcomes that is fixed across trials. This results in equal importance of all data points, and thus no serial position effects are possible. The recency effect emerges when the possibility of change in the underlying associations is introduced into the generative model (Daw et al., 2008). Undoubtedly, the possibility of change in the underlying associations is one of the essential properties of our natural environment, and this naturally leads to greater importance being placed on more recent data. However, it is definitely not a rational solution to the task in question. Moreover, the number of variations on the basic generative model that are plausible in the natural environment is virtually unbounded. Therefore I find the choice of including this assumption into the model a profound breach of the rational approach making the model no more theoretically motivated than any of the algorithmic models the authors criticise for lack of theoretical grounding (e.g. Kruschke, 2006). Moreover, this approach opposes another popular view of learning, spelled out by David Shanks as "to a first approximation, associative judgements are unbiased at asymptote" (Shanks, 1995, p. 33) , because it suggests that the learning is fundamentally biased.

Similarly to the recency effect, the primacy effect also only emerges in rational models when the rationality is bounded. It is entirely possible to derive a generative model that would produce a primacy effect, e.g. by assuming that the amount of noise in the system increases over trials. This time Daw and colleagues (2008) attempted to explain the inefficiency (primacy effect) by the need for approximation because the rational model is computationally infeasible. Exactly as in the case of their explanation of recency effect, the argument of Daw and colleagues is hard to disagree with; however, because the number of ways that can be used to approximate rational inference is unbounded and not all of them produce the recency effect, this semi-rational approach has little value in practice.

I think the reason that the additional assumptions of the semi-rational approach seem so natural is that exact statistical inference requires re-evaluation of all the data ever encountered, in order to update ones' beliefs. This becomes inefficient with large data sets, since storing each data point in memory poses significant cost to the neural system. At some point, the benefit of performing the exact inference is outweighed by its computational cost. To make the learning practical, an iterative algorithm is needed that only considers a limited number of statistics of the data previously observed (along with the new observation), to form a new posterior. The problem here is that there is an infinite number of ways for the statistical inference to be approximated.

While the need for approximation is apparent and relatively straightforward to derive once the limits of the system are known, it is extremely hard to find out what the limits are. There is a multitude of possible bounds on computation, such as memory capacity, processing power or energy requirements, none of which are yet known. Inferring the bounds on processing from the sub-optimality of learning performance is tricky, because the sub-optimality might be caused by applying the wrong generative model.

The most common approximations to optimal statistical inference take a form similar to the Rescorla-Wagner rule (e.g. Nassar, Wilson, Heasly, & Gold, 2010; Daw et al., 2008), which entails the computation of PE. As the Rescorla-Wagner rule is identical to gradient descent with a squared error cost function (Rescorla, Wagner, et al., 1972), it is therefore guaranteed to be the best iterative approximation based purely on weight matrix. If, for instance, not only on a weight matrix (current state of associations) was conserved, but also a selection of previous data points, it would be possible to arrive at even better approximations (e.g. by combining gradient descent with particle filtering: Doucet, De Freitas, & Gordon, 2001). The optimal approximation given the particular computational bounds of the system may or may not involve PE computation depending on what precisely the computational bounds are. Indeed, we know that some of the most recent datapoints can be conserved in memory as episodes, and used for belief update (Mazur & Wagner, 1982).

1.2 Algorithmic level

Analysis at the algorithmic level does not make explicit assumptions about the environment, nor computational limitations of the system; these theories merely specify the computation performed. There are three ways these models are derived: a) top-down, very much in the way algorithms are developed in computer science, b) by specifying bounds on rationality in a rational model, or c) derived post-hoc to fit the data.

While there has been a large number of various algorithmic theories of learning, the vast majority of them can be conceptualized in the connectionist framework (Rumelhart et al., 1988). The main tenet of connectionism is that associations can be represented as weights in an associative network and learning is a change to these weights. The few theories that do not lend themselves to the connectionist framework, such as exemplar learning (Shepard, 1958), are not well supported by the data (Shanks, 1995) and therefore will not be considered in this thesis.

In the early 20th century, Edward Thordike (1927) postulated that the behaviours providing favourable outcomes will become more likely. This became known as the *law of effect*. When interpreted in the connectionist framework, this postulate becomes identical to Donald

Hebb's neural doctrine, which is an implementational theory derived from Hebb's observations of synaptic plasticity. In terms of weight matrices, we can define Hebb's learning rule as:

$$\Delta w_{ij} = k a_i t_j \tag{1.1}$$

where a_i refers to activation of input unit / presence of cue i, t_j is a presence of the (target) outcome j, and k is a real-valued learning rate. We will look closer at properties of the many theories based on the principles derived by Thorndike and Hebb (Hebbian learning) in Chapter 2.

The dominance of Hebbian theories of learning was challenged in 1969 with the introduction of blocking paradigm by Leon Kamin (Kamin, 1969). Blocking is a compound conditioning paradigm in which a novel cue A is presented in a compound with cue B, together with a reward (outcome). In the experimental condition, cue B has already been conditioned to predict the reward, while in the control condition, cue B is novel too. Subsequently, reward anticipation caused by cue A is compared between the two groups, and the blocking effect refers to the finding that this anticipation is greater in the control condition than experimental condition, because cue B has "blocked" learning of cue A [1].

Hebbian theories cannot explain this effect because in Hebbian learning, the update on weights relating to one cue are entirely independent of the weights relating to the other cues. Blocking has been used as an argument in favour of learning rules that instead modulate the amount of learning by PE. PE offers a simple explanation to the blocking effect: In the experimental condition, when the compound cue is presented, there is relatively little PE because the reward is already predicted by cue B. This results in little learning and hence little subsequent anticipation for reward when cue A is presented. On the other hand, the control condition results with relatively high PE when the compound is presented, since the reward is not predicted by either cue, resulting in a greater amount of learning (to both cues) and thus more subsequent reward anticipation by cue A.

Having said this, in Chapter 2 I show that augmenting Hebbian learning with a scaling parameter (learning rate) that depends on the *informativeness* of cues can also produce a blocking effect.

[1] Though note that this difference is not always found (Maes et al., 2016).

Despite the equivocal nature of the evidence from blocking experiments, there is little doubt that the Rescorla-Wagner (Rescorla et al., 1972) rule has become the most influentional algorithm for associative learning (or conditioning) (Siegel & Allan, 1996). For our purposes, this rule can be specified as

$$\Delta w_{ij} = ka_i(t_j - \sum_{i'} w_{i'j}a_{i'}) \ , \tag{1.2}$$

where the bracketed term represents the PE, i.e, difference between target outcome and outcome predicted by current weights.

Despite being the optimal solution involving a single weight matrix from the rational perspective, the standard Rescorla-Wagner rule cannot explain a number of other findings from associative learning, which has led to a number of adjustments. One of these adjustments is stimulus *associability*, which refers to the consistency to which a cue has been associated with reward in the past. This adjustment was originally proposed by Nicholas Mackintosh (1975) to account for associative history effects. For our purposes this can be formalised by the addition of a variable α_i that represents the associability of cue i:

$$\Delta w_{ij} = \alpha_i ka_i(t_j - \sum_{i'} w_{i'j}a_{i'}) \tag{1.3}$$

where α is increased on a given trial if:

$$|t_j - a_i w_{ij}| < |t_j - \sum_{i' \in I, i \neq i'} w_{i'j}a_{i'}| \tag{1.4}$$

(i.e, when cue i predicts the outcome better than all other cues), and decreased if:

$$|t_j - a_i w_{ij}| > |t_j - \sum_{i' \in I, i \neq i'} w_{i'j}a_{i'}| \tag{1.5}$$

However, these equations increase the computational complexity of the algorithm, because N^C values for α need to be updated on every trial and stored in memory.

Subsequently, John Pearce and Geoffrey Hall attempted to explain associative history effects by an algorithm that is an interesting fusion of Rescorla-Wagner and Hebbian learning (Pearce & Hall, 1980). Their approach was essentially a Hebbian learning model, but with a variable learning rate that is defined as the absolute value of PE on the previous trial:

$$\Delta w_{ij\tau} = ka_{i\tau}a_{j\tau} \left| t_{j(\tau-1)} - \sum_{i'} w_{i'j(\tau-1)}a_{i'(\tau-1)} \right| \ . \tag{1.6}$$

This modification explains both associative history effect and blocking effects. Nonetheless, yet other findings could not be explained, leading to a combined model (Pearce & Mackintosh, 2010). However, consideration of these effects is beyond the scope of the present thesis.

1.3 Implementational level

Theories couched at the implementational level are constrained by considerations of the physical instantiation of algorithms, which for present purposes are the neural mechanisms in the human brain. There has been a substantial amount of research describing the mechanisms of synaptic plasticity, and how such synaptic processes underlie behavioural evidence of associative learning, at least in simple organisms (Kandel, 2001). The link from synaptic processes to behavioural learning in humans is less direct, mainly owing to limits on the invasive methods of measuring plasticity. However, we proceed under the minimal assumption that the cellular mechanisms in primitive animals are preserved in humans, and also underlie probabilistic associative learning.

The Hebbian doctrine of synaptic plasticity states that neurons that "fire together, wire together", as expressed formally in Equation 1.1.

There are a few problems with this definition, even before we consider behaviour of neural systems. First, w_{ij} is not bounded; second, there is no mechanism for w_{ij} to decrease. These two issues have been addressed by Oja (1982) by the simple addition of a decay term:

$$\Delta w_{ij} = k a_i a_j - d w_{ij} \tag{1.7}$$

As more biological detail was discovered over the years, new theories of synaptic plasticity were derived. In respect to the topic of this thesis - the role of PE in learning - virtually all of the biologically inspired theories are Hebbian, i.e. do not ascribe any role to PE at the level of single synapses. One of the most influential contemporary models is named BCM after its authors Bienstock, Cooper and Munro (1982). Interpreted in the connectionist framework, BCM is essentially Oja's rule with a special postsynaptic activation function that depends not only on the current presynaptic activation, but also time-averaged presynaptic activity:

$$\Delta w_{ij} = \phi(a_i, \bar{a_i})a_j - dw_ij \ , \tag{1.8}$$

where

$$\phi(a, \bar{a}) = a(a - \bar{a}). \tag{1.9}$$

While virtually all biologically-inspired models of synaptic plasticity are Hebbian in their nature, it has been demonstrated that a proportion of neurons in ventral midbrain compute PE (e.g. Schultz, Dayan, & Montague, 1997). While natural selection implies eventual loss of features that do not increase fitness, it is possible that PE computation improves fitness in some way other than guiding associative learning. In other words, those neurons may not necessarily contribute to learning. It is challenging to link the activity of these neurons to learning in a way analogous to the work that linked synaptic plasticity to behavioural change (as done by Kandel, 2001) because these neurons have not been found in lower species, and finding analogous evidence in higher species is more difficult due to increased dimensionality of cortical representations and practical considerations for measuring learning. To counter this problem, a number of researchers resorted to the use of non-invasive neuroimaging techniques in humans (e.g. Gläscher, Daw, Dayan, & O'Doherty, 2010; Nassar et al., 2010). This approach however suffers from a number of problems, such as pooling over large populations of neurons, and alternative metabolic contributions, as discussed in depth in Chapter 3.

In conclusion, evidence at the neural implementational level does not sufficiently support the notion that PE is the driving force behind learning; nonetheless, learning is currently the only good explanation for the existence of neurons that signal PE.

1.4 Summary

The vast majority of the development in the field of associative learning is focused around PE-based theories. The main drivers behind the popularity of these theories are the empirical blocking effect (Kamin, 1969) and rational analyses of learning (Anderson, 1990). However, the classic blocking experiments of Kamin (1969) are not always reproducible (Maes et al., 2016), and in Chapter 2, I introduce an alternative explanation for the blocking effect that does not rely on PE. Moreover, while rational models of learning are often seen as an example of inductive reasoning (Gelman & Shalizi, 2013), thereby offering a less biased view of learning, without specifying bounds of rationality, these rational theories are not testable. This renders the rational approach hypothetico-deductive rather than inductive, exactly as the theories at the algorithmic level. On the other hand, building a learning model at the implementational level, i.e. based on the descriptions of synaptic plasticity (e.g. Hebb, 1952; Bienenstock et al., 1982; Toyoizumi, Kaneko, Stryker, & Miller, 2014), could be considered an inductive approach, because it takes the biological properties of neurons and assumes only that their consequences are reflected at higher levels of description. However, this approach results in learning theories that are essentially modifications of Tolman's law of effect (1932), without necessitating a role of PE in learning.

In conclusion, there is a top-down argument for the role of PE in learning, which relies on the assumption that the computational limitations of the neural system are in a regime that favours approximations to statistical inference based on gradient descent (PE). The behavioural evidence supporting the role of PE in learning has been recently found to be less robust than originally thought (Maes et al., 2016). On the other hand, there is a good bottom-up argument for Hebbian learning, as its origins are in the description of synaptic plasticity. This thesis aims to find whether the principles implicated by the implementational or the computational level are reflected at the algorithmic level.

1.5 Overview of the thesis

In Chapter 2, I look at the ability of of various algorithmic learning theories to account for associative history effects as defined by (Mackintosh, 1975). As to my knowledge there is no rational theory that could account for this effect, Chapter 2 includes its derivation. The rational theory was further used to derive an algorithmic approximation that explains the associative history effect of Mackintosh, while being better theoretically motivated and less computationally complex. This algorithm is essentially Hebbian learning scaled by the relative informativeness of a cue.

In Chapter 3, I reconsider the neuroimaging evidence used to implicate the role of PE in associative learning. While the fact that PE is computed in the brain is well established (e.g. Schultz et al., 1997) the link between the neural PE signal and learning has not been well established. By means of analytical proof I demonstrate that parameter estimation tasks can not be used to distinguish between PE and non-PE learning. This proof is extended by numerical methods to the general associative learning context.

In Chapter 4 I identify the lack of direct observability of the subjective probability distributions as the main barrier to distinguishing PE and non-PE learning theories. This chapter includes an experimental paradigm and accompanying statistical methods that allow for inference of subjective probability distributions in participants.

These methods are utilised in Chapter 5 on a large online dataset collected to investigate whether associative learning is driven by PE. The results of this experiment strongly suggest that PE does not have the role it has been ascribed by the Rescorla-Wagner theory. In contrast, the algorithm based on relative informativeness derived in Chapter 2 provided significantly better fits to the data.

Chapter 2

Stimulus associability effects

There are many possible algorithms that can achieve reasonably good learning (e.g. WIDROW & HOFF, 1960; Hebb, 1952). However, these algorithms tend to have many degrees of freedom, and it is therefore possible to find an algorithm simulating almost any data. The best way to dissociate between them is to look for situations when they fail to account for the data. There is a group of behavioural effects that can be used for this purpose. One of them is associative history, which is an effect of previous learning on current learning (LePelley & McLaren, 2004). One specific associative history effect is stimulus associability, as originally described by Mackintosh (1975).

Here I offer a detailed analysis of the mechanisms that can give rise to an effect observed in one paradigm from our lab (Greve, Cooper, Anderson, & Henson, 2014). That paper described a number of behavioural experiments purported to show that one-shot human associative learning is driven by prediction error. Experiment 2 was the only paradigm that explicitly manipulated associability of cues, through varying the *consistency* of the $C \to O$ mapping.

The first phase of that experiment - the *training* phase – varied the consistency of associations between cues and outcomes. An example of cues with different consistency of associations after training can be seen in Figure 2.1. Learning during this training phase was not measured directly, though was inferred indirectly from the speed-up in reaction times that was found for consistent but not inconsistent cues. During the second phase of the experiment - the *study* phase - each cue was paired with a completely new (*unseen*) out-

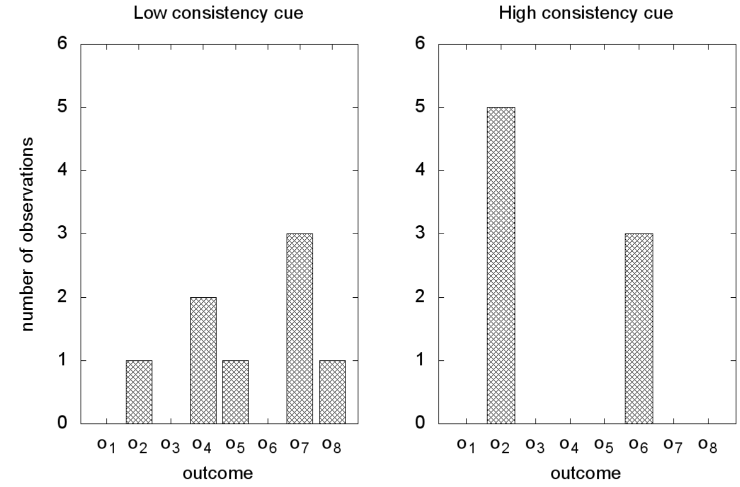


Figure 2.1: Two different example observations of cue-outcome pairing. Note that both examples entail the same number of observations, but they have different consistency of associations.

come. Lastly, in the final *test* phase, a three-alternative-forced-choice [3AFC] tested memory for which outcome had been paired with a cue in the study phase. The two other 3AFC choices were also from the study phase, but had been paired with different cues. The crucial finding of this experiment was that accuracy on 3AFC was higher for consistently paired cues than for inconsistently paired cues. The authors explained this result in terms of supposedly greater PE when consistent versus inconsistent cues were paired with new outcomes in the study phase, causing better learning of those new associations. Here I investigate these claims and explore an explanation of the observed effect that consistent cues have higher associability.

I analyse both rational and algorithmic models of the effect of cue consistency, but ultimately I aim to demonstrate how those two approaches can complement each other, in a way where the rational model explains the main principles to be used in an algorithmic model. The algorithmic model might then in turn explain the instances where human performance departs from rationality. A good algorithmic model can be therefore seen as an approximation to the ideal solution described by a rational model. This is echoed in the more mathematical literature, where it is well known that neural networks are often an approximation of statistically optimal inference (for formal proofs see White, 1989; Ruck, Rogers, Kabrisky, Oxley, & Suter, 1990). Still, this complementary approach is less common in cognitive science, as evident in the passionate debate about the superiority of rational versus algorithmic approaches (e.g. Jones & Love, 2011).

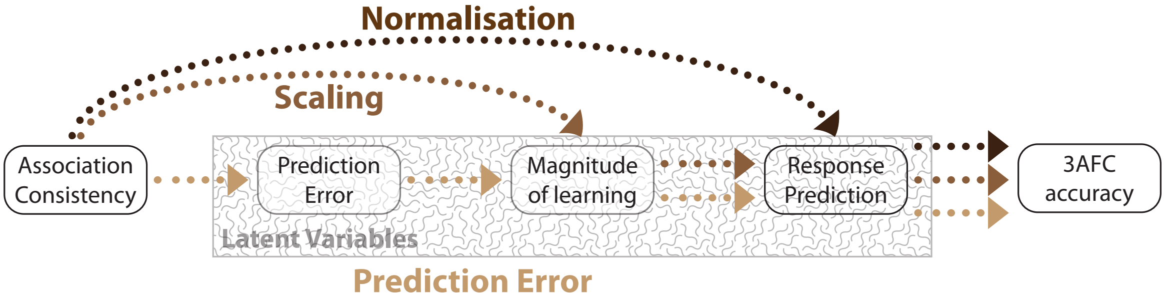


Figure 2.2: The three different causal models of the behavioural effect observed by Greve and colleagues (2017). Note that the rational model is based on the same causal theory as normalisation.

The rational model I propose here identifies an association consistency as important for predicting the posterior probability distribution across outcomes given a cue. The main idea is that when a cue has low consistency, the posterior probability across outcomes is even more uniform than their respective frequency in the data. In Section 2.2, I show how consistency can be determined from the data and rationally used to compute the posterior probability across outcomes.

Next, I go on to prove that neither a simple Hebbian algorithm nor a basic Rescorla-Wagner algorithm can implement this rational approach, produce this pattern of results and therefore cannot explain the results of Greve at al. (2014). I then consider Mackintosh's (1975) modification of learning in Section 2.3.3, in which weight updates are scaled by a measure of cue consistency, formalized as associability (α) for each cue, and extend this idea with a simpler and more tractable scaling factor that is determined by the informativeness of the weights associated with each cue. Finally I consider a second algorithm that can equally explain cue consistency effects, via normalisation of response selection. (These different algorithms are illustrated in Figure 2.2).

The original account of Greve at al. (2014) assumes that the predictions are affected directly by associative history, and errors in these predictions (PE) drive the weight updates. The associability theory assumes that associative history affects the magnitude of learning, without necessarily computing PE. Lastly, the normalisation account assumes that cue consistency affects the selection of responses from outcome predictions, without necessarily scaling learning per se, which has an identical causal structure to the rational model. It is difficult to distinguish between these algorithms since the amount of PE, the magnitude of learning and the outcome predictions are all latent variables. We can however investigate their internal consistency and inherent limitations.

All of the algorithms are formalised into computational models that can be used for rigorous mathematical analysis. All the analytical arguments presented are supported by numerical simulations, which were qualitatively compared to the results of the behavioural experiment (see Figure 2.5; Greve et al., 2014).

2.1 Formalisation of associative memory

2.1.1 Notation

For the purpose of rigorous analysis, it is necessary to fully define the system of interest. In the experiment described here (Greve et al., 2014), the associations learned were between scenes and faces; however for generality we will refer to these sets as cues (scenes) and outcomes (faces). Because there was no systematic relationship (e.g. similarity) within the sets of cues and outcomes, we will treat them as discrete variables (or orthogonal representations for the purpose of a neural network). The participants were exposed to trials of cue-outcome pairings, and performed an incidental task on the outcome (decide whether the face was male or female). They were not told to intentionally learn the pairings, but being able to predict the outcome (which occurred shortly after the cue) would help their task of responding as quickly as possible. However it is impractical to keep in memory all of the instances, so in our algorithmic models, we assume that knowledge is integrated into a structure summarising associations between cues. This is necessary because at some point, the cost of storing extra items in memory will exceed the benefit of better prediction (e.g. Simon, 1972).

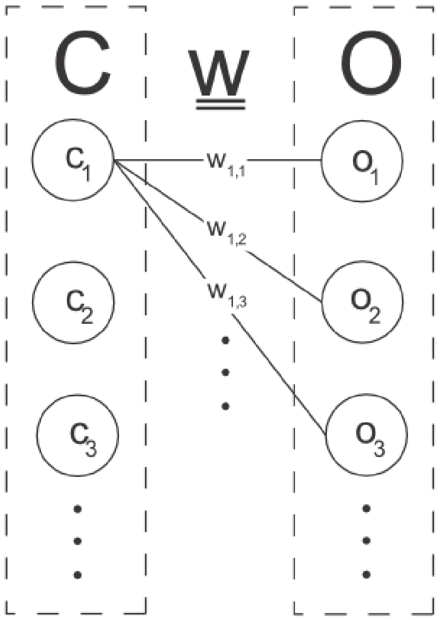


Figure 2.3: Graphical model of associative memory. Knowledge about cue-outcome associations is stored as a weight matrix $\underline{\underline{w}}$.

We therefore conceptualise associative memory in terms of a graphical model (see Figure 2.3). Each c_i node corresponds to one level of cue C, each outcome o_j to one level of O. For the purpose of modelling, states of variables will be encoded as vectors[1] $\underline{O}$ and $\underline{C}$ specifying states of the individual nodes. The weight matrix $\underline{\underline{w}}$ specifies the association between C and O nodes. Since cues are mutually exclusive and there is no uncertainty in cue identification, the probability of outcomes is fully defined by the weight vector corresponding to the currently observed cue.

2.1.2 Formalised experimental procedure

Using the notation we introduced, Greve's experiment (2014) is illustrated in figure 2.4, and formalised as follows: During the *training* phase, the cues are divided into three subsets with different consistency of $C \rightarrow O$ associations. Items from a *consistent* subset were shown three times to participants, each time paired with the appropriate cue. A *baseline* subset was shown only once, therefore there was no information about the consistency of its associations.

[1] A single underline is used to denote a vector, while a double one denotes a matrix.

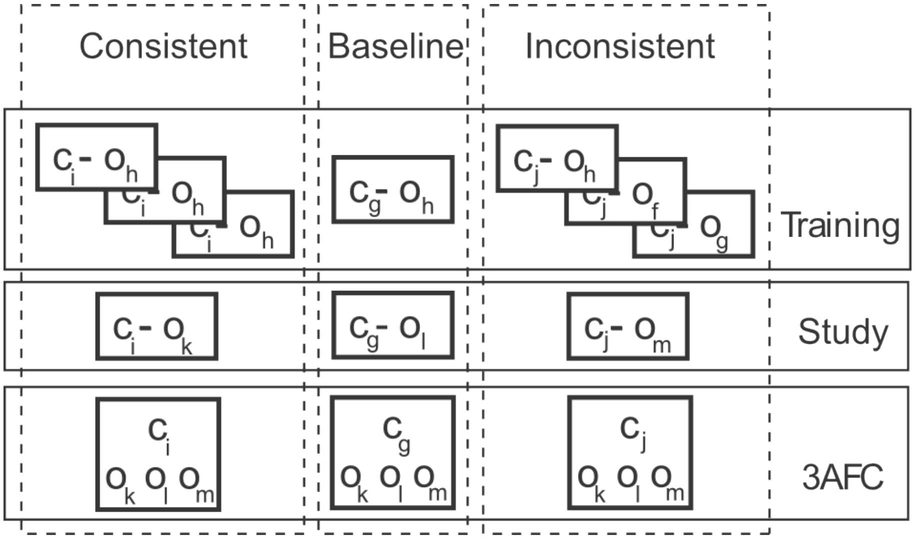


Figure 2.4: Schematic of design of Experiment 2 in Greve et al. (2017) that manipulated associative consistency.

An *inconsistent* subset was shown three times, but a different outcome was shown with each presentation of the same cue. Presentations of cues from different subsets were intermixed. As described earlier, the effect of cue consistency was measured by re-presenting each cue with a completely new outcome (in the study phase) and then later testing memory for this association using 3AFC.

Greve et al.'s results are shown in Figure 2.5, where memory for the new associations was best for consistent cues and worst for inconsistent cues.

2.2 Rational model

We assume that during learning, a rational agent should minimize surprise derived from observing each cue-outcome pairing. Surprise can be defined as the negative log probability of the observed outcome (Shannon & Weaver, 1949). A rational agent should calculate the expected probability distribution for $O_{\tau+1}$, given the cue $C_{\tau+1}$ just seen and all past cue-outcome pairings $\{C_{1:\tau}\}, \{O_{1:\tau}\}$, where τ indexes trial number (pairing). Therefore the

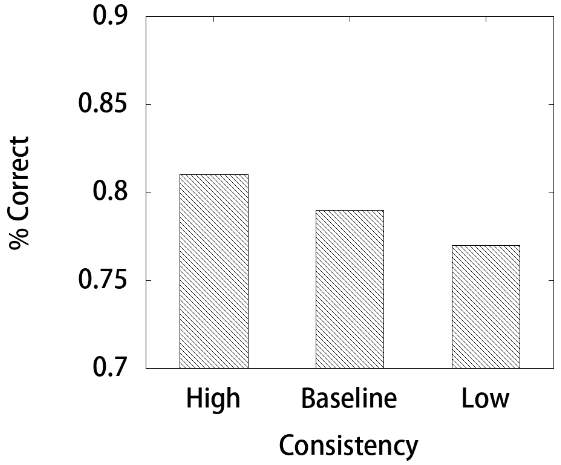


Figure 2.5: Pattern of results on 3AFC after manipulation of association consistency.

surprise in our situation can be defined as:

$$-log_2 P(O_{\tau+1}|C_{\tau+1}, \{C_{1:\tau}\}, \{O_{1:\tau}\}) \tag{2.1}$$

If we have no useful priors $P(C)$ and $P(O)$, the posterior probability is equivalent to a normalised likelihood. Nonetheless, consistency of associations is useful information for future learning. When consistency is high, the agent should consider the stimulus informative. For the case shown in Figure 2.1, for example, when a participant observes a number of inconsistent (relatively stochastic) pairings, then the surprise from seeing a yet unseen stimulus (e.g. o_3) should be lower than after observing a pairings with high consistency (relatively deterministic).

However, to be able to compute the posterior probability across outcomes given an observed cue, we need to obtain an estimate of consistency first. In fact, for each cue we will estimate a parameter, γ, called concentration, which is inversely proportional to cue consistency. If we have no prior knowledge about γ, we can estimate its most likely value $\hat{\gamma}$ by simple maximisation of its likelihood[2].

$$L(\{O_\tau\}, \gamma|\{C_\tau\}) = P(\{O_\tau\}|\{C_\tau\}, \gamma) = \frac{1}{\beta(\gamma)} \prod_{j=1}^{N^O} \beta(\underline{\gamma} + \underline{n}^{(j)}) \tag{2.2}$$

[2] Full derivation of this formula can be found in appendix A.

where $n^{(j)}$ refers to how many times outcome o_j was observed, N^O is the number of outcomes and β is the beta function.

The estimated value $\hat{\gamma}$ may be then used to define a prior distribution for associations corresponding to a cue. This is identical to assuming that for each cue, the association with outcomes is defined by a discrete probability distribution drawn from a Dirichlet distribution. The Dirichlet distribution is usually parametrised by a vector $\underline{A} = \{\gamma_{o_1}, \gamma_{o_2}, ..., \gamma_{o_N}\}$. However, since we assume no general bias in associations, we can use a simplified symmetrical Dirichlet distribution, specified by one hyperparameter γ, in which case $\underline{A} = \{\gamma, \gamma, ..., \gamma\}$. Using the definition of the Dirichlet probability density function (provided in appendix A, Equation 9), it is apparent that the posterior across outcomes can be defined by the number of times the current cue was paired with each outcome ($\underline{n}^{(C_{\tau+1})}$) and by the hyperparameter γ.

$$P(O_{\tau+1}|C_{\tau+1}, \{\underline{C}_{t=1:\tau}\}, \{\underline{O}_{1:\tau}\}, \underline{A}) = Dirichlet(\underline{A} + \underline{n}^{C_{\tau+1}}) \tag{2.3}$$

Using this formula with a particular set of observations will always result in the same $\hat{\gamma}$, because γ is a property of the dataset. But for the sake of illustration, we can consider how the posterior probability across outcomes changes with $\hat{\gamma}$ (despite that they are not separable). The effect of this procedure is best illustrated on yet unseen outcomes ($O_{\tau+1} \notin K_+$, where K_+ means already seen outcomes and K means all outcomes), for which we can define the posterior by:

$$P(O_{\tau+1} \notin K_+|C_{\tau+1} = j, \hat{\gamma}) = \frac{(K - K_+)\ P(O_{\tau+1} = g|n^{(g)} = 0, \hat{\gamma})}{(K)\ P(O_{\tau+1} \neq g|n^{(j)}, \hat{\gamma})}\ . \tag{2.4}$$

The numerator in this equation refers to the probability of an unseen outcome g multiplied by the total number of unseen outcomes ($K - K_+$), while the denominator refers to the probability of any outcome that has already been observed with a particular cue. Note that $n^{(g)}$ refers to the number of observations of outcome g while $n^{(j)}$ refers to cue j. Substituting Equation 2.3 and integrating out the nuisance variables, we get:

$$P(O_{\tau+1} \notin K_+|C_{\tau+1} = j, \hat{\gamma}) = \frac{(K - K_+)\hat{\gamma}}{K\hat{\gamma} + n^{(j)}}\ . \tag{2.5}$$

while for the seen outcomes it is:

$$P(O_{\tau+1} \in K_+|C_{\tau+1} = j, \hat{\gamma}) = \frac{n^{(j)} + \hat{\gamma}}{K\hat{\gamma} + \bar{n}^{(j)}}\ , \tag{2.6}$$

where

$$\bar{n}^{(j)} = \sum_{i=1}^{N_O} n_i^{(j)} \; . \tag{2.7}$$

To see the effect of cue consistency, we can examine the limits (see Table 2.1) of these functions, since these functions are clearly monotonic with respect to $\hat{\gamma}$. These limits indicate how a rational agent would exhibit the same pattern of behaviour as was observed in the experiment of Greve and colleagues: For highly consistent cues (in the limit), the probability of outcomes will approach their relative frequency, while for inconsistent cues, all outcomes will approach equiprobability. In the test phase of the experiment of Greve and colleagues, the outcome from the study phase (seen once) is tested against two outcomes as yet unseen with this cue. Since the relative frequency of unseen outcomes is 0, it is easy to see why the correct recall in the consistent case is relatively higher than in the inconsistent case, where the probabilities for seen and unseen outcomes are closer to equal. This behaviour was confirmed when this rational model was computationally simulated.

	γ	unseen	seen
inconsistent	$\hat{\gamma} \to \infty$	$\frac{K-K_+}{K}$	$\frac{1}{K_+}$
consistent	$\hat{\gamma} \to 0$	0	$\frac{n_i^{(j)}}{\bar{n}^{(j)}}$

Table 2.1: Posterior likelihood of seen or unseen outcome in the limits of γ for a cue j and outcome i.

2.2.1 Discussion

The rational model provided here defines the statistically optimal solution to Experiment 2 of Greve and colleagues (2014), which showed that learning of new associations was better for cues that had a more consistent pairing in the past. This model produces the same pattern of results as was observed in the experiment, which suggests that people indeed infer the consistency of stimuli when completing this task. However, this does not mean that the behavioural effect is the result of statistically optimal procedure as described here. There are other processes which might produce the same pattern of results, despite not being the optimal mechanism. These algorithmic models are the topic of the rest of this chapter.

This rational model assumes that the distributions across outcomes are independent among cues. While this assumption is likely false as confusion of cues can happen this effect bears no relevance to present analysis as cue confusion was not systematically manipulated in the experiment (Greve et al., 2017).

The rational model requires storage of all instances of learning in memory, but this framework can be easily adapted to step-by-step Bayesian updates to pose a realistic constraint on memory. Still, the example provided here has severe limitations in physiological interpretation. Maximisation of the function given by Equation 19 in appendix A requires keeping an extra statistic about each cue, and is not a simple process, but requires advanced computational capabilities, since searching for the maximum likelihood value is demanding. This kind of computation in its exact form is generally impossible in neural systems. However, there might be good approximations to this process which are possible in a neural system.

2.3 Algorithmic models

I shall briefly define the basic properties of Artificial Neural Networks (ANNs) used in the following analysis. The architecture of an ANN follows the graphical model devised earlier (Figure 2.3), but requires some additional mechanisms. The weight matrix is initialized before learning to a roughly flat distribution, by drawing each weight from a Gaussian distribution ($\mu = \frac{1}{N}, \sigma = \frac{1}{N^2}$) defined by number of possible outcomes $N = N^O$. It is necessary to distinguish between 1) the value of unit activation a_j (the posterior probability of observing outcome o_j), and 2) its target value t_j defined by the observed $C \rightarrow O$ pairing in the environment.

The activation across outcome units does not necessarily follow the properties of a probability distribution. There are a number of ways in which such a distribution can be interpreted; however, to avoid increasing the number of free parameters, a fixed transform can be used to convert the activation distribution into a proper probability distribution. Because all variables in the model are discrete and the states are mutually exclusive, the *softmax* scheme can be used (Denker & Lecun, 1991; Rumelhart et al., 1988). The 3AFC task results are then modelled as a softmax transformation of the activations across the three O units presented in each test trial.

To make the analysis simpler, I assume that learning occurs only after the exposure to a cue-outcome pairing, as a single update $\Delta w_{i,j}$ to the association between c_i and o_j . Thus for the collection of all associations, the weight matrix W is:

$$\underline{\underline{W}}_{\tau+1} = \underline{\underline{W}}_{\tau} + \Delta \underline{\underline{W}}_{\tau} \tag{2.8}$$

All of the learning rules presented below use a constant k controlling the learning rate. Unless otherwise stated, I assume k is a positive number smaller or equal to 1.

2.3.1 Hebbian learning

The Hebb rule (Hebb, 1952) captures probably the simplest idea about how learning might happen in biological systems. Here I show that various extensions of the Hebb rule - Oja's rule (Oja, 1982) and the BCM rule (Bienenstock et al., 1982) - cannot account for the experimental evidence discussed earlier in this chapter. See Section 1.2 and 1.3 for definition of these learning rules and discussion of their properties.

Let c_i and c_h be two cues with different associative histories and o_k, o_l and o_m be outcomes never seen by the participant before (as in Figure 2.4). The proof that variants of the Hebb rule cannot explain the results of Greve et al. has two parts. First, we need to show that all weights corresponding to the outcomes o_k, o_l and o_m ($\underline{w}^{(k)}$, $\underline{w}^{(l)}$ and $\underline{w}^{(m)}$) will be identical as long as these outcomes are not seen. Secondly, we need to show that when c_i is paired with o_k and c_h with o_m, the associative change is identical in both cases no matter what the associative history of c_i and c_h ($\Delta w_{ik} = \Delta w_{hm}$). This concludes the proof since it shows that o_k and o_m are identically associated to their corresponding cues at the point of the 3AFC (and likewise for o_l).

1) For Hebb's and Oja's learning rules (given by Equations 1.1 and 1.7), it is sufficient to show that a weight will not be changed unless a corresponding outcome is presented:

$$o_\tau \neq o_x \implies a_x = 0 \implies \Delta \underline{w}^{(x)}_{\tau} = 0 \tag{2.9}$$

In the case of the BCM (Equation 1.8), the ka_ia_j term will be unchanged for the same reason, however the dw_{ij} will cause a change. Nonetheless, while the individual weights corresponding to unseen cues will be changed, this will be the same for all unseen outcomes (at least on average, given randomly initiated weights).

2) If we compare the conditions during the study phase, we find that activation of both cue and outcome for each condition is identical, resulting in identical change of weights.

This means that none of the terms found in Hebb's, BCM and Oja's rules (Equations 1.1, 1.8 and 1.7) differ across the conditions. Therefore, the learning of the outcomes o_k, o_l and o_m will be identical. Thus in general, it is impossible for Hebbian learning to account for the cue consistency effect.

2.3.2 Rescorla-Wagner

The Widrow-Hoff learning rule (WIDROW & HOFF, 1960), is probably the most popular ANN implementation of the Rescorla-Wagner rule (Rescorla et al., 1972). It implements a gradient descent algorithm with squared error function:

$$E_j = \frac{1}{2}(t_j - a_j)^2 \tag{2.10}$$

and corresponding error derivative of

$$\frac{dEj}{dw_{ij}} = t_j - \sum_{i=1}^{N^O} a_i w_{ij} \tag{2.11}$$

Because the error function does not refer to weights relating to outcomes other than that presented (o_j in relation to above equations), it is apparent that this approach cannot account for Greve et al.'s findings for the same reason as the proof given for the Hebb rule in Section 2.3.1. In other words, the weights after training for unseen outcomes will be identical on average. Augmenting the Widrow-Hoff rule with a decay term (like Oja/BCM extensions of Hebb rule) will not help, for the same reasons as for Hebb rule in previous section.

2.3.3 Factors scaling the learning

In previous sections, I have shown that neither variants of the Hebb rule, nor the Widrow-Hoff rule implementation of PE-driven learning, can explain the cue consistency effects found by Greve et al. This is because these rules operate at the *local* level of individual weights, so weights for unseen outcomes are identical, regardless of the prior associative history of previously seen outcomes. However, it is possible to introduce an extra term into the learning equations that will scale the learning in a *global* manner, which if suitably defined can make the consistency of seen outcomes affect the learning of subsequent unseen outcomes. First, I investigate the original explanation of Greve et al. Secondly, I discuss an approach taken by Mackintosh (1975), identify its problems and then derive a theoretically better motivated and computationally less expensive alternative based on entropy of the weight vector relating to a particular cue.

Global PE

Greve and colleagues (Greve et al., 2014) hypothesised that the results they obtained were due to learning being scaled by global prediction error. However, it is not clear whether their results are actually consistent with this hypothesis. Their inconsistent condition [inc] can be defined by constantly changing $C \rightarrow O$, while their consistent condition [con] has $C \rightarrow O$ changed only on the last trial. Since two trials is the lowest number necessary to establish different consistency levels between conditions, we can consider just the difference between conditions at $\tau = 3$. In other words, the observed outcomes for each condition can be defined as ordered sets $O^{con} = \{1, 1, 3\}$ and $O^{inc} = \{1, 2, 3\}$.

Considering the two conditions separately in a simplified scenario where only one cue exists, Hebbian learning (Equation 1.1) can be scaled by the total absolute error E as

$$\Delta w_j = kEa_j \ , \tag{2.12}$$

where

$$E = \sum_{j' \in O} |t_{j'} - w_{j'}| \ . \tag{2.13}$$

After the first trial, the weight vectors will be identical across conditions as the conditions are identical until $\tau = 2$. During the second trial the error E and hence weight change will be larger for the inconsistent condition, where the pairing changed. More learning after $\tau = 2$ will in turn again result in greater E for the Inconsistent condition when the pairing changes (for both conditions) at $\tau = 3$, predicting the opposite pattern to that found in the data. Thus, at least for the above definition of global PE, Greve and colleagues' behavioural pattern remains unexplained (future work could examine whether this conclusion holds for the summation of higher moments of the difference between target and outcome, i.e, more convex error functions).

Associability

Mackintosh (1975) extended the Rescorla-Wagner (1972) learning theory by adding a variable learning parameter α that is defined by Equations 1.3 to 1.5. This is obviously not a useful approach for the behavioural data discussed here, since the experimental design (Greve et al., 2014) involved only one cue at a time, therefore the $\Delta\alpha_j$ would take exclusively negative values. Most importantly, the approach proposed by Mackintosh (1975) requires the organism to keep track of an extra statistic - associability - for each cue and update it on every trial.

Nonetheless, the Mackintosh's main idea provides a good starting point for derivation of a new learning rule based on the consistency of associations. As identified in Section 2.2, the best way to quantify consistency of a cue is to find the maximum likelihood hyperparameter of a Dirichlet distribution generating the observed data. The likelihood formula we have provided is impractical because it requires a record of outcome counts ($n^{(i)}$) and it does not have a closed-form solution (Minka, 2000). The lack of closed-form solution makes this task computationally demanding and unlikely to happen in the brain; however there are alternative metrics with similar properties.

Cue informativeness

The consistency of a cue for the purpose of scaling the learning can either be evaluated as an additional dynamical variable (as Mackintosh suggested) or, to avoid increasing the complexity of the model, determined from the information already encoded in our ANN. The only structure in the ANN that contains information about the consistency of cue c_i is the vector of corresponding weights $\underline{w}^{(i)}$. In information theory, a measure of information called entropy. Entropy is a measure of disorder in a system in terms of the distribution of its states (Sethna, 2006) therefore it is a useful inverse metric to quantify *informativeness* of a cue (how specific is the prediction made upon the cue, see Equation 2.14). Moreover, for a given probability distribution, entropy is a monotonic transformation of the most likely Dirichlet hyperparameter, linking well into the rational model. Because of the properties of our model (see Section 2.3), the normalized $\underline{w}^{(i)}$ vector can be used as a probability distribution with elements p_j.

$$I = \frac{-1}{\sum_j p_j log(p_j)} \tag{2.14}$$

Both Hebb and Widrow-Hoff rules can be scaled by the informativeness of the weight vector, e.g for Hebb rule:

$$\Delta w_{ij} = kIa_it_j \tag{2.15}$$

Cues with higher consistency will have higher informativeness, and therefore larger weight updates, consistent with the results of Greve et al.

Note also that some of the most influential sources of evidence for the models of associative learning are based on compound learning, where more than one cue is paired with an outcome (Kamin, 1969; Mackintosh, 1975). To account for this evidence, we can define a relative form of informativeness, where α is:

$$\alpha = \frac{I(w_\omega)}{\sum_{c \in C} I(w_c)} \tag{2.16}$$

where ω refers to a cue presented on a given trial.

In conclusion, classic learning rules scaled by a measure of consistency derived from a weight vector, such as informativeness, are able to account for the findings of Experiment 2 of Greve and colleagues (2014). Moreover, this approach is better theoretically grounded and less computationally expensive than the approach of Mackintosh (1975).

2.3.4 Normalisation

The last algorithmic model is closely tied to the rational model. A closer look at the rational model reveals that it is the distribution of expectations that is directly affected by cue consistency. However, rather than augmenting the learning rule with information about this distribution, this information can be used directly in the transformation of output unit activations into response probabilities. In other words, this information can be used to adjust the softmax function (cf. Section 2.3). The softmax function can be parameterised by a temperature parameter[3] T:

$$P(O_{\tau+1} = o_j) = \frac{exp(a_j/T)}{\sum_{k=1}^{R} exp(a_k/T)} \tag{2.17}$$

For high values ($T \to \infty$) the posterior will be almost flat, while for low values ($T \to 0$), the outcome with highest activation will approach a posterior probability of 1.

By making the temperature value inversely related to cue consistency, i.e. proportional to the entropy as defined in the previous sections, it can be seen that the Greve et al. Experiment 2 results can again be reproduced.

2.4 Discussion

This chapter analysed the theoretical implications of an experiment conducted by Greve and colleagues (2014). The experiment manipulated the consistency of cues in an associative learning task, and examined the effect on subsequent learning of the same cues paired with new, unseen outcomes. They argued that their observation of better learning of new outcomes for cues with high past consistency is consistent with the hypothesis that PE drives learning (even in one-shot, explicit memory tasks).

I derived a rational model to explain this result, and then considered various algorithms that can be used to approximate the rational model within an artificial neural network framework. More specifically, I showed how local learning rules, including those driven by PE, are not consistent with the rational model. One solution is to scale learning by a global measure of cue informativeness such as entropy, derived from the weights associated with each cue. This represents a more efficient and plausible implementation of Mackintosh's idea of cue associability (1975). Another solution is to normalise the mapping from output activations to response probabilities, making them sensitive to the same measure of informativeness.

Quantitative fitting of models to the data was not performed because this would involve another level of modelling (mapping the subjective probability of an outcome to the probability of outcome selection, as discussed in Chapter 4), which would result in the model

[3] i.e. this was set to 1 until now.

being over-parametrized, i.e, insufficient data in the accuracy levels reported by Greve and colleagues (2014) in order to distinguish different learning rules. In other words, it is not possible to quantitatively distinguish between the scaling and normalisation algorithms based on behavioural data because the magnitude of learning is a latent variable. However, it may be possible to distinguish them by simultaneous recording of brain activity: according to the scaling account, the effect of cue consistency arises during learning, i.e. during the study phase of Greve et al.'s experiment. According to the normalisation account on the other hand, the effect of cue consistency should happen during response selection, i.e, during the test phase of Greve et al.'s experiment. However, it is also possible that normalisation occurs at some point in the period between learning and test, for example as some form of weight normalisation (e.g, pruning).

This chapter has argued, at both Marr's computational and algorithmic levels, that the recent data used by Greve et al. to support PE in human associative learning is not conclusive. Indeed, I proved that neither local learning rules like the Widrow-Hoff rule, nor learning driven by global prediction error, can reproduce these data. In the next chapter, I consider the neural evidence for PE in human learning, i.e, at Marr's implementational level.

Chapter 3

Neuroimaging evidence for PE

Chapter 2 analysed stimulus associative history effects at the computational and algorithmic level, and showed that they do not provide support for learning being driven by Prediction Error [PE]. In the present chapter, I will question neural evidence implicating the role of PE in learning at the implementational level. I model associative learning in artificial neural networks using Hebbian (non-PE) learning algorithms to investigate whether the data used to implicate PE in learning can arise without actual PE computation. I conclude that the metabolic demands of synaptic change during Hebbian learning would produce a PE-correlated component in functional magnetic resonance imaging (fMRI), which suggests that the research used to imply PE in learning is currently inconclusive.

There is a considerable body of evidence that PE is computed by dopaminergic neurons in ventral midbrain. Single-cell recordings have shown neurons that are excited by unexpected reward, and depressed by unexpected lack of reward (Schultz et al., 1997). This response implies reward PE computation takes place in the brain; however, it does not imply that the PE signal is utilized during learning, and no single-cell study, to our knowledge, has demonstrated this link to learning. Furthermore, these findings have only been obtained with regard to rewarded behaviour, while the majority of learning in humans happens in absence of reward (Tolman, 1932).

These concerns can be potentially addressed in fMRI studies by relating a PE-related component of fMRI to subsequent memory, with or without overt rewards. Unfortunately, most fMRI research has focused simply on replicating the single-cell findings by identifying a correlate of PE in the human brain (e.g. McClure, Berns, & Montague, 2003; D'Ardenne, McClure, Nystrom, & Cohen, 2008; Abler, Walter, Erk, Kammerer, & Spitzer, 2006), without assessing its effect on behaviour. I am only aware of two fMRI studies that attempted to go beyond the single-cell recording findings by demonstrating an effect of PE-related component in fMRI on learning (McGuire, Nassar, Gold, & Kable, 2014; Gläscher et al., 2010). Both of these studies identify a component of the fMRI signal that is correlated with trial-by-trial estimates of PE from an assumed learning model, and then link that component to subsequent decision-making.

3.1 The nature of PE-correlated signal in fMRI

However, a PE-correlated fMRI signal does not necessarily originate from PE computation: the BOLD signal measured by fMRI may relate to metabolic changes that are only indirectly related to neural activity. One of the major factors contributing to the BOLD signal is cellular respiration associated mainly with ATP metabolism (Aubert & Costalat, 2002), which is elicited by a large number of cellular processes. Synaptic plasticity has several components working at different timescales (Collingridge, Isaac, & Wang, 2004), but there are four notable processes that operate at the timescale of these studies: a) synaptic transmission of signal, b) facilitation, which is an important form of short-term synaptic plasticity (Kandel, 2001), c) migration of receptors, which is a crucial components of long-term potentiation and depression (Collingridge et al., 2004), and d) fast forms of homeostatic activity, which serve as a form of global synaptic scaling and metaplasticity (Pérez-Otaño & Ehlers, 2005). While synaptic transmission (a) is the main energy expense during signalling (up to 55% of signalling cost, Harris, Jolivet, & Attwell, 2012), synaptic plasticity (b-d) can increase signalling efficiency up to hundred-fold (Harris et al., 2012) and therefore be expected to have a significant energy budget. Many of these synaptic processes occur rapidly (Collingridge et al., 2004), and could therefore take place within the same timewindow (resolvable by fMRI) as any neural activity related to PE. Thus while the actual energy consumption of synaptic plasticity is unknown (Harris et al., 2012), I conclude that there is a distinct possibility that it is sufficiently large to contribute to the BOLD response.

The outstanding question for this alternative explanation is why synaptic plasticity would correlate with PE, unless PE were computed and used to update synapses. In what follows, I model synaptic plasticity as the magnitude of Hebbian weight update in associative networks, and demonstrate that this quantity correlates with PE even when the learning algorithm does not compute PE.

3.1.1 Analysis

I consider a modified Hebbian learning rule that includes a weight decay term, also called Oja's rule (Equation 3.1, Oja, 1982). This learning rule does not use the current state of the network (e.g, predictions) to inform learning in any way. The only modification from the classic Hebbian algorithm is that the weights decrease linearly at each time step, which is the minimal modification necessary to obtain stable and biologically plausible learning dynamics. I contrast this variant of Hebbian learning with the Widrow-Hoff learning algorithm, also modified to include decay to increase its biological plausibility (e.g. Rumelhart et al., 1988) as shown in Equation 3.2. The formulation of theWidrow-Hoff learning rule used here is essentially Hebbian learning scaled by PE. In these equations, w_{ij} refers to the weight between unit i (representing the cue) and unit j (representing the outcome), a_i/a_j refer to the activity of unit i/j, t_j refers to a desired output of unit j, $0 \leq k < 1$ is the learning rate, $0 < d \leq 1$ is the decay rate and the H and WH superscripts refer to Hebbian or Widrow-Hoff learning rules respectively.

$$\Delta w_{ij}^{H} = -d^{H} w_{ij} + k^{H} a_i a_j \tag{3.1}$$

$$\Delta w_{ij}^{WH} = -d^{WH} w_{ij} + k^{WH} a_i (t_j - \sum_{i'} (w_{i'j} a_{i'})) \tag{3.2}$$

First, I address the relationship between learning under Hebbian and Widrow-Hoff rules in an experiment conducted by McGuire and colleagues (2014). The parameter estimation task they used is effectively associative learning with a single cue, because the participants' task was simply to predict the value of a parameter during each trial. As only one stimulus exists in this paradigm, the i subscript becomes redundant, therefore we can say that $a_j = w_j$

and both H and WH learning rules can be simplified to

$$\Delta w_j^{H'} = -d^{H'} w_j + k^{H'} a_j \tag{3.3}$$

and

$$\Delta w_j^{WH'} = -d^{WH'} w_j + k^{WH'} (t_j - w_j). \tag{3.4}$$

By equating $\Delta w_j^{H'} = \Delta w_j^{WH'}$, we can see that this statement is true whenever $k^{H'} = k^{WH'}$ and $d^{WH'} + k^{WH'} = d^{H'}$. This means that in parameter estimation tasks, learning according to the Widrow-Hoff rule can be perfectly mimicked by a Hebbian rule. Therefore, performance on this task cannot be used to argue for PE learning.

This proof cannot be extended to experiments with multiple cues, such as the one by Gläscher and colleagues (2010). I therefore turn to computational simulations to investigate whether there is a correlation between Hebbian weight update and prediction error.

3.1.2 Simulations

In computational simulations of multi-cue learning I ask whether PE correlates with weight update. Because fMRI observes entire populations of neurons, in contrast to single-cell recordings, we need to specify the variables of interest at the population level too.

I only consider the magnitude of the population weight change, $|\Delta W^H|$, because the fast decreases in synaptic strength are likely to require a similar amount of ATP as increases (Kandel, 2001; Collingridge et al., 2004) thus producing the same BOLD signal. Therefore the change associated with trial τ is:

$$|\Delta W_\tau^H| = \sum_i \sum_j |w_{ij\tau} - w_{ij(\tau-1)}| \; . \tag{3.5}$$

Likewise, I only consider the magnitude of the population PE, given that both positive and negative PE is likely to have metabolic consequences. I define this quantity, $|PE|$, as the sum of the absolute values of differences between predictions for each possible outcome, $\|a_j\|$, and the corresponding target values t_j, on the current trial:

$$|PE| = \sum_j \left| t_j - \|a_j\| \right| , \tag{3.6}$$

where the prediction $\|a_j\|$:

$$\|a_j\| = \frac{\sum_i a_i w_{ij}}{\sum_j \sum_{i'} a_{i'} w_{i'j}} \tag{3.7}$$

is a normalized activation vector as most parametrisations of Hebbian learning do not produce predictions that can be interpreted directly as probabilities.

Another quantity of interest is the classification error after learning $\mathcal{E}$. This is defined as the magnitude of the difference between prediction and true (noiseless) outcome for each cue $\mathcal{C}$, thus not only capturing how well the learning model can remember observations, but also how resilient it is to noise during learning:

$$\mathcal{E} = \sum_{\mathcal{C}} \left(\sum_j t_j^{\mathcal{C}} - \|a_j^{\mathcal{C}}\| \right) . \tag{3.8}$$

Simulations were conducted for a number of possible experimental designs, for both categorical and continuous associative learning, with various degrees of stochasticity and various numbers of cues/outcomes. The simulations were run across the range of values for learning rate and weight decay that produce plausible learning dynamics (figure 3.1). I recorded $|\Delta W^H|$ and $|PE|$ on each trial, and calculated the correlation between them.

The resulting correlations, plotted as a function of learning rate and weight decay, reveal that most of the parameter space results in strong correlations (figure 3.1). Moreover, the classification error $\mathcal{E}$ is almost identical across the parameter space (except for a region in the bottom left where both parameters are near zero), and therefore almost all parameter combinations are equally plausible for a real learner that tunes its learning parameters to the task. In other words, it is not the case that situations in which $|PE|$ and $|\Delta W^H|$ are highly correlated are non-optimal.

3.1.3 Discussion

I conclude that, while there is convincing evidence that PE is computed by some neurons, the current evidence used to implicate this neural PE signal in learning has alternative explanations. There are a few fMRI studies that correlate brain activity with PE, a subset of which go further and link this to learning outcomes. However, due to the nature of BOLD signal

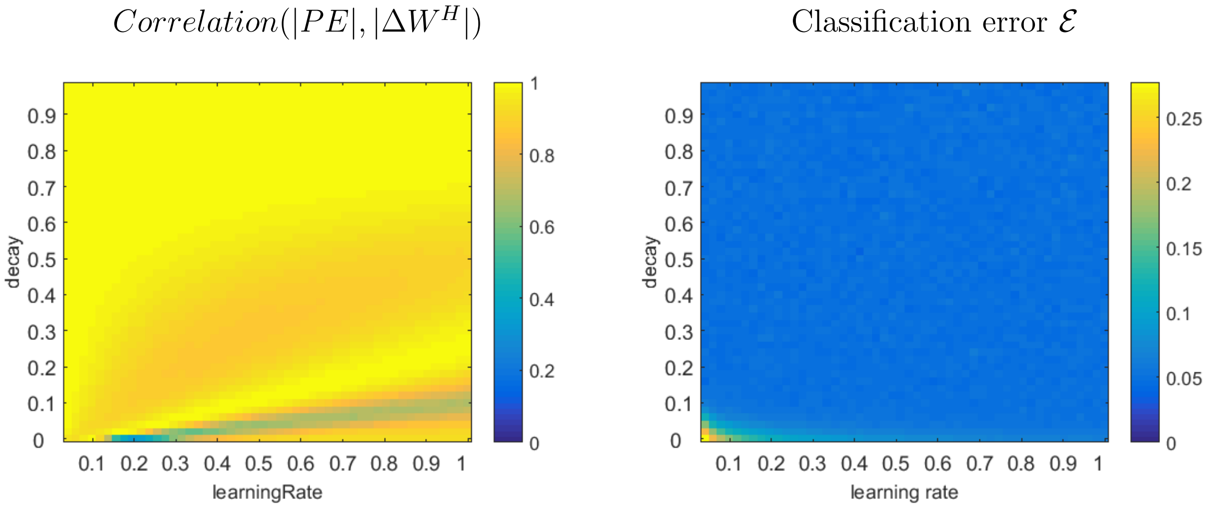


Figure 3.1: Left plot shows the correlation coefficient between $|PE|$ and $|\Delta W^H|$ as a function of learning rate and decay parameters of Hebbian learning during a quasi stochastic associative learning task. Right plot shows the average classification error $\mathcal{E}$ on the task after 50 learning trials. These particular plots reflect a learning situation where 4 cues are alternately associated with 4 distinct outcomes. 90% of the stimulus-outcomes pairs followed a particular bijective mapping, while the other stimulus-outcome pairs violated this mapping to introduce stochasticity.

measured by fMRI, the correlation with PE may not be a result of actual PE signalling, but rather a result of metabolic processes related to synaptic plasticity: computational modelling demonstrates that the magnitude of synaptic plasticity is highly correlated to PE, even when no PE computation takes place during learning. This conclusion especially affects studies such as Fletcher and colleagues' (2001) because if the possibility of observing plasticity in fMRI is accepted, then the results of this research become entirely consistent with Hebbian learning theory. Further modelling and experimental paradigms are therefore needed to establish the principles governing human associative learning at the implementational level. In the next two chapters, I return to the algorithmic level to test further behavioural evidence for PE in the context of blocking effects, and provide some novel data that is more consistent with Hebbian learning scaled by the relative informativeness of cues.

Chapter 4

Statistical inference of subjective probability distributions

In a probabilistic associative task, an agent learns a subjective probability distribution S across the outcomes following a cue. The main interest of this thesis - learning - can be seen simply as a change of the subjective distribution with exposure to a new stimulus-outcome pair, i.e. $S_{t+1} = S_t + \delta S_t$. As apparent from previous chapters, a large number of learning theories (e.g. Rescorla et al., 1972) assume that δS is not only a function of cue and outcome, but also S itself. Other theories disregard the role of S in updating of itself (e.g. Hebb, 1952). These two views of learning seem to be very different, but it has proven difficult to delineate between them. One of the main reasons for this is that subjective distributions are difficult to infer. In other words, there is a considerable gap between behaviour we observe and the statements the theories make.

The simple solution to the problem would be to ask people about their subjective probability distribution; however, when asked to provide a direct judgement of probability, people generally perform poorly (e.g. Kahneman & Tversky, 1973). Also, there is evidence for a dissociation between direct judgement and indirect choice behaviour (Franco-Watkins, Derks, & Dougherty, 2003). Therefore direct judgements are not a suitable method to observe subjective distributions.

More indirect estimates of subjective distributions include cued recall, yes/no recognition, free choice and N-alternative forced choice [NAFC]. The information about S obtained from these tasks is however limited, for the following reasons. During cued recall tasks, the participant is required to produce outcomes associated with a given cue. The information gained from such a procedure is limited to outcomes that have subjective probabilities that exceed some (unknown) threshold for memory retrieval. Yes/no recognition paradigms require the participant to judge whether a specific cue-outcome pair has been observed, but is still subject to a memory threshold, even if that threshold is lower than for recall. In free choice, participants select an outcome from all possible outcomes, but the information gained is limited to which outcome has the highest subjective probability. In the NAFC, participants select one of a subset of N-alternative outcomes. By providing control over which alternatives are offered, the experimenter can obtain more information about specific aspects of the subjective distribution (not just the peak). The resulting information is still, however, only a comparison of subjective probability of N unique outcomes. In the modification to NAFC introduced below, the N choices can include combinations of multiple outcomes, providing yet further information about the nature of S.

Alternatively one can make assumptions about the form of S, for instance we can assume that subjective distribution is the relative frequency of outcomes observed, $\mathcal{S}$, or distributions sampled from a model exposed to the same data as the participant. This approach has been adopted by a number of studies (e.g. Gläscher et al., 2010). However this approach is heavily biased by the assumptions made.

To my knowledge, a robust, assumption-free method of estimating S for individual participants is lacking in the literature. In this chapter, I propose an experimental paradigm and analytical techniques that enable this.

4.0.1 Task design

It is important to probe S during learning, i.e, interleaved with learning trials, rather than only after learning. Moreover, with NAFC, it is important to present multiple probes after each learning trial, with different choices, to better estimate S. Furthermore, the choices should include combinations of possible outcomes, which allow more precise estimation of S. In other words, not only could 2AFC be used to compare pairs of outcomes (e.g. $S_a < S_b \land S_b < S_c$), revealing the rank order of individual outcome probabilities, but it can also be used to compare the combined probability of outcomes (e.g. $S_a + S_b > S_c$) to gain extra quantitative information about S. From a set theoretical perspective, repeated NAFC can be exploited to define the smallest set of subjective distributions consistent with a participant's responses, together with a set that is not consistent. In general, the greater the number of alternative choices $NAFC$, the more combinations of outcomes can be included in one choice and thus finer information about S obtained.

S (like any other probability distribution that must integrate to 1) exists on a simplex, which, for N^O outcomes, is a $(N^O - 1)$-dimensional triangle positioned within N^O-dimensional space. When querying the distribution by NAFC, we effectively partition the set of all subjective distributions (the simplex) into a part that complies with the participant's response and a part that does not, via a $(N^O - 2)$-dimensional surface. As a result, the proportion of compliant to non-compliant space that can be defined from one NAFC trial will exponentially increase with N^O, and exponentially more NAFC trials will be needed to find the smallest identifiable subset of S compliant with the responses. If an experiment is to be used with human participants, we need to keep the number of queries to a reasonable number, i.e. ensure N^O is not too large. In the experiment described in section 4.1.3, I used $N^O = 3$ and the number of choices in NAFC to be two (2AFC) for practical purposes.

Each trial of an experiment of this type involves presentation of a cue and an NAFC task for participants to select the outcome they expect. In learning trials, their choice is followed by an outcome. If their choice matches the outcome, the participant is rewarded[1]. One or more probe trials can then be interspersed with learning trials. Probe trials involve a cue and NAFC choice, but these are not followed by an outcome, in order to minimize updating of S during the probe trials themselves.

Assuming participants are reward maximisers, they will always select the more probable alternative during 2AFC based on S. Therefore, repeating 2AFC with different configurations of choices will lead to the smallest identifiable subset of probability distributions that contains the actual subjective distribution of the participant at a given time. This

[1] For practical purposes, I used points as reward, and the participant has to collect a certain number of points to finish the experiment. It has been demonstrated that effort and time are minimized by participants (e.g. Shenhav, Botvinick, & Cohen, 2013), therefore the points are a suitable reward for present purposes.

approach has, however, two problems. First, the set of distributions we can identify by repeated NAFC will still include an uncountably infinite number of (continuous) subjective distributions without any means of distinguishing between them. Secondly, participants sometimes contradict themselves and therefore if we treat them as deterministic agents, the set of compliant subjective distributions might be empty. To counter both of these problems, I adopted a probabilistic approach to infer S. In the following section, I define a generative model for participants' data and then invert it to calculate the likelihood across subjective distributions.

4.0.2 Generative model

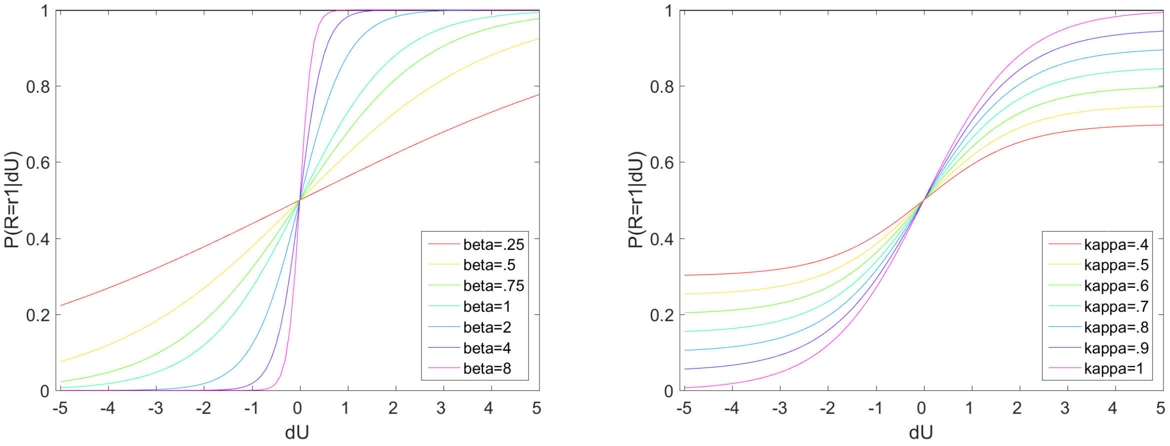
The rational approach to the task is simple: participants should pick the option that has the higher expected utility. As the reward function is binary in our experimental paradigm, the expected utility is simply the likelihood of the choices as estimated by S. However, participants are not perfect deterministic agents. To counter this problem, we formalize the decision-making model in a way that allows for quasi-stochastic decision making. Firstly, we assume that the participants are agents sensitive to the difference in the expected utility, being more likely to select the better option as the difference in expected utility between the option increases. The sensitivity can vary between participants and is characterised by a parameter β. We assume that the cause of less-than-perfect sensitivity in decision making comes from Gaussian noise in the "read out" of the subjective distribution. The softmax function can be used to model this, which in the case of two alternatives becomes a simple sigmoidal function:

$$P(R = R_1|dU) = \frac{1}{1 + e^{-\beta dU}} \tag{4.1}$$

where R is the actual response made by the participant, R_1 is a response 1 and dU is a relative difference between the expected utilities defined as

$$dU = \log \frac{S(R_1)}{S(R_2)}\ . \tag{4.2}$$

This model assigns probability of 1 and 0 to the responses for extreme dU which rarely matches human performance. This motivates extending the model to account for resid-



(a) $\kappa = 1$ **(b)** $\beta = 1$

Figure 4.1: Demonstration of the effect of decision-making parameters.

ual randomness in the decision-making by adding another participant-specific parameter κ, which corresponds to the proportion of responses that are drawn from a Bernoulli distribution, $R \sim B(\frac{1}{2})$, in other words:

$$P(R = R_1|dU) = \frac{1-\kappa}{2} + \frac{\kappa}{1+e^{-\beta dU}} \tag{4.3}$$

The effect of β and κ is shown in figure 4.1.

The resulting model defines the likelihood of responses R to be produced by an agent with decision-making parameters β and κ and a subjective distribution S.

Frequency matching

The decision-making literature, however, describes another decision-making model as well. Frequency matching [FM] is clearly not rational, yet its use by humans is well documented (for review see Brehmer, 1999). When participants use FM, they effectively match the probability of outcome with their responses. This decision-making model can be defined as

$$P(R = R_1|S(R_1), S(R_2))) = \frac{S(R_1)}{S(R_1)+S(R_2)}\ , \tag{4.4}$$

which is also a sigmoidal function. In fact, if we substitute Equation 4.2 into Equation 4.3 and equate it to 4.4, we can solve for β and κ.

$$\frac{1-\kappa}{2} + \frac{\kappa}{1+e^{-\beta \log\frac{S(R_1)}{S(R_2)}}} = \frac{S(R_1)}{S(R_1)+S(R_2)} \tag{4.5}$$

which is true when $\beta = 1$ and $\kappa = 1$. This means that FM is just a special case of the rational model with noise, and therefore if we use the rational model with free parameters for noise we can account for either of the decision-making models or their mixture.

4.1 Methods

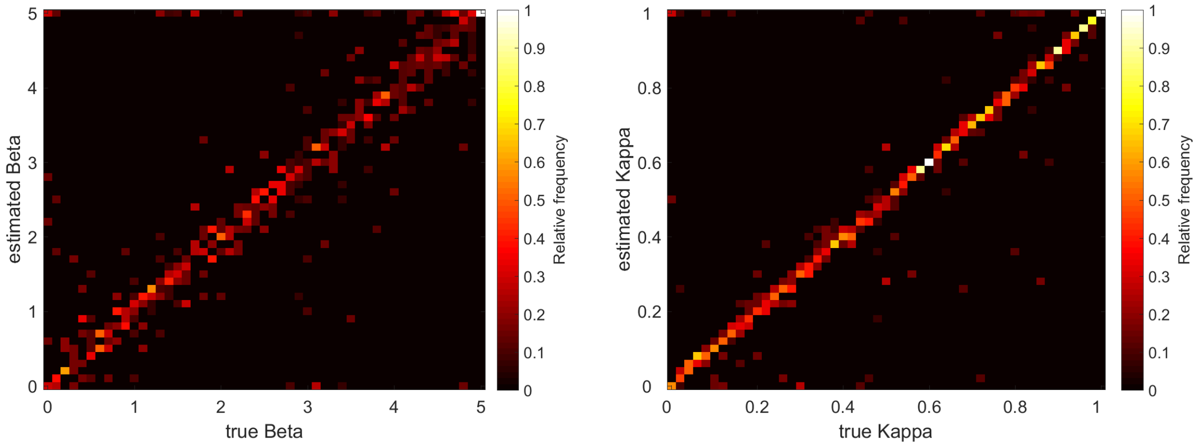
The task is to find the likelihood function across the subjective distribution space S for any point being the true subjective distribution $\bar{S}$ of participant p after learning trial t. It is useful to consider the collection of all subjective distributions being inferred as $S_{p,t}$. The generative model described by Equation 4.3 can be inverted for this purpose. To constrain the analysis, I assume that β and κ are fixed for each participant during the entire experiment. I can exploit this assumption to infer likelihood functions for all learning trials from one participant, $S_{p,\bullet}$.

The best approach would be to find $L(S_{p,\bullet}, \beta_p, \kappa_p | R_{p,\bullet}, E_{p,\bullet})$, where E is a description of the experimental setup (i.e. history of cues, outcomes and arrangement of possible responses), and marginalize β_p and κ_p. However the lack of upper bound on β makes the integral go to infinity. To circumvent this problem, I find $ML(\beta_p, \kappa_p | \mathcal{S}_{p,\bullet}, R_{p,\bullet}, E_{p,\bullet})$ and use the resulting estimates $\hat{\beta}_p$ and $\hat{\kappa}_p$ in subsequent analysis. $\mathcal{S}_{p,\bullet}$ is an initial guess on $\bar{S}_{p,\bullet}$ based on the rational approach to our task. It is necessary to start with a guess because any estimate $\hat{S}_{p,\bullet}$ would have to be function of the true values of the decision making parameters $\bar{\beta}_p$ and $\bar{\kappa}_p$ which are not yet estimated.

4.1.1 Estimation of decision-making parameters

Since there is no closed-form solution to this problem (it is over-parametrised), we need to find $argmax(L(\mathcal{S}_{p,\bullet}, R_{p,\bullet}, E_{p,\bullet} | \beta_p, \kappa_p))$ by methods of numerical optimisation. I used a non-gradient based solver from MATLAB Optimisation Toolbox - *fminsearch* - for this task.

The initial values for fminsearch used were $\beta = 1$ and $\kappa = .94$. During validation of this procedure, it became apparent that the solver was failing to converge to a correct solution. Investigation of error surfaces revealed that for low values of $\bar{\beta}_p$ and $\bar{\kappa_p}$, the error surface around the initial value is flat, preventing the solver from finding the true minimum. To



(a) Quality of β estimation

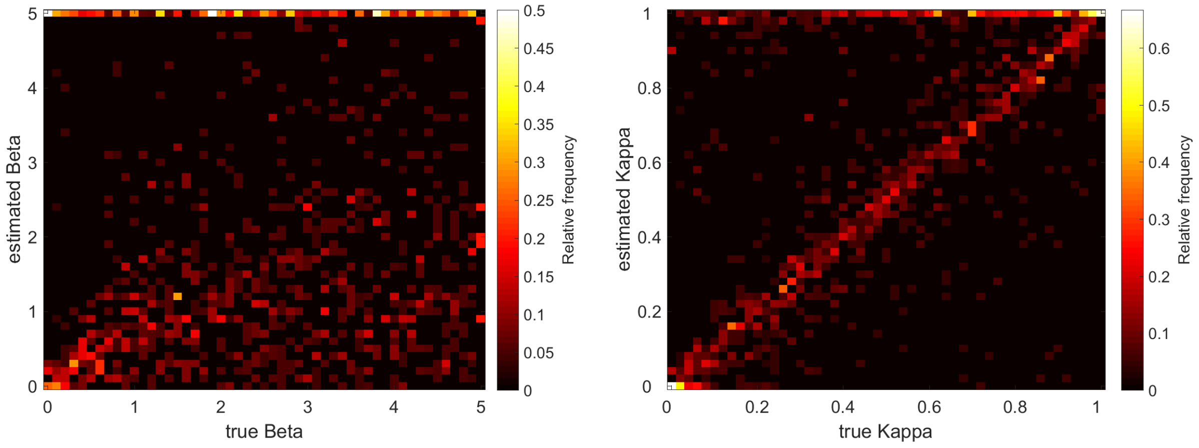
(b) Quality of κ estimation

Figure 4.2: Demonstration of the validity of the decision-making parameters estimation procedure. The datasets used were produced by rational learner and consisted of 100 instances of sampling of 100 different $\bar{S}$.

overcome this issue, I attempt optimisation 15 times with different initial parameters. On the first attempt, I still initialize $\beta = 1$ and $\kappa = .94$; on the second attempt, I set $\beta = .5$ and $\kappa = .94$; and thereafter I sample $\beta \sim U(0, 10)$ and $\kappa \sim U(.5, 1)$. The motivation for the hard limits on sampling is that $\beta < 0$ corresponds to participants' intentionally deciding against their belief, which I assume does not happen. For the increasing values of β, the decision-making model quickly approaches a step function if the subjective distributions are similar to $\mathcal{S}$, which I assume they are. κ is bounded by $0 \leq \kappa \leq 1$ by definition 4.3, and I ignore the lower half of the range because the error surface for these values is very flat. It is generally not difficult for the solver to move from high κ to low κ, while it is almost impossible for the solver to move in the other direction. Additionally the parameters are provided in appendix B.

Validation

I provide two validations of this model. Both of them use an artificial dataset to allow us to compare $\bar{\beta}$ and $\bar{\kappa}$ with $\hat{\beta}$ and $\hat{\kappa}$. The validation in Figure 4.2 comes from a large dataset and an agent following our assumptions about $\mathcal{S}$. This demonstrates that the inference



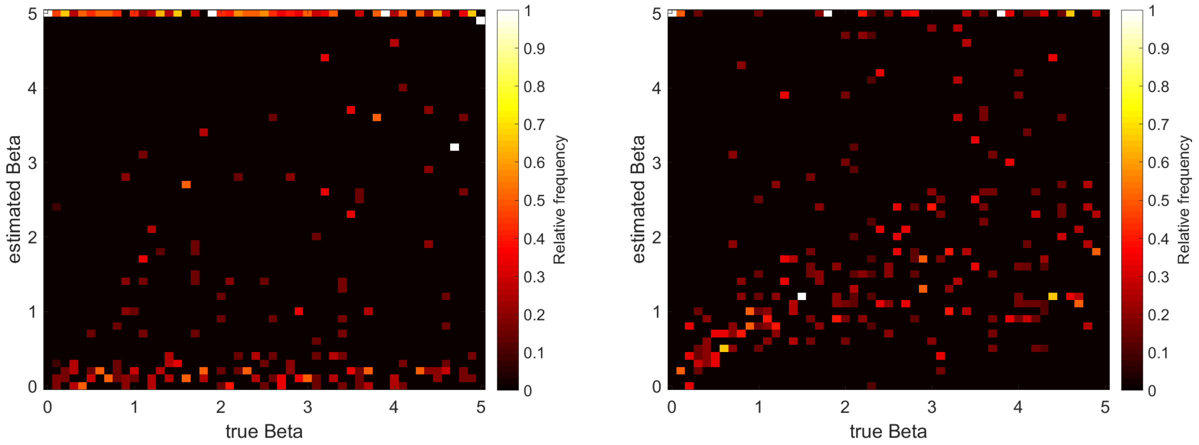
(a) Quality of β estimation

(b) Quality of κ estimation

Figure 4.3: Demonstration of the quality of decision-making parameters estimation for a realistically sized datasets (10 times sampling 60 different $\bar{S}$) produced by non-rational learners.

procedure is correct. The validation shown in Figure 4.3 uses a dataset of a size similar to what can be realistically obtained from human participants, and the agent producing the dataset did not use the normative approach to the task, i.e. $\mathcal{S}$ was not obtained by calculating the relative frequency, but by an alternative, suboptimal learning process: The learning process was either Widrow-Hoff learning or Hebbian learning with equal probability. The free parameters of these learning models were set as $k \sim B(4, 10)$ and $d \sim B(2, 20)$.

There are two reasons why Figure 4.3a is unsatisfactory, which at the same time are reasons why the apparently poor estimation of $\bar{\beta}$ does not pose a problem for our ultimate aim of calculating the likelihood of subjective distributions. The first reason is that a large proportion of the mis-estimation in β is due to a low value of $\bar{\kappa}$, and this issue disappears for larger values or κ (Figure 4.4). When κ is small, β has only a limited impact on decision-making (see Figure 4.1a). The second reason is that as β increases, the impact of any variance in β on decision-making decreases due to its exponential nature (Figure 4.1a).



(a) $\bar{\kappa} < .25$

(b) $\bar{\kappa} > .75$

Figure 4.4: Demonstration of the relationship between $\bar{\kappa}$ and the quality of β estimation.

4.1.2 Calculating likelihood of subjective distributions

The second step of the inference procedure consists of using $\hat{\beta}_p$ and $\hat{\kappa}_p$ to calculate the likelihood of a subjective distribution being the true subjective distribution of a participant, i.e:

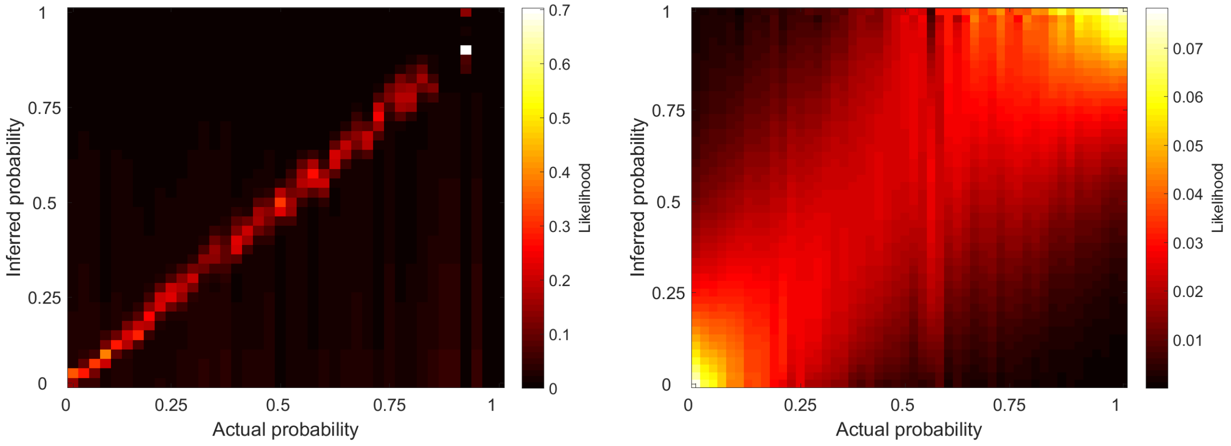
$$L(S_{p,\bullet} = \bar{S}_{p,\bullet} | \hat{\beta}_p, \hat{\kappa}_p, R_{p,\bullet}, E_{p,\bullet}) \ . \tag{4.6}$$

As there is no closed form solution to this equation, the likelihood mass is computed numerically by computing the likelihood across the whole probability simplex using Equation 4.3.

Validation

Similarly to Section 4.1.1, I provide two validation reports. These compare
$L(S_{p,t} = \bar{S}_{p,t} | \bar{\beta}, \bar{\kappa}, R_{p,t}, E_{p,t})$ with $\bar{S}_{p,t}$. Since $S_{p,t}$ exists in a two-dimensional space, I cannot demonstrate an elegant validation of the inference like in Section 4.1.1 because of the high dimensionality of the resulting plot. Instead I separately plot the probability of each of the three states the outcome can take. The resulting validation plots in figure 4.5a show how well the inference procedure finds the true probability of an outcome.

Similarly to Section 4.1.1, the first validation shown in Figure 4.5a is a result of the S inference procedure performed on a large, normative dataset. The dataset consisted of 1000



(a) large dataset

(b) 'realistic' dataset

Figure 4.5: Demonstration of the validity of the inference procedure for subjective distributions.

samples from 100 different $\bar{S}$, where $\bar{\beta}$ and $\bar{\kappa}$ were used by the inference procedure. Figure 4.5a therefore shows only that the inference procedure is correctly implemented, but not that it will be actually useful with data gathered from humans.

Figure 4.5b demonstrates validation with a realistically-sized dataset. This dataset consisted of 10 samples from 60 $\bar{S}$, while the decision-making parameters were not known, but instead estimated by the first step of our inference procedure. The inference validates quite well, despite the severe mis-estimation of β seen in figure 4.3a. The area of higher mis-estimation seen for the high values of $\bar{S}_j$ on the likelihood function in Figure 4.5b is caused by the fact that probability distributions with such values are less frequent in the dataset due to uniform sampling of probability distributions from a flat Dirichlet distribution.

4.1.3 Experimental paradigm

Now I have a means to estimate the likelihood of subjective distributions from multiple 2AFC probes, the next task is to design an experiment that can make use of these analytical techniques to reveal the nature of learning. Learning can be understood as a transfer from one subjective probability distribution to another, given some data.

The web-based experimental paradigm I developed for this purpose consists of repeated

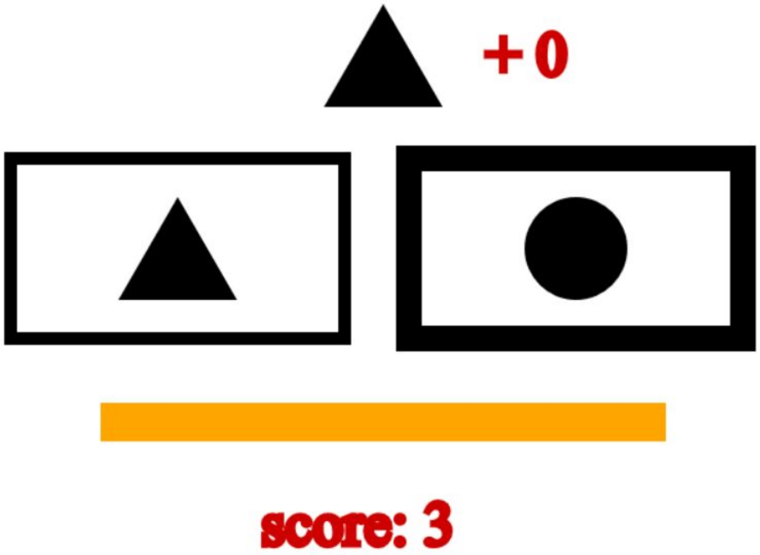


Figure 4.6: Screenshot of experiment as presented to the participant. Orange bar represents the cue on the trial. The two boxes represent the 2AFC options, the boldened one being the one selected by the participant. The triangle at the top is the outcome on the trial that was displayed once the participant selected a response. +0 signifies that the response did not match the outcome and therefore no points were gained on this learning trial.

exposure to new data and subsequent probing of (the resulting changes in) subjective distributions. The full experiment exactly as it was presented to the participants can be found at https://learning.mrc-cbu.cam.ac.uk. Participants first go through a brief training phase during which the experiment is explained to them in an interactive manner. On each trial, participants are presented with either blue or orange colour (cue) which has a relationship to a shape (circle/triangle/square) that will later appear on a screen (i.e, $N^O = 3$ outcomes). Participants are offered two boxes that contain one or more shapes (Figure 4.6) and asked to pick the box that they think contains the outcome. After the participant makes their choice, and if the trial is a learning trial, the outcome appears, and the participant's score is updated and shown. If the trial is a probe trial, then no outcome nor score is presented.

After the initial training, participants proceed with the task in its most basic form. Initially, all trials are single-cue learning trials. After participants reach 10 points however, compound trials are introduced with a probability of .25. Compound trials contain two cues

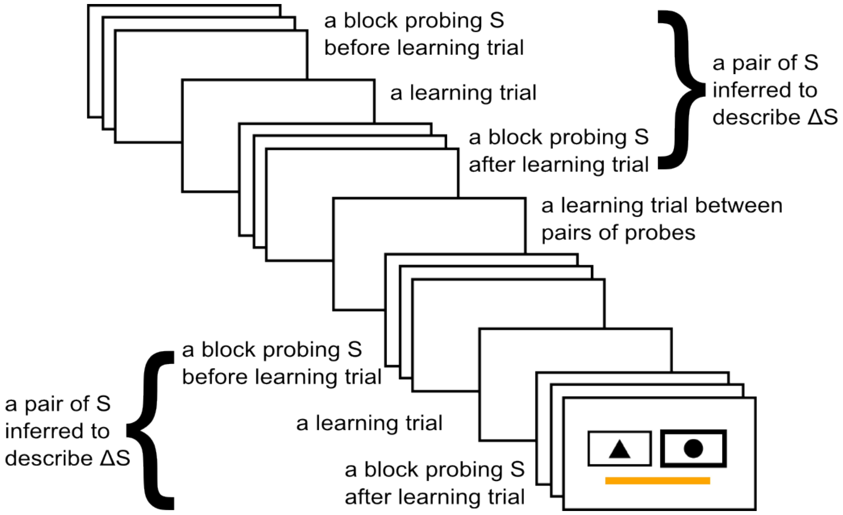

a block probing S before learning trial
a learning trial
a block probing S after learning trial
a pair of S inferred to describe ΔS
a learning trial between pairs of probes
a block probing S before learning trial
a pair of S inferred to describe ΔS
a learning trial
a block probing S after learning trial

Figure 4.7: Illustration of the experimental procedure proposed in this chapter.

presented simultaneously, and are to test various types of blocking effects, as explained in Section 5.3. After participants reach 20 points, the main part of the experiment begins.

The main phase illustrated by figure 4.7 consists of pairs of blocks probing the subjective distributions and a learning trial between them (i.e. within the pair). This allows us to infer S before and after the learning trial and therefore look at the change of S caused by the learning trial. An extra learning trial is introduced between the pairs of probing blocks (i.e. outside the pair) for practical purposes. The change in S is not inferred for this trial. The learning trials within a pair of probing blocks are compound learning trials (both cues presented at the same time) with probability of .5 else they are single-cue trials. The learning trials outside pairs of probing blocks are always single-cue trials to make the task easier. The learning trials are the only trials after which the outcome is displayed to the participant.

To maximise the information gain, each block starts with 2AFC between two random single outcomes. The subsequent 2AFC options are selected to gain the maximum amount of information about the participant's subjective distribution from the set-theoretical point of view described in Section 4.0.1, simply by forcing the participant to decide between options that will eliminate the largest set of subjective probability distributions until the combinations that can further decrease the set compliant with responses are exhausted[2]. These trials

[2] This approach assumes the participant's decision-making is deterministic and is not optimal in a probabilistic context. Designing an optimal information gain procedure in a probabilistic context would be a

are intermixed with random trials to increase the amount of information collected and to make it harder for the participant to detect a pattern in the probing blocks. On average, each probing block consists of 10 trials equally distributed between the two cues. Probing of both cues is randomly intermixed to further remove structure from the task. Both probing blocks before and after a learning trial consist of identical trials, therefore on average a pair of probing blocks consists of 20 probing trials with a learning trial in the middle. This is necessary to compare the likelihood density before and after a learning trial as the quality of the density estimation depends on the trials used for probing. However, both the trial order and arrangement on the screen are permuted before the probing block is repeated to decrease the participant's ability to recall their own responses. None of twelve pilot participants realized that the two probing blocks in a pair consisted of identical trials.

Data collection

The real data were collected and analysed according to the procedures described in Chapter 4. The experiment is readily accessible at https://learning.mrc-cbu.cam.ac.uk including the source code. The dataset was collected online, with participants recruited via FaceBook advertisements to maximise the number of participants. These adverts asked participants to "help researchers learn about brains". No payment was made to participants in order to make the motivation more similar to the naturalistic latent learning. Because of constraints on ethical approval for this recruitment, no personal information was collected. The only demographic information on the sample is that: 1) the advertisement was only displayed in English speaking countries, 2) 78% of people who saw the advertisement were female and 3) the most represented age-group was 45-55 years old. No demographic information on the people who actually participated (as opposed to just saw the advertisement) are available.

There is no doubt that the sample was self-selected to a degree and that there are ways in which they systematically differ from the population; however, we believe that the sampling bias exists in many other studies too, such as those done only on student volunteers studying for psychology degrees.

Many participants left the experiment before it has finished, meaning that the amount of

useful extension of the project.

	N
participants	1,990
trials	214,508
subjective distributions inferred	20,618
compound learning instances	5,132

Table 4.1: Amounts of data collected and analysed. Each participant contributed a different amounts of data and the main focus was on compound learning trials.

data differs greatly between participants. This is not a problem issue for the analysis methods, because I only examine aggregate performance of the entire sample. Various counts of the sample are provided in Table 4.1. Note that the number of participants is not as relevant as the total number of trials – more specifically, the number of compound learning trials from which we can infer the subjective probability distribution before and after learning. The average performance on the task was 63.37%, with chance performance being 50%. I did not exclude any participants, even when they performed significantly worse than chance. The performance metric was obtained by comparison against a simple frequentist model, that is rational in the context of our task.

4.2 Discussion

I have described an experimental procedure for probing subjective distributions, combined with a method for statistical inference, which I believe offers better insight into the nature of human learning. Simulations show that even if our assumptions of rational behaviour are violated, and there is a limited amount of data, the procedure still provides a reasonable probabilistic estimate of the subjective distributions held by an agent. This approach, that is fully probabilistic, provides significant advantage over existing approaches such as Kalman filter, which are not suitable for inference over probability space, not parameter-free and inadequate for capturing non-parametric distributions.

These estimates can be useful for research into the theory of learning in probabilistic associative tasks because they are not biased by the process that generated them, and therefore might provide new insights about the processes that drive learning. Moreover as the estimates are probabilistic, they provide opportunity for probabilistic approach in further analysis by preserving entire distributions as opposed to only point estimates.

The main limitation of the method presented here is that I assume subjective distributions to be static during the periods of testing when no outcomes nor reward are present. This is likely not to be true in human participants (e.g. Bridge & Paller, 2012). However I am not aware of any method of estimation of subjective distributions which does not suffer from this problem.

In the next chapter, I apply this inference method to a large online dataset to delineate between learning theories driven by PE and those that are not.

Chapter 5

Is probabilistic associative learning driven by PE?

Chapter 4 described a probabilistic associative learning paradigm and inference method for estimating subjective probabilities. In this chapter, I describe how certain types of compound trials can produce different types of blocking, which are able to distinguish between PE-based and non-PE based learning (unlike conventional paradigms and analyses). In particular, I define a new type of blocking, which is more discriminative than conventional blocking, and demonstrate how the results from a large online dataset are inconsistent with PE-based learning.

5.1 Blocking

Ever since Kamin's (1969) blocking effect became mainstream, learning theory became dominated by models that learn by correcting PE (e.g. Rescorla et al., 1972). This was seen as necessary given the inability of older non-PE models of learning (e.g. Hebb, 1952) to explain the blocking effect. This paradigm-shift makes blocking probably the most influential effect in the history of learning theory.

However, there are two reasons why I do not consider blocking to be sufficient evidence for PE as the driving force behind learning. Firstly, as elaborated in Chapter 2, Kamin's classic findings can potentially be explained by non-PE learning scaled by relative informativeness of a cue α_i (defined by Equation 2.16).

Secondly, recent research indicates that blocking as described by Kamin (1969) is not as replicable as many believe (Maes et al., 2016). Through the lengthy series of 17 experiments, Maes and colleagues (2016) demonstrated that the blocking effect either has important boundary conditions or requirements that are not described by the classic definition of blocking.

Furthermore I see a fundamental problem in the connection between learning theories and behaviour. While almost all mechanistic theories of learning describe learning as a change of weights, it is unclear what the weights are, how can we measure them, or how they translate to behaviour. Therefore there is a need to redefine blocking and other predictions of learning theories in a context that is invariant to the process of translating weights into behaviour.

The methodology introduced in Chapter 4 enables the translation from behaviour into subjective probability distributions. In the first, theoretical part of this chapter, I first discuss the connection between weights and subjective probability distributions. This allows the identification of the properties of learning theories with respect to subjective probability distributions, explicitly linking the learning theories to the behaviour and data. I then redefine blocking effect in terms of motion in the space of subjective probability distributions. Crucially I extend the blocking effect to compound learning with any two subjective probability distributions involved, as opposed to only one specific configuration considered in Kamin's classic blocking paradigms. When generalized in this way, PE and non-PE learning theories predict qualitatively different, but still correlated outcomes during blocking. Next, I describe a novel type of blocking for which the predictions of the two classes of learning theories differ in a more fundamental way.

In the second, empirical part of this chapter, I investigate whether these two blocking effects actually occur in a large dataset collected on-line, using a paradigm like that described in Chapter 4 (I also test for Kamin's classic blocking effect). I then determine whether the results are better explained by PE-driven learning, or the Hebbian learning scaled by relative informativeness that was proposed in Section 2.3.3.

From weights to responses

Most of the prominent theories of associative learning (e.g. Rescorla et al., 1972) describe learning as a process by which weights in an associative network are changed. However, the weights do not have a clear correspondence to any property of biological systems that are responsible for associative learning, and are not directly measurable. Instead, in human associative learning, we observe responses. Therefore to reconcile the theory with the data we need to bridge the gap between weights and responses.

The main issue in this endeavour is that there are multiple processes that need to take place in the brain to produce responses based on weights. There are also processes that take place at the same time as learning and affect how the learning process will manifest in the data. Firstly, the responses we observe in our experiments and that make up our data are a result of decision-making. Secondly, the decision-making is not based on weights, but rather on some probability-like interpretation of weights, i.e. the *subjective probability distributions* described in Chapter 4. These subjective probability distributions are a result of a *read-out* function applied to the weights. Lastly, we need to account for homeostatic processes that cause forgetting and weight normalization that we can capture as a single *decay* process. Neither of these processes has been described well enough to allow us to simply take a model of that process from the literature.

In Chapter 4, I proposed a method of inferring the basic properties of a participant's decision making relevant to our task. Combined with the experimental procedure described also in Chapter 4, this allows me to infer the likelihood of subjective probability distributions at a given point in time. This means effectively inverting the decision-making function.

Unfortunately, the read-out function can not be inverted. This is because weights are bounded by the decay process that acts gradually on the weights to normalize and equalize them, thus while weights are approximately bound to a constant, the bound is soft. On the other hand, the subjective probability distributions must be bounded by the axioms of probability to allow for effective decision-making. This bound can be considered hard in contrast with the bound on weights. As a result, for any subjective probability distribution there is an infinite number of weight vectors that can produce it through the read-out function. In the theoretical part of this chapter, I consider the learning dynamics of systems with two broad classes of read-out function that I find plausible. The first one is a simple linear normalisation. As non-linear normalisation is too large a class of normalising functions, for the second one, I limit analysis to the softmax function.

The decay process acting on the weights, being up-stream in the processing pipeline from read-out, cannot be uniquely determined for the same reasons as above. To keep the analysis relatively assumption-free, I consider both linear and non-linear decay. Non-linear decay is not a standard feature of learning models, despite the fact that it seems more likely than the standard linear decay process. The particular implementation I used in the simulations here is:

$$W_{i,t+1} = W_{i,t}(1-d) + softmax(W_{i,t}, \tau)d \tag{5.1}$$

where $W_{i,t}$ is a weight vector corresponding to cue i at time t, τ is temperature parameter of the softmax function and d is the decay ratio. This process was applied to all weight vectors after each trial (even when a particular cue was not shown). The effect of this decay algorithm is to move the weights towards their softmax transformation by proportion d.

5.2 Theory of learning in subjective probability space

Any point in subjective probability space corresponds to a particular subjective probability distribution S, therefore this space can be understood as a probability simplex. Due to the axioms of probability, any discrete probability distribution with N states is located on an $N-1$ dimensional hyper-triangle. To offer the reader an intuitive understanding that is only possible in a two-dimensional space (and in keeping with the three outcomes used in the later experiment), consider the case of $N = 3$. Any discrete probability distribution $S^{[o_1,o_2,o_3]}$ over the probabilities of the three outcomes can therefore also be expressed by its location on the probability simplex $S^{[x,y]}$. The $[x, y]$ coordinate system I use here has the origin halfway along the centre of the bottom side of the triangle.

With this 2D representation, learning and decay are characterised by movements between points on the simplex. As illustrated in Figure 5.1, the learning function, $\mathcal{L}$ is a function of weights W before learning, cue c, outcome o and other unknown variables, ?, at time t that

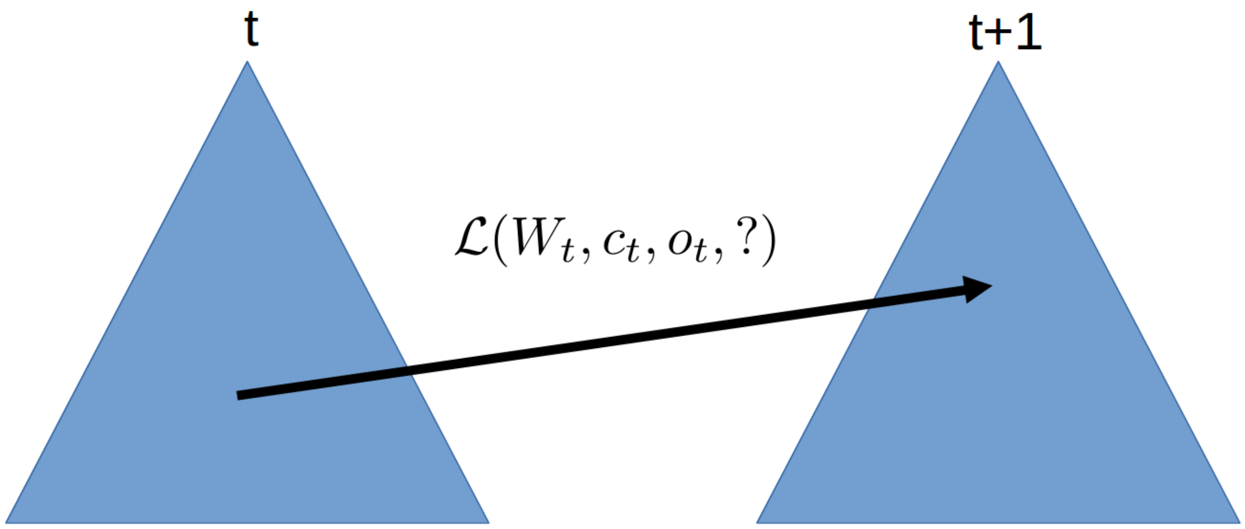


Figure 5.1: The learning function $\mathcal{L}$ can be understood as a vector movement along the probability simplex defined by the weights, cue, outcome and other unknown variables at the time t.

produce an R^2-valued motion vector, ΔS, along the simplex:

$$\mathcal{L}(W_t, c_t, o_t, ?) = \Delta S_t = S_t - S_{t+1} \in R^2 \tag{5.2}$$

A single vector ΔS_t does not tell us much about learning; however, a set of ΔS_t vectors across the entire simplex provide an approximate description of $\mathcal{L}$. The main problem with this approach is the curse of dimensionality; as described in Chapter 4, a significant number of trials (5-20 depending on required accuracy) are required to describe a subjective distribution at a single time-point, thus necessitating an impractical number of trials per participant. We will address this issue by replacing W_t, a 3D unobservable unbounded vector with the 2D observable bounded vector $S_t^{[x,y]}$. While this requires assumptions about the read-out function, I will do this for a range of read-out functions and focus on those properties of $\mathcal{L}$ that hold across the different realizations of read-out.

However, we have yet another dimension: the specific cue presented on a trial. We can further decrease the dimensionality by coregistering the simplices between two cues: one for the cue presented (learning) and one for a cue not presented (forgetting). Both of the simplices have the outcome presented at a given trial at their top corner (indicated in subsequent figures by a red dot).

5.2.1 Learning as a flow

The set of the coregistered vectors across simplices constitute a flow diagram identical to sampling from the function $\mathcal{L}$. It is useful to first produce synthetic flow diagrams for various configurations of single-cue learning, before looking at compound learning or the real data. Figure 5.2 shows the flow across simplices for Rescorla-Wagner learning with various configurations of read-out and decay processes. The most simple learning dynamics using linear decay and linear read-out shown in sub-plot a) simply demonstrate that the subjective probability distributions for the presented cue move towards the outcome presented. Sub-plot b) shows the corresponding flow for a non-presented cue: here, the read-out and decay processes perfectly counteract each other and the only movement left is random noise.

Sub-plots c) and d) show the same learning and forgetting dynamics with linear read-out and softmax decay processes. This time, forgetting is apparent in non-linear motion towards the centre of the simplex. In other words, the probability distribution becomes flat for non-presented cues. Interestingly, the non-linear motion depends on the τ parameter.

Finally, sub-plots e) and f) show how the motion for softmax read-out and linear decay becomes non-linear for both learning and forgetting. An interesting feature of sub-plot e) is that the attractor on the simplex has now moved from the top corner towards the centre. This indicates that extreme probability distributions can no longer be supported under this read-out process. The y-coordinate of the attractor is dependent on the decay parameter d and τ.

In conclusion, there is a significant degree of variation in learning dynamics from the Rescorla-Wagner learning rule incurred by changing assumptions about memory that are not explicitly mentioned in the model, but necessary to simulate human performance. The learning dynamics are characterised by the y-coordinate of the attractor and the non-linearity of approach to the attractor for the simplices corresponding to the presented cue. The forgetting dynamics are characterised by the presence of an attractor and approach towards the attractor for the simplices corresponding to a non-presented cue.

Figure 5.3, on the other hand, shows the dynamics arising from Hebbian learning. It is immediately apparent that the characteristics of dynamics are very similar to those for

Rescorla-Wagner learning. While the relationship between the free parameters in the learning models and the characteristics of the dynamics is slightly different for Hebbian and Rescorla-Wagner learning, the variation in those dynamics is the same for both types of learning models. Indeed, it can be analytically demonstrated that in the probability space any dynamics resulting from Rescorla-Wagner learning can also be a result of Hebbian learning. While this statement is not true for the weight-space (see proof in Section 3.1.1), we can only observe probability-space in behavioural data.

Studying the flow dynamics across simplices for PE-driven and non-PE learning algorithms demonstrates that single-cue learning cannot help us to delineate between these two classes of learning algorithms. However, understanding single-cue learning as a flow across a probability simplex equips us with analytical tools helpful for similar investigation in compound learning.

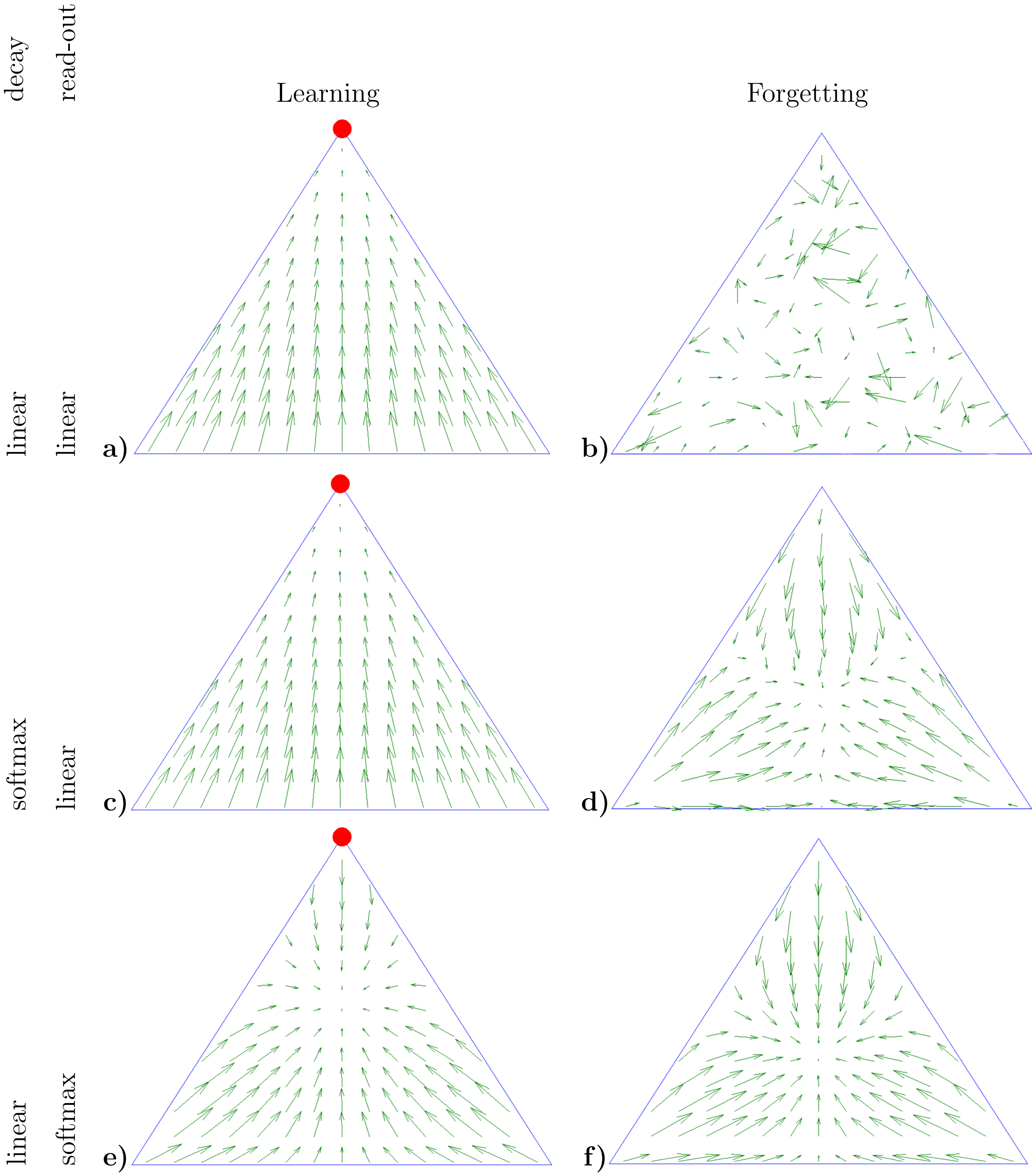


Figure 5.2: Simulated Rescorla-Wagner learning dynamics across probability simplices for different assumptions about decay and read-out processes. Learning and Forgetting labels correspond merely to whether or not the given cue was present on the particular trial. Red dots mark the outcome that was presented on the learning trial.

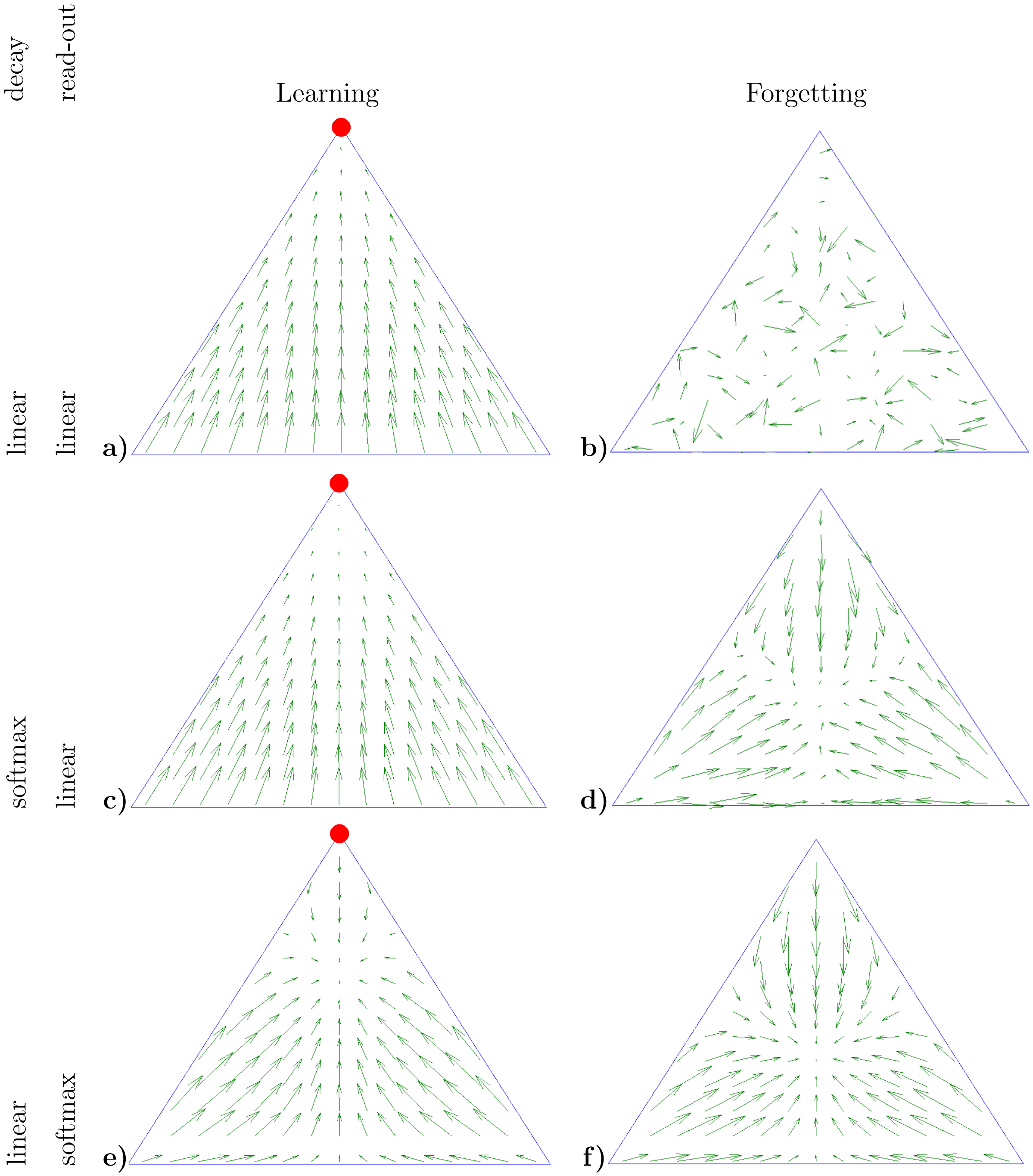


Figure 5.3: Simulated Hebbian learning dynamics across probability simplices for different assumptions about decay and read-out processes. Learning and Forgetting labels correspond merely to whether or not the given cue was present on the particular trial. Red dot marks the outcome that was presented on the learning trial.

5.3 Compound learning

To look beyond the above case of single-cue learning, I first generalize the blocking effect to a general associative learning situation and define all variables involved. Then I describe another effect that can be helpful to delineate the PE and non-PE classes of learning theories but to our knowledge has not been studied before.

Blocking as described by Kamin (1969) is an effect quantified by the difference in learning to associate an outcome with a cue (cue A) that is presented in compound with another cue (cue B), as a function of whether the other cue (B) has previously been paired with that outcome (experimental condition) or not (condition condition). The rationale behind the blocking effect is that in the experimental condition the outcome is already predicted, therefore there is less PE, therefore there is less learning of the new cue.

This interpretation rests on the following assumptions: 1) the probability distribution across outcomes for a novel cue is flat; 2) the probability distribution across outcomes for an already-associated cue is predictive of the outcome to some degree.

In terms of the above simplex conceptualisation, cue A has a subjective probability distribution located in the centre of mass of the simplex (i.e. distribution is flat), while cue B is either located at the attractor of the learning simplex if it has been already learnt (experimental condition) or in the centre of mass if it is novel (control group). Therefore Kamin's statement can be reformulated as: The flow of cue A's subjective probability distribution from the simplex centre towards the attractor is slower when cue B's subjective probability distribution is located at the attractor than it is when cue B's subjective probability distribution lies at the centre of the simplex.

From this definition, it can be seen that blocking can equally be explained by the relative informativeness account, as well as the classic PE account: Cue B has higher informativeness when it lies away from the centre of the simplex (i.e, towards the attractor in the experimental condition), and therefore cue B effectively reduces the velocity of flow for any subjective probability distribution associated with cue A.

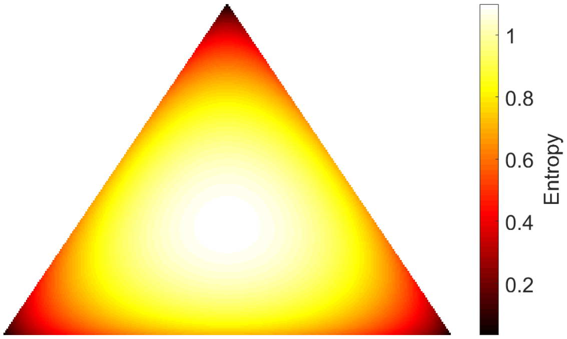


Figure 5.4: Entropy across probability simplex.

5.3.1 Generalised blocking

The classic blocking effect is just one example of blocking. Blocking can be generalized as the distribution of the motion of subjective distribution S in the y direction, which is a function of the subjective distribution associated with the other cue in the compound, $\neg S^{[x,y]}$. Because looking at ΔS (a 4D object) as a function of $\neg S$ (a 2D object) is impractical, it is essential to derive a lower-dimensional property of ΔS that captures the effect of interest. Now the presence of an attractor within the simplex, which follows from axioms of probability, dictates that the mean y component of motion across a simplex must be 0. Therefore, rather than affecting the mean, the influence of $\neg S^{[x,y]}$ should affect the skewness, γ, of the distribution obtained by sampling motion vectors. Under this definition of blocking, PE learning always dictates a decrease in $\gamma(\Delta S^{[y]})$ as $\neg S^{[y]}$ increases. On the other hand, relative informativeness implies that $\gamma(|\Delta S^{[y]}|)$ is scaled by the entropy of $\neg S$; in other words, as $\neg S^{[x,y]}$ moves away from the simplex centre of mass (see Figure 5.4 for illustration), $\gamma(|\Delta S^{[y]}|)$ decreases.

Because $\gamma(\Delta S^{[y]})$ and $\gamma(|\Delta S^{[y]}|)$ are different, rendering a direct comparison between the two hypotheses difficult. Nonetheless, because Hebbian learning predicts that the direction of the motion is towards the attractor located above the centre of mass, the conjunction of relative informativeness with Hebbian learning actually predicts a decrease of $\gamma(\Delta S^{[y]})$ as $\neg S^{[x,y]}$ moves away from the centre.

5.3.2 False blocking

Given that the distribution of motion in the y dimension as a function of $\neg S^{[x,y]}$ seems to be an interesting way to dissociate Rescorla-Wagner learning from Hebbian learning scaled by informativeness, it may also be fruitful to investigate motion in the x dimension. I call this effect "false blocking", as it has important similarities and differences from the standard blocking effect. Similar to the above generalized blocking effect, it is difficult to illustrate generalised false blocking effect as 4D flow that is a function of 2D cues. For intuitive understanding of this effect, we have to consider specific conditions, analogous to Kamin's (1969) work.

The crucial difference between normal blocking and false blocking is that, in false blocking, both of the cues presented have already been associated with outcomes. Consider a compound learning situation, illustrated in Figure 5.5, in which cue A has been associated mostly with outcome 2 and cue B mostly with outcome 1. The question is what happens when the compound cue AB is presented along with outcome 2. According to both Rescorla-Wagner and Hebbian learning theories, the association between both of the cues and outcome 2 should increase and the association with other outcomes should decrease. However, the theories differ on the relative change in associative strength between cue A and outcomes 1 and 3. According to Hebbian learning, the association between cue A and both outcomes 1 and 3 should decrease equally due to decay. This is also true even when we introduce relative informativeness as a scaling factor on learning rate, because relative informativeness is the same for both cues. Rescorla-Wagner learning, on the other hand, predicts that the association between cue A and outcome 1 should decrease more than the association between A and outcome 3. The reasoning behind this is that outcome 1 was more strongly predicted than outcome 3 during the compound learning trial because of its prior association with cue B. But because neither outcome 1 or 3 occurred, there was more PE for outcome 1, which implies more learning.

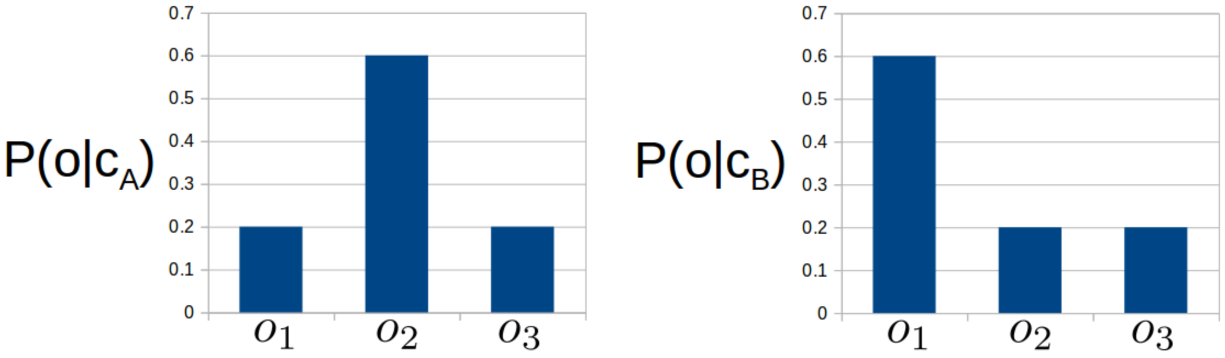


Figure 5.5: Example distributions associated with the cues A and B to outcomes 1-3 prior to the learning trial in false blocking paradigm.

To generalise false blocking a step further, and in terms of simplex flow, consider any distribution associated with cue A while keeping the cue B distribution constant at $P(o|c = B) = [2/3, 1/6, 1/6]$. Figure 5.6 shows that PE-driven learning causes the flow to be "pushed away" from the corner of the triangle that corresponds to the outcome predicted by cue B. On the other hand, it is impossible for non-PE learning to break the symmetry of flow with respect to the y axis even when relative informativeness is introduced as a scaling factor as it is symmetrical across the simplex in respect to the y axis.

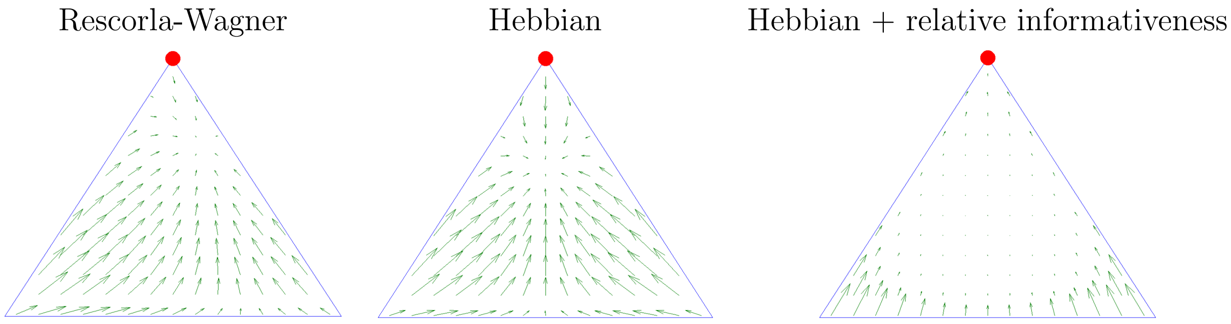


Figure 5.6: Learning dynamics across different learning rules in AB-o_2 compound learning scenario for any distribution associated with cue A and a distribution that predicts o_1 (bottom left corner of triangle) associated with cue B.

The asymmetry of flow in respect to the y axis is a defining feature of PE-driven learning for the false blocking effect. Similarly to our conceptualisation of blocking, this asymmetry can be characterised as the skew γ of the distribution of the x-component of ΔS as a function of $\neg S$. Therefore I treat $\neg S$ as the independent variable, but collapse the flow (dependent variable) into a single dimension by only considering its skewness in the dimension of our interest (y for generalised blocking and x for false blocking). The prediction of Hebbian learning for false blocking is identical to its prediction for the generalized blocking effect.

5.4 Methods

As discussed above, it is informative to examine the skew of the distribution of movement vectors for a cue as a function of the probability distribution for the other cue, formally:

$$\gamma(\Delta S) = f(\neg S) \ . \tag{5.3}$$

First, however we need to find out ΔS. The methods introduced in Chapter 4 were developed to make these variables observable. However, instead of actual measurements, we only obtain likelihood functions of the actual subjective distribution lying on a certain portion of the probability simplex. It is possible to obtain reasonable estimates of S by Maximum Likelihood estimation, and then determine the other variables from those estimates. However it is much more accurate to marginalize the nuisance variables. The distribution of motion vectors is

$$M^{\mathcal{D}}(d, \neg S^{[x,y]}) = \int_t \int_{S_t^{[x]}} \int_{S_t^{[y]}} \int_{S_{t'}^{[x]}} \int_{S_{t'}^{[y]}} \delta\left(S_{t'}^{[\mathcal{D}]} - S_t^{[\mathcal{D}]}, d\right)$$
$$L(S_t^{[x,y]}) L(S_{t'}^{[x,y]}) L(\neg S_t^{[x,y]}) \, dS_{t'}^{[y]} \, dS_{t'}^{[x]} \, dS_t^{[y]} \, dS_t^{[x]} \, dt \tag{5.4}$$

where δ is a Kronecker delta function, $\mathcal{D}$ is a dimension of ΔS of our interest (i.e. y for blocking and x for false blocking) and d is an index in $\mathcal{D}$. Dimension t refers to all pairs of subjective distributions, collapsed across participants, such that t is the state before learning trial and t' is the state after the trial. This function provides us with a three dimensional output, where two dimensions correspond to $\neg S$ and the third one to the ΔS in dimension $\mathcal{D}$.

Because relative informativeness does not affect the directionality of motion, just its magnitude, we have a complementary function:

$$|M|^{\mathcal{D}}(d, \neg S^{[x,y]}) = \int_t \int_{S_t^{[x]}} \int_{S_t^{[y]}} \int_{S_{t'}^{[x]}} \int_{S_{t'}^{[y]}} \delta\left(|S_{t'}^{[\mathcal{D}]} - S_t^{[\mathcal{D}]}|, d\right) \\ L(S_t^{[x,y]}) L(S_{t'}^{[x,y]}) L(\neg S_t^{[x,y]})\, dS_{t'}^{[y]}\, dS_{t'}^{[x]}\, dS_t^{[y]}\, dS_t^{[x]}\, dt \quad (5.5)$$

that describes the L1 norm of the motion.

The metric of interest over $M^{\mathcal{D}}$ and $|M|^{\mathcal{D}}$ is its skewness as a function of $\neg S^{[x,y]}$. However, our data are not data points, but rather a distribution, therefore we need to generalise Pearson's moment coefficient of skewness to probabilistic contexts to get the third standardized moment of any arbitrary distribution A:

$$\gamma(A) = \frac{\int_x (A(x) - \mu(A))^3 \ dx}{\int_x A(x)\ dx}. \quad (5.6)$$

where

$$\mu(A) = \int_x xA(x)\ dx\ . \quad (5.7)$$

Finally we can look at the skewness of the functions $M^{[\mathcal{D}]}$ and $|M|^{[\mathcal{D}]}$ which will be taken along the d dimension (index of $\mathcal{D}$) while the two dimensions corresponding to coordinates on $\neg S$ are conserved.

I used Monte Carlo [MC] methods to remove any possible biases created by the experimental paradigm and to approximate the null distribution during hypothesis testing. I repeat the whole process 100 times with permuted $\neg S_t$ along the t dimension, and then take the mean of the metrics of interest for each point on $\neg S$ across the MC samples and subtract it from the metrics obtained from the human data. The resulting surfaces are fit by a linear regression using the L1 norm of a difference between model and data as the error metric.

To estimate the probability of type I error, I fit the models to each MC sample and then fit a Gaussian distribution to the best-fit parameters. The probability of type I error is the value of the cumulative density function of the fitted distribution at the parameter value best-fitting the real data.

Model fitting

Firstly, I attempt to replicate the classic blocking effect as defined by Kamin (1969). In the present framework, this means testing the difference in skew of the distribution of the

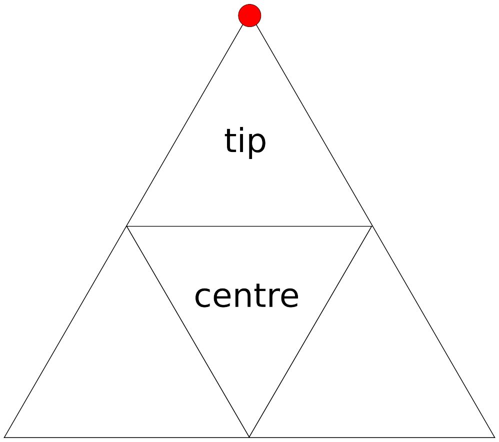


Figure 5.7: The sections of simplex used to do analysis analogous to classic Kamin's (1969) blocking experiment. The red dot corresponds to the outcome that has been presented on a trial.

y-component of motion across S between the centre of mass of $\neg S$ and the tip of $\neg S$ that corresponds to the outcome observed on the trial. This is illustrated in Figure 5.7. The likelihood-weighted mean is taken of both the tip and centre sections of the simplex.

After looking for the classic blocking effect, I tested for blocking and false blocking effects generalized to the entire $\neg S$. As explained above, PE-learning theories predict that, the more an outcome is predicted by $\neg S$, the less motion towards the outcome happens on S. Purely Hebbian learning predicts that the mean skew should be positive but constant across $\neg S$. Finally, Hebbian learning scaled by relative-informativeness predicts that $|\Delta S|$ is inversely related to the entropy of $\neg S$. Based on these predictions I formulated four hypotheses in Table 5.1, expressed as linear regressions of skewness as a function of x or y. These hypotheses are tested by whether there is evidence that the slope of the regression, a, is significantly negative.

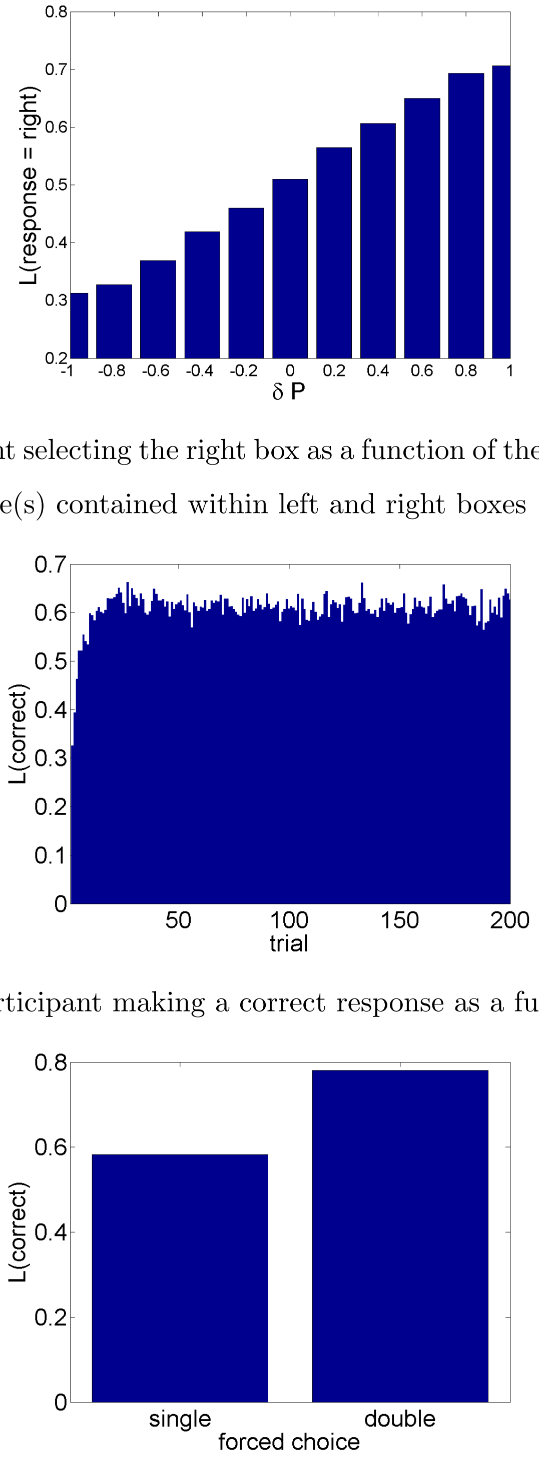
Hypothesis	Effect	Model	Prediction
PE learning	blocking	$\gamma(\Delta S^{[y]}) = ay + c$	$a \in R^{-}$
	false blocking	$\gamma(\Delta S^{[x]}) = ax + c$	$a \in R^{-}$
RI	blocking	$\gamma(\lvert\Delta S^{[y]}\rvert) = a \times H(\neg S^{[x,y]}) + c$	$a \in R^{-}$
	false blocking	$\gamma(\lvert\Delta S^{[x]}\rvert) = a \times H(\neg S^{[x,y]}) + c$	$a \in R^{-}$

Table 5.1: Hypotheses formulated as linear models. Blocking and false blocking effects referred to in this table relate to effects generalized across probability space. RI refers to relative informativeness. H is the entropy function.

As the methods are novel and largely untested, I verified that the predicted pattern of results holds using artificial participants. The artificial participants were programmed to follow various learning algorithms when completing the experiment and their performance was then analysed with the techniques presented here. In all cases, I was able to correctly identify the learning algorithm used by the artificial participants.

Data exploration

Before analysing the dataset for the main effects of interest, I performed basic checks to confirm the validity of the experimental paradigm. Firstly, the likelihood of correct responses increased orderly with the difference between the relative empirical probability of the options in the 2AFC task, as demonstrated by Figure 5.8a. Secondly, participants' performance improved rapidly during first few dozen trials until it reached asymptote of approximately %60 correct responses around trial 25 (see Figure 5.8b). Thirdly, the 2AFC task is significantly easier when two outcomes are present in one of the boxes, Figure 5.8c, as expected.



(a) Likelihood of participant selecting the right box as a function of the difference between empirical probabilities of the outcome(s) contained within left and right boxes (δP).

(b) Likelihood of participant making a correct response as a function of trial number.

(c) Likelihood of participant making a correct response split between trials that had two outcomes as one of the response options (double forced choice) and those that didn't.

Figure 5.8: Basic tests confirming validity of the experimental paradigm.

5.5 Results

Models were tested by comparing the slope parameter a estimated from the data against the null distribution of values estimated by MC methods. In all cases, a is predicted to be less than zero. For the classic blocking effect a difference in mean y-direction skew was compared between the centre of triangle and its tip, with the PE account predicting the tip of triangle to produce larger skews.

There was no evidence for a classic blocking effect ($a = \mu_{centre} - \mu_{tip} = 13.39$, $\mu(a^{MC}) \approx 0$; $\sigma(a^{MC}) = 46.78$, $p(a \sim a^{MC}) = .39$), therefore the null hypothesis of no difference between those two areas (as for example predicted by simple Hebbian learning) was favoured.

The results for the further hypotheses described in Table 5.1 are shown in Table 1. The PE-based models did not fit the data well, with the estimate for false blocking actually being of opposite sign (positive) to the predictions. The Hebbian models scaled by relative informativeness, on the other hand, produced negative values of a whose probability of occurring by chance (from the null distribution) approached 0. These patterns have held up even when the data was split into high and low performing participants.

Hypothesis	Effect	a	$\mu(a^{MC})$	$\sigma(a^{MC})$	$p(a \sim a^{MC})$
PE learning	blocking	-.04	-.01	.12	.39
	false blocking	.26	0	.01	1
RI	blocking	-0.42	0	0.06	0
	false blocking	-.16	0	0.03	0

Table 5.2: Results of hypothesis testing. a is a free parameter fitted and MC refers to distribution of values of a from Monte Carlo sampling. Blocking and false blocking effects referred to in this table relate to effects generalized across probability space. RI refers to relative informativeness.

5.5.1 Post-hoc analysis

I have explored whether the pattern demonstrated in Table 1 holds across the entire dataset. In particular it is of interest whether this pattern holds separately for high-performers and low performers. For this purpose I have identified top and bottom halves of participants according to their performance. However as a large number of participants with low score are those who have left the experiment early (haven't completed many trials) the participant with less than 100 trials completed were excluded from the dataset before the split. The median performance in the reduced dataset is 58.6% correct and the two groups used for post-hoc analysis contain 376 participants each.

5.6 Discussion

The present chapter showed that Hebbian learning scaled by the relative informativeness of cues is better than Rescorla-Wagner (Rescorla et al., 1972) learning based on PE in terms of fitting data from a compound cue learning experiment.

I derived a novel approach for understanding learning as a flow in probability space. This approach allowed me to demonstrate that the location of the attractor in the probability space, and non-linearity of the flow to the attractor, are good characterisations of learning. In the theoretical section, I showed that, while it is impossible to distinguish between the two classes of learning theories in context of single-cue learning, compound learning has a potential to resolve this issue.

Blocking as defined by Kamin (1969), which was introduced in Chapter 1 as key evidence in favour of PE-driven learning, can also be explained by Hebbian learning when scaled by relative informativeness that was introduced in Chapter 2. The independent variable in Kamin's definition of blocking has only two levels, which leaves lot of space for alternative explanations. To avoid this drawback, I generalized the blocking effect to the entire probability space. This definition of blocking provided us with an observable effect on which predictions of the two classes of theories, though correlated, can differ qualitatively. Moreover, the value of the independent variable in blocking experiments is traditionally not observed but rather just assumed based on the conditioning schedule. Here I introduce techniques to observe the independent variable.

The similarity of predictions made by the two classes of learning theories even for generalized blocking led me to derive a novel effect that I called false blocking, for which the predictions of the two classes of learning theories now become distinct.

In the empirical part of the chapter, I described data I have collected from a large online study. Despite the non-conclusiveness of the classic blocking effect as defined by Kamin (1969), I looked for this effect in our data because of recent concerns over its replicability (Maes et al., 2016). This classic effect was not significant in our data, though it was in the predicted direction numerically.

More importantly, I tested blocking and false blocking, generalized to the entire probability space, against the two hypotheses outlined in the theoretical section: PE-driven learning and Hebbian learning scaled by relative informativeness. For both effects, PE-driven learning was not supported (and for false blocking, the pattern of data was numerically opposite to what would be expected).

The fits provided by the relative informativeness model were highly significantly better than chance for both generalized blocking and false blocking. Therefore, I conclude that relative informativeness and Hebbian learning are a better explanation of our data than PE-driven learning (or pure Hebbian learning alone).

While the lack of blocking effect could be due to poor sensitivity of the present analysis, this seems unlikely given that the analysis was sufficiently sensitive to detect the predicted effects of relative informativeness. Moreover the false blocking effect was detected, which opposes the very notion on which blocking is based.

The results presented in this chapter are heavily reliant on methods that I designed for this particular task and introduced in Chapter 4. The methods could not have been tested before because they require a specific type of dataset that was collected for the first time (though I validated the whole analysis pipeline by simulating artificial participants). The reason for this novel and untested method is that comprehensive exploration of the relationship of motion across S as a function of $\neg S$ cannot be done in any other way. Conventional methods only test specific points in the probability space, so that many individual experiments probing different points in the probability space would be required to provide support to the relative informativeness hypothesis presented here. Since the exploratory step is finished, the logical next step is to replicate our results using conventional methods for the points in probability space that provide greatest difference between the hypotheses of interest.

Another avenue of future research would be to use neuroimaging to identify components of brain activity correlated with relative informativeness. This can be done by replication of the experiment in either fMRI or MEG. The brain correlates of relative informativeness correlated could be subsequently used to explain variance in learning and compared in their ability to do so with any brain correlates of PE.

In conclusion, relative informativeness provides a good model of the learning observed in the present data, while there is no evidence for a role of PE.

Chapter 6

Discussion

This thesis aims to evaluate the role of Prediction Error [PE] in human probabilistic associative learning. David Marr's levels of analysis (1982) provide a useful framework to structure this investigation, as the answer to the main question of this thesis might be different at each level.

Chapter 2 started by investigating the associability effect described by Mackintosh (1975), who modified the algorithmic-level Rescorla-Wagner (Rescorla et al., 1972) model of associative learning to explain his experimental findings. I derived a rational model (computational-level) of the task as well as a less computationally demanding algorithmic mechanism that explains the associability effect without using PE, based on scaling Hebbian learning by the informativeness of a cue. In Chapter 3, I examined the evidence for PE in learning on the implementational (neural) level and concluded that the evidence is less robust than often assumed.

Associative learning theories describe a transition from one state of memory to another, but the experiments used to test the theories look only at the combined effect of many learning trials. In Chapter 4, I addressed this problem by introducing a paradigm and an inference method for how to measure participants' beliefs – the *subjective probability distribution* - before and after each learning trial. This method was exploited in Chapter 5 to test more directly whether PE is the driving force behind learning. The results from a large online dataset showed that learning, particularly with compound cues, is better explained by the idea of relative informativeness introduced in Chapter 2 than by a PE-driven learning rule. In the process, I described a way to generalise Kamin's blocking effect, and developed a new type of blocking – false blocking – that is much better suited to distinguishing these types of learning theories.

I review these findings in more detail below, before considering future directions.

6.1 Summary

Chapter 2 analysed a set of experiments by Greve and colleagues (2014) that investigated the effects of cue consistency on learning. After giving a rational (computational-level) description of the effects of cue consistency, I proved that standard and popular algorithmic-level approaches - Hebbian learning and Rescorla-Wagner (PE-driven) learning - cannot explain these data. I then considered Mackintosh's theory to explain such effects, which introduces an associability parameter into the standard Rescorla-Wagner model (Mackintosh, 1975). However, the weaknesses of Mackintosh's theory lie in necessitating the estimation of an extra associability parameter for each cue present in the environment. This parameter has to be estimated with every exposure to a cue and kept in memory, which is computationally demanding. To provide an alternative to this approach, I considered an approximation of the rational description which is based on scaling learning by informativeness (inverse entropy) of a cue. This is easy to compute at any instance simply from the strength of associations with the cue. Moreover, this approach can be extended to the situations of compound learning - scaling the learning by informativeness of a cue relative to the other cues present on a trial – which was used in Chapter 5 to explain the classic blocking effects of Kamin (1969) without the need for PE (i.e. Hebbian learning scaled by relative informativeness).

In Chapter 3, I reviewed the single cell recordings first obtained by Schultz, Dayan and Montague (1997) that show that a proportion of neurons in the ventral midbrain signal PE. This evidence is often used to implicate the role of PE in learning, however there is actually very little research directly linking this PE signal to learning. Demonstrating this link is very difficult because no neurons that compute PE have been found in lower animals so far, thus a clear link between PE computation, synaptic plasticity and behaviour cannot be established in a simple model species in the same way that Eric Kandel (e.g. Kandel, 2001) established the link between synaptic plasticity and behaviour. Demonstrating this link in higher species is problematic because of the high dimensionality of cortical representations, which make linking PE from environment, neural signal and synaptic plasticity and/or behaviour very difficult. To counter this issue, some researchers have resorted to using neuroimaging instead of electrophysiology.

There have been only a few studies that have attempted to identify neuroimaging correlates of PE and link those correlates to behavioural change (e.g. Gläscher et al., 2010; Nassar et al., 2010). They suffer from a number of problems however. Firstly, the PE correlate is being identified at the level of neuronal populations, not individual neurons, therefore they are not directly comparable with the results of Schultz and colleagues (1997). This is a critical issue, because the argument of Schultz and colleagues that those particular neurons signal PE is critically reliant on demonstrating the decrease in spiking with negative PE. However on the level of neural populations, negative PE cannot be observed, since a homeostatically-regulated system will have an approximately identical amount of positive and negative activity. Secondly, the neuroimaging correlate of PE may not actually correspond to the spiking activity of PE-signalling neurons.

The alternative hypothesis I proposed in Chapter 3 is that the PE-correlated signal being observed in fMRI recordings may actually correspond to the energy demands of synaptic plasticity rather than PE-signalling. The energy demands of synaptic plasticity are not well understood, so it is difficult to say whether they can actually cause a signal observable on fMRI. However, the signalling cost can be decreased up to a hundred-fold by appropriate synaptic adjustments (Harris et al., 2012), therefore there should be a significant energy budget assigned to this purpose. These adjustments could well occur in the time-frame of BOLD response lag (Collingridge et al., 2004). Therefore I concluded that there is a distinct possibility that synaptic plasticity might be the source of the PE-correlated signal observed in fMRI.

I further support this argument by a formal proof that, in the context of a task such as that used by Nassar and colleagues (2010), synaptic plasticity driven purely by non-PE learning (Hebb rule) is proportional to PE. The relationship between synaptic plasticity and PE for more complex experimental paradigms such as that used by Gläscher and colleagues

(2010) is relatively complex. Therefore I resorted to numerical simulations of synaptic plasticity in Gläscher's task. These showed that, for a large area of the parameter space, the correlation between synaptic plasticity and PE is very high, even though PE was not used to update the synapses.

Chapter 3 therefore concludes that evidence for PE-driven learning at the level of neural implementation is inconclusive. I therefore returned to the algorithmic level to see if there was any evidence, beyond Greve et al.'s work on cue consistency, that necessitates PE-driven learning. The most striking behavioural effect linked to PE is blocking. However, before addressing blocking, I needed to develop methods to more directly probe the subjective probability distributions that people update during learning experiments.

Conventional methods for comparison of learning theories rely on long experimental schedules at the end of which performance on various items is compared for each participant. The comparison of learning theories is then model-based in the sense that models are fitted across long runs of learning trials and compared to the performance of the participants. This approach is limited by the fact that learning theories specify how beliefs change with exposure to every new data point, which can be lost in cumulative summaries at the end of learning blocks. To address this issue, in Chapter 4 I introduced an experimental paradigm and accompanying statistical methods that allow inference about changes in participants' subjective probability distributions before and after a single learning trial.

The experimental paradigm that I designed for this purpose consists of multiple N-Alternative Forced Choice [NAFC] queries before and after each learning trial, to learn about each subjective probability distribution of interest. No feedback is given, to avoid changing the probability distribution itself.

The statistical method developed to analyse these data utilizes a decision-making model that is parametrized by the participant's sensitivity to the difference in utility between the two choices, and a residual level of randomness in decision-making that is independent of the difference in utilities. These two parameters are first estimated from data assuming that the participants follow optimal statistical inference during learning. Subsequently, each participant's subjective probability distribution is inferred from their responses using the decision-making parameters. To demonstrate the ability of this method to recover the true subjective probability distributions, I successfully validated it on artificial datasets. Moreover I demonstrated that I can obtain unbiased estimates of subjective probability distributions, even when the assumptions of the inference procedure are not met.

This methodology for inferring subjective probability distributions is utilized in Chapter 5 to answer the main question of this thesis: whether learning is driven by PE. Despite the developed and validated system to infer subjective probability distributions before and after a learning trial, addressing this question would still depend on fitting learning models rather than directly demonstrating the effect of PE (or lack thereof). While this model-comparison approach is often used, I wanted to test the role of PE in learning more directly. I therefore derived two effects that should be observable in data if PE is the driving force behind the learning. The first of these effects is a generalization of Kamin's blocking paradigm (1969) to continuous probability space. Kamin's seminal result is that there is less learning when the outcome has already been predicted. By characterising learning as a movement of the subjective probability distribution across the probability simplex induced by the learning trial, I was able to see directly how much learning happens in relation to how much an outcome is predicted. Correlating these quantities generalises Kamin's original approach, which just compares learning between two groups depending on whether or not a group was pre-exposed to the cue-outcome pair. This generalisation had a profound impact, besides increasing statistical power. It revealed that Kamin's original findings can be explained by Hebbian learning scaled by relative informativeness, in addition to PE-driven learning, because the two data points (one per group) do not sufficiently constrain the hypotheses. When the entire space is investigated, the predictions of these two theories differ: Hebbian learning scaled by relative informativeness predicts that learning from one cue in a compound will be relatively small not only when the other cue is highly predictive of the outcome, but also when it is highly predictive about any other outcome.

Besides this generalized blocking effect, I also derived a novel effect that has not, to my knowledge, been described in the literature, yet is a necessary consequence of PE learning. I call this effect *false blocking*, because it has both similarities and differences to Kamin's original blocking. The effect considers compound learning where both cues were pre-exposed, but each was associated with a different outcome. It leverages the prediction of PE learning that when one of the cues in a compound predicts an outcome that has not occurred, the association with this outcome decreases, even for another cue that was present but not predictive of that outcome. While Hebbian learning does not predict any difference in learning dependent on the associations of the other cue in the compound, Hebbian learning scaled by relative informativeness predicts that the learning will be smaller as the predictiveness of the other cue increases, irrespective of what it is predicting.

To test these predictions, I collected a large on-line dataset (approximately 2000 participants completing 5000 compound learning trials), in order to provide sufficient data to approximate type I error probability by Monte Carlo methods (rather than making assumptions about the form of the noise in our data). I statistically tested two sets of hypotheses: 1) whether the generalized blocking and false blocking effects are in line with PE learning where simple Hebbian learning was used as null hypothesis, 2) whether these effects are better explained by Hebbian learning with relative informativeness or by standard Hebbian learning. PE did not provide a significantly better fit to the generalized blocking effect, and more importantly, it produced a very significant fit in the wrong direction for the false blocking effect. Hebbian learning scaled by relative informativeness, on the other hand, provided highly significant fits in the predicted direction for both effects.

This chapter thus clearly demonstrated that, at least in the context of my task, learning is not driven by PE. Instead, I conclude that there is good evidence that learning is driven by a signal similar to relative informativenes.

6.2 Future directions

There were several limitations and issues that arose during this thesis that could be explored in future studies. For example, in Chapter 2, it was proposed that neuroimaging might be able to identify the time at which cue consistency affects brain activity associated with learning, which could potentially tease apart whether these effects scale weight updates at the point of learning, or normalise response selection at retrieval. In Chapter 3, the question was raised about the neural causes of the BOLD signal measured by one neuroimaging method (fMRI). More specifically, the possibility was raised that the BOLD signal (in learning contexts) is dominated by the energy demands required for synaptic plasticity, rather than neural activity per se related to PE. However, the precise energy demands and timescales associated with synaptic plasticity do not appear to be fully known, requiring more basic research in molecular and cell biology. Answering this question would be beneficial not only for the support or refutation of the argument proposed here, but also a valuable addition to the understanding of the nature of signals observed in fMRI in general, potentially confounding results in fMRI research outside of the field of associative learning.

The other proposition made by Chapter 3 is that PE correlates with non-PE driven synaptic plasticity. I questioned the conclusions of Gläscher and colleagues (2010) by suggesting that the PE-correlate they identified might in fact be a correlate of non-PE driven synaptic plasticity. If this suggestion is correct, then their findings that this PE-correlate is related to behavioural change do not support their conclusion that PE drives learning. While I demonstrated that PE and non-PE driven synaptic plasticity are in many cases highly correlated, they are not aligned perfectly. Indeed, if they are sufficiently decorrelated, it may be possible to compare the variance in behaviour explained by the neural correlate of PE with that explained by the neural correlate of non-PE driven synaptic plasticity, and hence support one theory over the other.

Chapter 4 proposed a set of statistical methods that identify the likelihood that a given probability distribution is the probability distribution held by the participant at a certain point in time. This method is critically reliant on the decision-making model that was adopted for this task. The two parameters of this model, β and κ, define the participant's sensitivity to the difference in utility of the alternatives offered on a particular trial and a residual level of randomness that is not sensitive to this utility difference. There are a large number of other parameters that might be potentially worth incorporating into the decision-making model, such as the tendency to persevere with one's past responses. Including these and other parameters may change the ability to accurately recover subjective probability distributions. Secondly, the decision making parameters implemented in the decision-making model are fitted for each individual participant and kept fixed throughout the experiment. However, it is likely that the properties of decision-making vary during the task for reasons such as fatigue, therefore it might be potentially beneficial to allow the parameters of our decision-making model to vary in time.

The results presented in Chapter 5 clearly favour non-PE theories, yet there is still a possibility that these results are driven by our unusual experimental paradigm and/or online data collection. It would be beneficial for the argument made by Chapter 5 to look for the false-blocking effect in a more conventional learning paradigm, such as a simple paired associate task.

Ideally, the false blocking effect would be investigated in conditioning paradigms with non-human species. Here I summarise what such a conditioning version of the false blocking paradigm might look like. As in the classic blocking paradigm, this paradigm would consist of three phases: pre-exposure, exposure and testing, and two groups of subjects: experimental and control. The major difference between blocking and false blocking is that false blocking requires three possible outcomes (types/locations of reward). The pre-exposure phase would consist of repeated exposures to cue A and outcome 1, as well as cue B and outcome 2. However, the subject must be aware that cue C and outcome 3 are both also possible, even though outcome 3 should never be paired with either A or B. During the exposure phase, an AB compound would be presented along with outcome 2. During the testing phase, the experimental group would be tested for the outcome associated with cue B, while the control group would be tested for the outcome associated with cue C. The subject must be prevented from selecting outcome 2: even though it might be preferred, it is not relevant to the predictions. Rather, the prediction from theories driven by PE is that the experimental group will prefer outcome 3 over outcome 1, because outcome 1 was predicted during the exposure phase (by cue A) but did not occur, and therefore the B-1 association should have decreased relative to the B-3 association (whereas in the control group, there should be no difference). According to Hebbian learning on the other hand (whether or not that learning is scaled by relative informativeness) there should be no difference between outcomes 1 and 3 for either groups.

While the model based on Hebbian learning scaled by relative informativeness provided fits to the data that were extremely unlikely to occur by chance (due to the Hebbian model alone), it is important to note that close investigation of the data shows that there are a number of other robust effects that are not predicted by any current theory of associative learning. Using the formalisation of learning as a flow across a probability simplex introduced in Chapter 5, it is possible for the first time to observe learning directly. I therefore suggest that these tools should be used for future exploratory investigations in associative learning - describing the learning dynamics rather than testing hypotheses. The alternative, a conventional approach to test a specific hypothesis on a very limited number of points in the vast space of learning situations, may take very long to converge on the true learning mechanisms.

6.3 Conclusion

This thesis offered a critical view of one of the most popular hypotheses in learning theory: viz, that learning is driven by Prediction Error. In two theoretical chapters, I investigated the evidence used to justify this hypothesis and concluded that conclusive proof of the role of PE in associative learning is yet to be found. In the two subsequent chapters, I offered statistical and experimental tools to infer the subjective probability distributions after each learning trial, which are critical to fully determine the type of learning, and applied them to a novel dataset. By examining compound trials that enable measurement of generalised and false blocking, these data rejected the hypothesis that learning is driven by PE, and instead support my alternative hypothesis that learning is Hebbian, but scaled by the relative informativeness of cues.

References

Abler, B., Walter, H., Erk, S., Kammerer, H., & Spitzer, M. (2006). Prediction error as a linear function of reward probability is coded in human nucleus accumbens. *Neuroimage*, *31*(2), 790–795.

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, New Jersey, US: Psychology Press.

Aubert, A., & Costalat, R. (2002). A model of the coupling between brain electrical activity, metabolism, and hemodynamics: application to the interpretation of functional neuroimaging. *Neuroimage*, *17*(3), 1162–1181.

Bienenstock, E. L., Cooper, L. N., & Munro, P. W. (1982). Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *The Journal of Neuroscience*, *2*(1), 32–48.

Bray, D. (2009). *Wetware: a computer in every living cell*. Yale University Press.

Brehmer, B. (1999). Reasonable decision making in complex environments. *Judgment and decision making: Neo-Brunswikian and process-tracing approaches*, 9–21.

Bridge, D. J., & Paller, K. A. (2012). Neural correlates of reactivation and retrieval-induced distortion. *The Journal of Neuroscience*, *32*(35), 12144–12151.

Collingridge, G. L., Isaac, J. T., & Wang, Y. T. (2004). Receptor trafficking and synaptic plasticity. *Nature Reviews Neuroscience*, *5*(12), 952–962.

Courville, A. C., Daw, N. D., & Touretzky, D. S. (2006). Bayesian theories of conditioning in a changing world. *Trends in cognitive sciences*, *10*(7), 294–300.

D'Ardenne, K., McClure, S. M., Nystrom, L. E., & Cohen, J. D. (2008). Bold responses reflecting dopaminergic signals in the human ventral tegmental area. *Science*, *319*(5867), 1264–1267.

Daw, N. D., Courville, A. C., & Dayan, P. (2008). Semi-rational models of conditioning: The case of trial order. *The probabilistic mind*, 431–452.

Denker, J., & Lecun, Y. (1991). Transforming neural-net output levels to probability distributions. In *Advances in neural information processing systems 3* (pp. 853–859). Morgan Kaufmann.

Doucet, A., De Freitas, N., & Gordon, N. (2001). An introduction to sequential monte carlo methods. In *Sequential monte carlo methods in practice* (pp. 3–14). Springer.

Fletcher, P. C., Anderson, J., Shanks, D., Honey, R., Carpenter, T. A., Donovan, T., ... Bullmore, E. T. (2001). Responses of human frontal cortex to surprising events are predicted by formal associative learning theory. *Nature neuroscience*, *4*(10), 1043.

Franco-Watkins, A., Derks, P., & Dougherty, M. (2003). Reasoning in the monty hall problem: Examining choice behaviour and probability judgements. *Thinking & Reasoning*, *9*(1), 67–90.

Gelman, A., & Shalizi, C. R. (2013). Philosophy and the practice of bayesian statistics. *British Journal of Mathematical and Statistical Psychology*, *66*(1), 8–38.

Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, *66*(4), 585–595.

Greve, A., Cooper, E., Anderson, M., & Henson, R. (2014). Does prediction error drive one-shot declarative learning? *unpublished*, *1*(4), 1–2.

Harris, J. J., Jolivet, R., & Attwell, D. (2012). Synaptic energy use and supply. *Neuron*, *75*(5), 762–777.

Hebb, D. O. (1952). *The organisation of behaviour: a neuropsychological theory.* Wiley.

Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? on the explanatory status and theoretical contributions of bayesian models of cognition. *Behavioral and Brain Sciences*, *34*(04), 169–188.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological review*, *80*(4), 237.

Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. *Punishment and aversive behavior*, 279–296.

Kandel, E. R. (2001). The molecular biology of memory storage: a dialogue between genes and synapses. *Science*, *294*(5544), 1030–1038.

Kruschke, J. K. (2006). Locally bayesian learning with applications to retrospective revaluation and highlighting. *Psychological review*, *113*(4), 677.

LePelley, M., & McLaren, I. (2004). Associative history affects the associative change undergone by both presented and absent cues in human causal learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *30*(1), 67-73.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, *82*(4), 276-298.

Maes, E., Boddez, Y., Alfei, J. M., Krypotos, A.-M., Hooge, R., De Houwer, J., & Beckers, T. (2016). The elusive nature of the blocking effect: 15 failures to replicate. *Journal of Experimental Psychology: General*, *145*(9), e49.

Marr, D., & Vision, A. (1982). A computational investigation into the human representation and processing of visual information. *WH San Francisco: Freeman and Company*, *1*(2).

Mazur, J. E., & Wagner, A. R. (1982). An episodic model of associative learning. *Quantitative analyses of behavior: Acquisition*, *3*, 3–39.

McClure, S. M., Berns, G. S., & Montague, P. R. (2003). Temporal prediction errors in a passive learning task activate human striatum. *Neuron*, *38*(2), 339–346.

McGuire, J. T., Nassar, M. R., Gold, J. I., & Kable, J. W. (2014). Functionally dissociable influences on learning rate in a dynamic environment. *Neuron*, *84*(4), 870–881.

Minka, T. (2000). *Estimating a dirichlet distribution.* Technical report, MIT.

Nassar, M. R., Wilson, R. C., Heasly, B., & Gold, J. I. (2010). An approximately bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *The Journal of Neuroscience*, *30*(37), 12366–12378.

Oja, E. (1982). Simplified neuron model as a principal component analyzer. *Journal of mathematical biology*, *15*(3), 267–273.

Pearce, J. M., & Hall, G. (1980). A model for pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological review*, *87*(6), 532.

Pearce, J. M., & Mackintosh, N. J. (2010). Two theories of attention: A review and a possible integration. *Attention and associative learning: From brain to behaviour*, 11–39.

Pérez-Otaño, I., & Ehlers, M. D. (2005). Homeostatic plasticity and nmda receptor trafficking. *Trends in neurosciences*, *28*(5), 229–238.

Rescorla, R. A., Wagner, A. R., et al. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning ii: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Ruck, D. W., Rogers, S. K., Kabrisky, M., Oxley, M. E., & Suter, B. W. (1990). The multilayer perceptron as an approximation to a bayes optimal discriminant function. *Neural Networks, IEEE Transactions on*, *1*(4), 296–298.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Ii. the contextual enhancement effect and some tests and extensions of the model. *Psychological review*, *89*(1), 60.

Rumelhart, D. E., McClelland, J. L., Group, P. R., et al. (1988). *Parallel distributed processing* (Vol. 1). IEEE.

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

Sethna, J. P. (2006). Entropy, order parameters, and complexity. *Statistical Mechanics, Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, NY*, 14853–2501.

Shanks, D. R. (1995). *The psychology of associative learning* (Vol. 13). Cambridge University Press.

Shannon, C. E., & Weaver, W. (1949). The mathematical theory of information.

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron*, *79*(2), 217–240.

Shepard, R. N. (1958). Stimulus and response generalization: Tests of a model relating generalization to distance in psychological space. *Journal of Experimental Psychology*, *55*(6), 509.

Siegel, S., & Allan, L. G. (1996). The widespread influence of the rescorla-wagner model.

In (Vol. 3, pp. 314–321). Springer.

Simon, H. A. (1972). Theories of bounded rationality. *Decision and organization*, *1*(1), 161–176.

Thorndike, E. L. (1927). The law of effect. *The American Journal of Psychology*, *39*(1/4), 212–222.

Tolman, E. C. (1932). *Purposive behavior in animals and men.* Univ of California Press.

Toyoizumi, T., Kaneko, M., Stryker, M. P., & Miller, K. D. (2014). Modeling the dynamic interaction of hebbian and homeostatic plasticity. *Neuron*, *84*(2), 497–510.

White, H. (1989). Learning in artificial neural networks: A statistical perspective. *Neural computation*, *1*(4), 425–464.

WIDROW, B., & HOFF, M. E. (1960). Adaptive switching circuits.

Appendix

A Derivation of likelihood of γ

The following derivation shows how a concentration hyperparameter, γ, be obtained from data. This derivation specifies the likelihood for a group of cues (or a specific context), but can equally well be used to find out concentration of a single cue.

$$L(\{O_t\}, \gamma | \{C_t\}) = P(\{O_t\} | \{C_t\}, \gamma) = \tag{1}$$

using summation rule

$$= \int P(\{O_t\}, \underline{\underline{w}} | \{C_t\}, \gamma) \, \mathrm{d}w \tag{2}$$

and product rule

$$= \int P(\{O_t\} | \{C_t\}, \underline{\underline{w}}) \; P(\underline{\underline{w}} | \gamma) \, \mathrm{d}w \tag{3}$$

Specifying for matrix columns/rows across T trials

$$= \int \prod_{t=1}^{T} P(O_t | \underline{w}^{(C_t)}) \; \prod_{i=1}^{N} P(\underline{w}^{(i)} | \gamma) \, \mathrm{d}\underline{\underline{w}} \tag{4}$$

$$\prod_{t=1}^{T} P(O_t | \underline{w}^{(C_t)}) \equiv \prod_{i=1}^{N} \prod_{t:\, C_t = i} P(O_t | \underline{w}^{(i)}) \tag{5}$$

or also across N possible outcomes

$$= \int \cdots \int \prod_{i=1}^{N} \left(\prod_{t:\, C_t = i} P(O_t | \underline{w}^{(i)}) \right) \; P(\underline{w}^{(i)} | \gamma) \, \mathrm{d}\underline{w}^{(1)} \cdots \mathrm{d}\underline{w}^{(N)} \tag{6}$$

which is

$$= \prod_{i=1}^{N} \left[\int \left(\prod_{t:\, C_t = i} P(O_t | \underline{w}^{(i)}) \right) \; P(\underline{w}^{(i)} | \gamma) \, \mathrm{d}\underline{w}^{(i)} \right] \tag{7}$$

given a multinomial distribution (which results from dropping the normalising factor in the Dirichlet distribution that is no longer needed as the distributions is already normalised)

$$P(o|\underline{w}) = \prod_j w_j^{\delta_{jo}} \tag{8}$$

and a Dirichlet distribution normalized by inverse of multinomial β function

$$P(\underline{w}|\gamma) = \frac{1}{\beta(\gamma)} \prod_j w_j^{\gamma - 1} \tag{9}$$

and the fact that

$$\prod_{t\colon C_t=i} P(O_t|\underline{w}^{(i)}) = P(\{O_t\}_{t\colon C_t=i}|\underline{w}^{(i)}) \tag{10}$$

substituting into 7

$$= \prod_{i=1}^{N} \int \left(\prod_{t:C_t=i} \prod_j {w_j^{(i)}}^{\delta_{jO_t}} \right) \frac{1}{\beta(\gamma)} \prod_j {w_j^{(i)}}^{\gamma-1} \, \mathrm{d}\underline{w}^{(i)} \tag{11}$$

$$= \frac{1}{\beta(\gamma)} \prod_{i=1}^{N} \int \prod_j {w_j^{(i)}}^{\gamma-1+\sum_{t:C_t=i} \delta_{jO_t}} \, \mathrm{d}\underline{w}^{(i)} \tag{12}$$

since

$$n_k^{(i)} = \sum_{t:C_t=i} \delta_{ko_t} \ , \tag{13}$$

where δ stands for Kronecker delta function,

$$= \frac{1}{\beta(\gamma)} \prod_{i=1}^{N} \int \prod_j {w_j^{(i)}}^{\gamma-1+n_k^{(i)}} \, \mathrm{d}\underline{w}^{(i)} \tag{14}$$

we define

$$\gamma' = \gamma + n_k^{(i)} \tag{15}$$

substituting 15 into 14

$$= \frac{1}{\beta(\gamma)} \prod_{i=1}^{N} \int \prod_j {w_j^{(i)}}^{\gamma'-1} \, \mathrm{d}\underline{w}^{(i)} \tag{16}$$

since the beta function is a normalising factor for the Dirichlet distribution, which is a probability distribution, and therefore W must lie on a probability simplex

$$\int \frac{1}{\beta(\gamma)} \prod_j w_j^{\gamma-1} \, \mathrm{d}w = 1 \ \longrightarrow \ \beta(\gamma) = \int \prod_j w_j^{\gamma-1} \, \mathrm{d}w \tag{17}$$

based on which we define

$$\beta'(\gamma) = \int \prod_i w_i^{\gamma'-1} \,\mathrm{d}w = \beta(\gamma') \tag{18}$$

which, given 18 and 15, results in

$$= \frac{1}{\beta(\gamma)} \prod_{i=1}^{N} \beta(\underline{\gamma} + \underline{n}^{(i)}) \tag{19}$$

B Decision-making parameters optimisation

The decision-making model from chapter 4.0.2 needs to be optimized for every participant as I allow for individual differences in decision-making. The subjective probability distributions used in this phase are the relative frequencies of outcomes for each cue. As the optimization initialized from a single point can lead to severely sub-optimal local minima I initialize from several values provided below. After obtaining best fitting decision-making parameters, the likelihood across subjective probability distributions are calculated using the generative model.

variable	description	initialization
β	decision-making temperature	1, .5 then 13 times $U(0, 10)$
κ	residual randomness indecision making	.94, .94 then 13 times $U(.5, 1)$

C High and low performing participants

To further explore the dataset, I split the participants that completed at least 100 trials into high and low performing halves. The trial number criterion was necessary because most participants who did not score highly were those that completed only very few trials, therefore the resulting trial-count would differ widely between groups.

No actual hypothesis testing was performed as I do not have any hypotheses about what the patterns of performance should be. Adapting the MC results from Chapter 5 was necessitated by very high computational demands of the sampling. While this approach is not entirely adequate, the very large difference between MC samples and best fit parameters shown in Table 1, together with the very large sample size, suggests that the effects discussed in chapter 5 exist even for these two groups separately.

Hypothesis	Effect	a^{high}	a^{low}	$\mu(a^{MC})$	$\sigma(a^{MC})$
PE learning	blocking	-0.03	-0.01	-0.01	0.12
	false blocking	0.25	0.17	0	0.01
RI	blocking	-0.52	-0.41	0	0.06
	false blocking -0.10	-0.16	0	0.03	

Table 1: Results of hypothesis testing for a high (a^{high}) and low (a^{low}) performing group of participants. a is a free parameter fitted and MC refers to distribution of values of a from Monte Carlo sampling that was performed over the entire dataset.