A LATENT CAUSE THEORY OF CLASSICAL CONDITIONING

Aaron C. Courville

Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213

June 2006

Submitted in partial fulfilment of
the requirements for the degree of
Doctor of Philosophy

Thesis Committee:
David S. Touretzky, Chair
James L. McClelland
Geoff Gordon
Peter Dayan, UCL (Gatsby)

ABSTRACT

Classical conditioning experiments probe what animals learn about their environment. This thesis presents an exploration of the latent cause theory of classical conditioning. According to the theory, animals assume that events within their environment are attributable to a latent cause. Learning is interpreted as an attempt to recover the generative model that gave rise to these observed events. In this thesis, the latent cause theory is applied to three distinct areas of classical conditioning, in each case offering a novel account of empirical phenomena.

In the first instance, the effects of inference over an uncertain latent cause model *structure* are explored. A key property of Bayesian structural inference is the tradeoff between the model complexity and data fidelity. Recognizing the equivalence between this tradeoff and the tradeoff between generalization and discrimination found in configural conditioning suggests a statistical account of these phenomena. By considering model simulations of a number of conditioning paradigms (including some not previously viewed as "configural"), behavioral signs that animals employ model complexity tradeoffs are revealed.

Next, the consequence of merging latent variable theory with a generative model of change are studied. A model of change describes how the parameters and structure of the latent cause model evolve over time. The resulting non-stationary latent cause model offers a novel perspective on the factors that influence animal judgments regarding environmental change. In particular, the model correctly predicts that the introduction of an unexpected stimulus can spur fast learning.

Finally a version of a latent cause model is developed that explicitly encodes a latent *timeline* to which observed stimuli and reinforcements are associated, preserving their temporal order. In this context, the latent cause model is equivalent to a hidden Markov model. This model is able to account for a theoretically challenging set of experiments which collectively suggest that animals encode the temporal relationships among stimuli and use this representation to predict impending reinforcement.

This thesis offers a unified theoretical framework for classical conditioning. It uses state of the art statistical methods to explore a novel theoretical account of a wide range of empirical phenomena, many of which have otherwise resisted a computational explanation.

ACKNOWLEDGEMENTS

First I would like to thank my advisor, Dave Touretzky, for his support and for giving me the freedom to pursue the line of research that ultimately lead to this thesis. I would also like to thank my thesis committee: Jay McClelland, Peter Dayan and Geoff Gordon. Each one has played an extremely important role in the development of the thesis and the research it contains. I deeply appreciated the support they have given me throughout this process. I am particularly indebted to Geoff Gordon, who's guidance and insights have significantly shaped the research described in this thesis.

I would like to acknowledge my friend and collaborator Nathaniel Daw. The best work found in these pages came about as a direct result of his input. I have thoroughly enjoyed working with him and I hope we will have the opportunity to work together in the future.

Finally, I would like to express my gratitude to my family for their support and patience throughout the preparation of this thesis. Sweetie, we made it.

TABLE OF CONTENTS

Bibliography . . . 147

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

Since the pioneering experiments of Pavlov, the empirical science of classical conditioning has produced a vast catalog of data. Today, precise experimental methods are used to study a wide range of phenomena in species ranging from invertebrates to humans. Alongside these empirical advancements, learning theorists have pushed forward our understanding of the learning process. With this thesis I endeavor to continue the efforts of those who seek an answer to the question: *what is learning?*

1.1. What is learning?

A simple question with no simple answer. The word *learning* is used to describe phenomena ranging from simple classical conditioning to human problem solving and language acquisition. The relative simplicity of classical conditioning offers a unique opportunity to directly probe the phenomenon of learning from a number of perspectives.

Within the conditioning literature there appears to be broad agreement with regard to an operational definition of learning: *a relatively permanent change of behavior resulting from experience* (Thorpe, 1956; Mowrer & Klein, 1989). The specification that we are concerned with relatively permanent change excludes changes in behavior resulting from changes in motivational states (such as hunger or thirst) or resulting from sensory adaption or fatigue (Hilgard & Marquis, 1961). Specifying changes due to experience excludes other sources of behavioral change such as development. While this kind of operational definition offers a fine description of how learning is measured in the laboratory, it does little to help us understand: *what is learning*? Through the course of this thesis we will encounter many theories of conditioning, each with its own unique answer to this question.

There are some who believe that to seek an understanding of the behavior of animals beyond an operational definition is nonsensical. They believe that the process underlying the observable manifestation of learning is not directly observable and can never be

uniquely inferred from the behavioral data. This school of thought is known as Behaviorism. Its most notable adherents include Thorndike, Hull and Skinner. Though it once dominated the American psychological landscape, Behaviorism is now largely discredited, a victim of a number of key empirical and theoretical developments that emerged almost simultaneously in the mid-1960's.

The modern perspective on classical conditioning is largely dominated by associative learning theory, and in particular the model of Rescorla and Wagner (1972), both of which emerged with the fall of Behaviorism. Although associative learning theory shares many characteristics with its predecessor, including an elementary stimulus-response style formalism, associative learning theories are more inclined to a mechanistic explanation of behavior. They envision learning as the formation of associations or links between internal representations of stimuli and reinforcement. While associative learning theory has contributed significantly to our present conceptualization of classical conditioning, in recent years theoretical development within the associative learning framework has not kept pace with the rate of empirical discovery emerging from ever more sophisticated experiments.

In this thesis I pursue a normative strategy to modeling animal learning and behavior within the context of classical conditioning. Normative analysis in based on the supposition that the behavior of organisms has been optimized through the process of evolution (Anderson, 1990). It has previously been applied to explain a wide range of human cognitive processes including decision making, similarity judgments (Tenenbaum & Griffiths, 2001a), causal induction (Tenenbaum & Griffiths, 2001b, 2003), medical diagnosis, etc. The normative perspective has also been used to explain conditioning phenomena (Dayan & Kakade, 2001) and other animal learning phenomena (Gallistel, 1990).

A normative approach to conditioning offers an important advantage over the more mechanistic associative learning theories. The assumption of optimality implies an important consequence: when there exists a unique optimal solution to a learning problem, the subjects will approximate this solution. Thus given any conditioning paradigm the role of the theorist simplifies to understanding what constitutes optimal behavior in that context. This kind of directed line of inquiry dramatically reduces the space of functions that might possibly emulate animal behavior. As I intend to show, this permits substantial progress in the development of models that begin to approach the level of sophistication that animals routinely demonstrate in modern conditioning experiments.

Normative analysis approaches behavior at a level of considerable abstraction. Here we are not looking to gain insight into the neural organization supporting learning, nor are we looking to describe the mechanisms that give rise to the computation. Rather we seek

to identify the nature of the computations that organisms face. We seek to understand *why animals behave as they do*.

1.2. A latent cause model of conditioning

The central tenet of the normative perspective is that the behavior of the system under consideration is the result of a process of optimization. However it is left to the modeler to specify precisely what this means—what is being optimized? In economics, rational behavior is taken to mean behavior that optimizes a personal utility function. In the context of classical conditioning, the normative standard is usually taken to be statistical, with rational behavior being defined consistent with a Bayesian treatment of the training data. It remains to specify the prior beliefs, including the modeling framework under consideration and the function describing the likelihood of the observations.

Often in models of conditioning, the goal of conditioning is to recover the conditional probability distribution over reinforcement given the observed pattern of stimuli. In this thesis, I take the perspective that animals are attempting to learn the "causal texture" of their environment.[1] Specifically that animals are attempting to recover the causal model that gives rise to all stimuli including, but not limited to, reinforcement.

The perspective that animals are learning about the causal structure of the environment is hardly new. Learning theorists have generally held to this view since the time of Tolman (in the 1930s). However, my conceptualization of cause is quite different than that commonly found in the conditioning literature. Typically direct causal links are assumed between the stimuli and the reinforcer. According to this view, Pavlov's dogs would interpret the sound of the metronome as the cause of the ensuing food delivery. This interpretation of the causal structure of classical conditioning lies at the heart of the associative framework and has stood as dogma among learning theorists for the past 40 years. Yet, is this really a valid interpretation of Pavlov's experimental procedure? Did the sound of the metronome *cause* the food to appear? What in the animal's experience would indicate that a sound is the cause of anything? Sound is the quintessential effect.[2] The actual causal structure almost certainly involves a third unobserved entity—the experimenter or possibly some sort of experimental apparatus—that is ultimately responsible for causing both the metronome to sound and the food to appear. Indeed, the causal reality of the vast

[1] Tolman first used the term "causal texture" (Tolman & Brunswick, 1935) to articulate his vision of causal learning. He credited S. C. Pepper as the originator of the term.

[2] Sound may *cause* the behavior of other animals in the environment to change, but these kinds of interactions are really beyond the scope of the causal relationships I consider.

majority of classical conditioning experiments may be characterized similarly, with an unobserved (or latent) cause giving rise to perhaps a pattern of stimuli together with some sort of reinforcing signal. I suggest not only that animals are capable of this kind of causal inference, but that in most cases, this is the prevailing interpretation of events.

The statistical calculations I here ascribe to animals are difficult, and in most cases computable only approximately. I intend no claim that animals are employing the same approximation schemes to answer these questions as I employ in my simulations. My purpose here is instead to shed light, at a more abstract and formal level, on the reasons *why* animals might sensibly make the predictions measured in classical conditioning experiments. Models more closely linked to plausible *implementations* (such as those of Pearce (1994) and of Rescorla and Wagner (1972)) can then be viewed as candidates for how computations analogous to aspects of our model might be approximated.

1.3. Thesis overview

This thesis is an exploration of the ramifications of this perspective on classical conditioning. The effects touch on a wide range of experiments and theories of learning. In Chapter 2, I review many of the key ideas that have influenced the development of the theory of classical conditioning over the past century, culminating in the theory of Rescorla and Wagner—currently the most successful model of conditioning. The following sections consider how to advance the theoretical understanding of classical conditioning beyond Rescorla-Wagner theory. To this end, rational analysis is introduced as a framework that has already proven useful in explaining conditioning experiments. In the last section of this chapter I develop the basic theoretical framework of the latent cause model of classical conditioning.

In Chapter 3, latent cause theory is considered in the context of a normative perspective on the factors at play in the tension between generalization and discrimination. I illustrate how generalization and discrimination represent two poles of a fundamental tradeoff that any learning system must negotiate. Using the latent cause framework, I demonstrate how Bayesian inference over the model structure is broadly consistent with animal behavior. In considering configural conditioning experiments—a set of empirical investigations performed with the specific goal of illuminating the relationship between discrimination and generalization—I demonstrate that the latent cause account is at least as successful as previous accounts in explaining this well studied collection of data. However, the latent cause theory also offers explanations for a range of other experiments for which the existing theories of configural conditioning offer no account.

In applying a Bayesian model inference formalism to the latent cause model in Chapter 3, trials are assumed to be independent, identically distributed (IID) draws from an unknown but static distribution. This assumption turns out to be a fairly poor approximation of the reality of conditioning experiments. Many conditioning phenomena are critically dependent on the order of the trials, and the IID assumption effectively prevents the model from capturing trial order effects. In Chapter 4, I remove the IID assumption by introducing a *model of change*. The model of change describes how the world changes in terms of the evolution of the latent cause generative model structure and parameters. Following the work of Dayan and Kakade (2001) with their Kalman filter model of conditioning, I define a generative model of parameter and structural change.

The empirical data bearing on the issue of how animals respond to changing events are considered. The well known model of Pearce and Hall (1980) successfully accounts for a number of these experiments with the claim that surprising reinforcement causes an increase in the learning rate. While the empirical data does support such a claim, considerations of a wider set of experimental results suggest that any surprising event, and not simply those having to do with reinforcement, results in an increased learning rate. Because the Pearce-Hall model is discriminative—learning only about the reinforcement predictions in the presence of a pattern of stimuli—it has no ability to detect surprising non-reinforcement events. The latent cause model, on the other hand, learns to predict all stimuli and thus can detect surprising non-reinforcement events. I show that incorporating a model of change within the latent cause framework offers a more complete accounting of phenomena relating to surprising events.

Chapter 5 concerns the application of the latent cause framework to modeling how animals interpret the temporal structure of a trial. In a series of fascinating experiments, Ralph R. Miller and colleagues demonstrate that animals build rich memory representations of the temporal order of stimulus presentations. The experiments are all versions of the second-order conditioning paradigm where, in a first phase of learning, a stimulus is trained as a conditioned excitor (repeatedly paired with reinforcement until the presence of the stimulus evokes a response, like Pavlov's salivating dog). In a second phase of training a second, behaviorally neutral stimulus is paired with the conditioned excitor. Subsequent testing reveals that this second stimulus also comes to elicit a response. Collectively, the experiments of Miller et al. suggest that this transfer of excitatory responding is a consequence of the integration of the memory representations of the two trial types. These experiments challenge existing theories of conditioning, including those specifically formulated to encode temporal information within a trial.

The latent cause framework is applied to this set of experiments by implementing a latent chain of states. The chain of states defines the temporal structure of a trial, with observations occurring at a particular moment becoming associated with the latent state representing that moment in the trial. In this context, with a latent chain structure, the model is simply an example of a hidden Markov model (HMM). When the two trial types of the second-order conditioning paradigm are presented to the model, the presence of the (first-order) excitor in both trials causes the state inference procedure to infer that both trials originate from the same latent chain of states. With the trial types merged on a single chain, inference over the hidden state leads to a prediction of imminent reinforcement with the presentation of the second stimulus in correspondence with the empirical evidence.

Finally, in Chapter 6, I briefly discuss some further implications of the present theoretical framework. I suggest that the latent cause framework offers a parsimonious account of a body of data indicating that stimuli interfere with each other irrespective of whether or not they are trained together. While such an observation is vexing for most existing models of conditioning, it is a natural, even unavoidable consequence of Bayesian model inference over the space of latent cause models. In this chapter, I also suggest directions for future work.

1.4. Contributions

There are four significant contributions of the work presented in this thesis. First, I introduce a new modeling framework: the latent cause theory of classical conditioning. The theory offers a conceptually parsimonious explanation for a wide array of conditioning phenomena, many of which have thus far resisted a computational account. The model is grounded in the intuitive notion that animals are simply attempting to recover the causal structure of their environment. Second, I explore the role of Bayesian model inference as a determinant of animal behavior. Leveraging this perspective, one can understand configural conditioning phenomena as the behavioral consequence of animal's actively negotiating the tradeoff between representational complexity and data fidelity. Third, by building on the generative models of change of Dayan and Kakade (2001) and of Dayan and Yu (2003), I demonstrate how the model of Pearce-Hall may be generalized to incorporate a broader and empirically better supported definition of surprise. This also contributes as one of the first cognitive models that incorporate inference over model structure and an assumption of a non-stationary environment. My final contribution of this thesis is to demonstrate the power of rational analysis and in particular the generative modeling formalism as a general approach to modeling learning phenomena. By employing these tools I am able to

add and remove simplifying assumptions with relative ease and exploit well-established practical algorithms for approximate inference.

CHAPTER 2

Perspectives on conditioning

This chapter is concerned with the development of the experimental and theoretical underpinnings of classical conditioning. The theory developed throughout this thesis contacts data from a broad spectrum of the empirical conditioning literature. These data are in turn addressed by a host of existing models. The study of classical conditioning phenomena is a mature field of scientific endeavor and is much too broad to be given a comprehensive treatment in a single chapter. Therefore, in my attempt to describe the previous empirical and theoretical contributions to the field, I have chosen to include those works that pertain most directly to the theory advanced in this thesis as well as those that I regard as significant determinants in the development of the field.

In Section 2.1 a brief overview of experimental conditioning phenomena is provided. In particular, I highlight the classic empirical results that have shaped our theoretical understanding of learning. Then in section 2.2 I present the levels of analysis framework of Marr (1982) and Anderson (1990) for modeling complex systems and cognition in particular. These levels are: 1) the computational level, 2) the algorithmic level and 3) the implementational level. Then, using these levels of analysis as a basic framework, I describe existing theories of conditioning. In Section 2.3, the central computational issues that have arisen in the development of classical conditioning theory are discussed. Section 2.4 explores some of the most influential theories of classical conditioning at an algorithmic level of analysis, along with their account of the empirical data. Section 2.5 describes models of learning that are based on considerations of the physiology of the brain. Then in Section 2.6, I review recent statistical approaches to modeling classical conditioning phenomena as well as other cognitive processes. Finally in Section 2.7, I introduce the basic modeling framework to be explored in the remaining chapters of this thesis. The model described here is another example of a statistical model of conditioning and as such builds on the models of Section 2.6.

2.1. The empirical phenomena of classical conditioning

In this section, I briefly introduce the key empirical phenomena that are today referred to collectively as classical conditioning phenomena. In particular, I focus on phenomena that highlight the various factors that are known to affect learning.

2.1.1. Acquisition and extinction

The empirical science of classical conditioning begins in earnest with the well known work of Pavlov (1927). Pavlov, a renowned physiologist, was awarded the Nobel Prize for his studies of the digestive glands of dogs. At some point in his investigations, he noticed that sham feedings (in which food never reaches the stomach and instead passes through an esophageal fistula) were nonetheless able to initiate gastric secretions in the stomach. Pavlov believed that the sensation of tasting the food and the act of swallowing had an influence on the operation of the stomach, calling these influences "psychic secretions" (Hilgard & Marquis, 1961). Later, Pavlov switched to a much more accessible experimental paradigm: salivary secretions. He demonstrated that the similar psychic secretions of saliva could be conditioned to occur at the sound of a metronome if the dog experienced repeated pairing of the ticking of the metronome and the presentation of food. It was with this experimental paradigm that Pavlov made his most significant contributions to psychology.

Pavlov's original discovery—that the pairing of a stimulus with reinforcement will eventually lead to characteristic responding to the stimulus—is known as *acquisition*. It forms the basis of all investigations of classical conditioning. Using the notation employed throughout this thesis, acquisition may be expressed as repeated trials of $A+$: pairings of an initially neutral *conditioned stimulus* A (such as a tone or light) together with an *unconditioned stimulus* or reinforcement $+$. The reinforcement may either be appetitive as was Pavlov's food stimulus or aversive such as an electric shock to the foot.

In addition to demonstrating the basic phenomenon of acquisition, Pavlov (1927) also established many other canonical conditioning phenomena, including *extinction*. Extinction is the loss of a conditioned response to a stimulus that occurs when the stimulus is repeatedly presented in the absence of reinforcement.

2.1.2. Latent inhibition

While extinction demonstrated that exposure to an isolated stimulus after acquisition training leads to a loss of conditioned responding, it is also true that exposure to a stimulus *before* acquisition training significantly slows learning. The phenomenon is known as latent

inhibition (or alternatively as the stimulus pre-exposure effect) and typically arises from a two phase training procedure. In the first phase, a subject experiences an extended phase of exposure to an unpaired stimulus $A-$; then in the second phase, the same stimulus is paired with reinforcement, $A+$. Subjects who experience both phases of training are slower to acquire an association between the stimulus and reinforcement (requiring more trials to achieve an acquisition criterion) than subjects that did not experience unpaired presentations of A.

2.1.3. Cue competition effects

The association between a stimulus and reinforcement may similarly be dependent upon the presence of other stimuli during training. The so-called cue competition phenomena demonstrate the interaction between multiple stimuli and their respective relationships with reinforcement. Overshadowing, blocking and the relative validity effect are all examples of cue competition effects.

Overshadowing: Kamin (1967) observed that if two stimuli, A and B, (say for example: tone and light) are presented together as a compound followed by reinforcement: $AB+$, then frequently the subject will come to respond almost completely to one of the stimuli and not the other. Typically the more salient stimulus *overshadows* the less salient stimulus, meaning the more salient stimulus elicits a response while the less salient stimulus does not.

Blocking: The classical procedure to establish blocking consists of two training phases. In the first phase, a stimulus A is paired with reinforcement: $A+$. In the second phase, A is paired with another stimulus, B, and the reinforcer: $AB+$. As originally observed by Kamin (1969), testing with B reveals a significantly lower response than in control subjects missing the initial $A+$ training. In this case, A is said to have *blocked* B from acquiring a significant association with reinforcement.

Relative Validity Effect: Wagner, Haberlandt, and Price (1968) demonstrated another important example of the interaction between stimuli trained together. In this experiment, a stimulus X is paired with one of two stimuli, A or B. One group of animals receives trials of $AX+$ (reinforced) and $BX-$ (not reinforced), while another group received trials of $AX\pm$ and $BX\pm$ (both partially reinforced). On testing, the first group came to elicit a greater response to A than to B or X. In contrast, subjects in the second group came to elicit a greater response to X than to either A or B. Table 2.1 summarizes the experimental procedure and results.

Training Procedure			Test Results		
Group 1	$AX+$	$BX-$	$A \rightsquigarrow Resp.$	$B \rightsquigarrow -$	$X \rightsquigarrow -$
Group 2	$AX\pm$	$BX\pm$	$A \rightsquigarrow -$	$B \rightsquigarrow -$	$X \rightsquigarrow Resp.$

Table 2.1. Summary of the experimental procedure to establish the relative validity effect. A, B, and X are stimuli and $Resp.$ indicates a conditioned response.

While both groups of animals experience the same number of reinforced and unreinforced X trials, they demonstrate distinctly different behavior on test presentations of X. The difference lies in the relative ability of each stimulus to act as a reliable predictor of reinforcement. In the case of the first group, A is most consistently paired with reinforcement. In the second group, no stimulus is more consistently paired with reinforcement than any other, however in this case X is presented twice as often than either A or B so it comes to *block* responding to either A or B. As a result of this interpretation of these findings this phenomenon has come to be known as the *relative validity effect*.

Overshadowing, blocking and the relative validity effect are collectively referred to as cue competition effects because one can readily interpret these phenomena as the result of a competition between the stimuli (or cues) to account for the reinforcement. The stimulus that is more salient, or has a longer or more reliable history of pairing with reinforcement, will fare better in the competition. When these phenomena emerged in the late 1960's, they brought about a surge of theorizing on the principles governing the establishment of classical conditioning phenomena. By far the most influential theory to emerge from this period was that of Rescorla and Wagner (1972) (as described in Section 2.4).

2.1.4. Conditioned inhibition

Stimuli paired together interact in ways other than through the competition effects described above. One of the more dramatic examples of stimulus interaction is the phenomenon known as *conditioned inhibition*. Repeatedly pairing a conditioned *excitor*[1] A with a second stimulus B in the absence of reinforcement ($AB-$) results in B becoming a *conditioned inhibitor* and able to suppress responding to other conditioned excitors. Conditioned inhibition is theoretically interesting because it appears as though animals are capable of learning a negative correlation between B and reinforcement that effectively cancels the behavioral influence of the excitor.

[1] An excitor is a stimulus able to elicit a conditioned response. Repeated trials of $A+$ establishes A as a conditioned excitor.

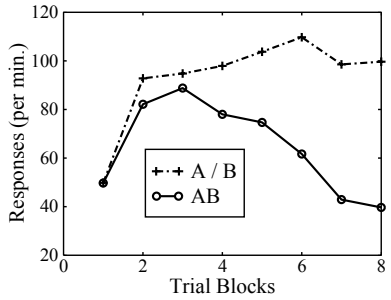


Figure 2.1. Negative patterning: $A+$, $B+$ and $AB-$. The empirical results (Pearce, 1994) showing that animals are able of learning negative patterning discrimination.

2.1.5. Second-order conditioning

The standard protocol for establishing second-order conditioning consists of two training phases. The first phase is a standard acquisition procedure with a stimulus A repeatedly followed by reinforcement ($A \rightarrow +$). The second phase involves a small number of pairings of A with a second stimulus B in the absence of reinforcement ($B \rightarrow A \rightarrow -$). Subsequent testing reveals that B is capable of eliciting a conditioned response. This phenomenon is known as second-order conditioning because B has been conditioned indirectly by way of A. Despite the similarity between the typical empirical protocols responsible for conditioned inhibition and second-order conditioning, these two procedures result in dramatically different behavioral outcomes. The relationship between these two phenomena is explored in some detail in Chapter 3.

2.1.6. Configural conditioning phenomena

Configural conditioning phenomena explore more complex interactions between stimuli and reinforcement than are found in either cue competition effects, conditioned inhibition or second order conditioning. These experiments seek to delineate animal behavior in the face of nonlinear contingencies between stimuli and reinforcement. The best known and perhaps simplest example of a configural conditioning effect is the XOR or negative patterning task. Here two stimulus elements, A and B, are each presented to the subject paired with reinforcement ($A+$ and $B+$), while a compound of the two is presented without reinforcement ($AB-$). As shown in Figure 2.1, subjects are eventually able to learn to discriminate the differential reinforcement contingencies of A, B and AB.

2.1.7. Other factors affecting learning

Stimulus timing: The relative timing between the stimulus and reinforcer also has a significant impact on the rate of acquisition or learning (Sutton & Barto, 1990). Acquisition

is achieved (to some preset behavioral criterion) in the fewest number of trials when the stimulus is followed immediately by reinforcement: the *delay* conditioning paradigm. In *trace* conditioning, where there is a gap between the stimulus offset and reinforcement onset, the rate of acquisition falls sharply with the size of the delay. Often no conditioning is achievable after more than a few seconds between stimulus and reinforcer.[2] Asymptotic conditioned response intensities are also diminished if the reinforcer is presented too early in the training trial, for example, if reinforcer onset co-occurs with stimulus onset.

Stimulus identity: The relative ubiquity of conditioning phenomena across species prompted early researchers including Pavlov (1927) and Skinner (1938) to think of conditioning phenomena as providing insight into a universal learning system—invariant of the choice of stimulus and reinforcer. This perspective appears to be an oversimplification. The work of Garcia and Koelling (1966) was one of the earliest experiments to call into question the universality of the laws governing the conditioned reflex. Contrary to the predictions of the universal mechanism perspective, Garcia and Koelling found that specific stimuli (CSs) were inclined toward associations with corresponding reinforcers (USs). Specifically they showed that *light+tone* formed a much stronger association with *shock* than with *illness*, while on the other hand a *flavor* stimulus quickly formed a strong association with *illness* but not with a *shock* reinforcer. Such findings led Seligman (1970) to propose the concept of "preparedness." According to Seligman, the biological context of a given conditioning paradigm determines to a degree the efficacy of training. The biological context predisposes animals to be prepared, unprepared, or contra-prepared to acquire specific stimulus-reinforcer associations.

2.2. Theories of cognition: levels of analysis

In the previous section I presented a broad overview of some of the best established empirical conditioning phenomena. With over a hundred years of experimentation, the classical conditioning literature presents a mass of data with a great number of distinctly labeled phenomena with numerous variations of each of these. In order to interpret and condense this growing wealth of empirical data, theorists have attempted to devise models that distill the essential principles of learning and present a coherent theory of how animals learn.

In classical conditioning, the system under study is a complex cognitive system. It is not immediately apparent what form a concise theory of conditioning should take. Should

[2]Interestingly, paradigms involving flavor-illness pairings are the exception, with strong conditioning achievable after just one trial and with hours separating the consumption of food and the onset of illness (Cheatle & Rudy, 1978).

it be a theory of behavior, concerned only with the stimuli and responses of a "black-box" system, or should it be a theory of brain function, tracing the neural pathways from sensory receptors, through the brain and then on to motor neurons? Which constitutes a theory of conditioning?

In studying the visual system, Marr (1982) emphasized the existence of multiple levels of analysis appropriate for the study of complex systems such as the brain. He identified three separate levels: the computational level, the representation and algorithm level, and the hardware implementation level. The computational level is concerned with specifying the goal of the computation to be performed and characterizes the system abstractly as a "mapping from one kind of information to another" (Marr, 1982, p. 24). The representational and algorithm level is concerned with how the computation is to be performed: specifying an appropriate representation of both input and output and the steps involved in transforming the input to the output. Finally, the hardware level of analysis is concerned with how the representation and algorithm are physically realized. Marr demonstrated that while these three levels are coupled – for example, the choice of the algorithm may depend on hardware limitations – the three may be profitably considered independently.

While Marr's own contributions to vision spanned all three of levels of analysis, he emphasized the importance of the computational level of analysis as a valuable perspective on developing theories of perception. He believed that the key to progress in model development was to start at the computational level. Framing the objective of the system under study provides necessary constraints for later representation and algorithmic development. He observed that "trying to understand perception by studying only neurons is like trying to understand bird flight by studying only feathers: It just cannot be done. In order to understand bird flight, we have to understand aerodynamics; only then do the structure of feathers and the different shapes of birds' wings make sense (Marr, 1982, p. 27)."

Anderson (1990) recognized the potential of Marr's perspective and considered its application to higher level cognitive functions such as categorization. According to Anderson, the adaptive character of Marr's computational level of analysis is fundamental, renaming it the rational level in his own hierarchy of analysis to highlight this point. Most cognitive theories, according to Anderson, are aimed at advancing our state of knowledge at the level of the algorithm or the implementation and, like Marr, he bemoaned this strategy as being woefully under-constrained and consequently often of little use in achieving a greater understanding of cognitive processes. As an alternative, Anderson emphasized the

role the rational level of analysis could play in developing theories of cognition. He introduced his *General Principle of Rationality* to codify this perspective: "The cognitive system operates at all times to optimize the adaptation of the behavior of the organism" (Anderson, 1990, p. 28). In what he called the "adaptionist principle" in Marr's argument, Anderson recognized that the evolutionary selective pressures could be viewed as a mechanism enforcing cognitive processing to adhere to rationality.

While Anderson sought to develop theories of human cognition, his ideas and the ideas of Marr apply equally to the development of theories of animals learning and behavior. In the following three sections, I consider existing theories of conditioning at each of the three levels of analysis.

2.3. Computational principles of conditioning

Unfortunately the theoretical discourse surrounding classical conditioning rarely transpires at the computational level of abstraction. Instead, conditioning theorist have proposed numerous competing theories of conditioning that specify detailed stimulus representations and learning algorithms. Nevertheless, these theories and their comparison do identify a number of important computational issues concerning the nature of learning.

At the computational level of analysis, classical conditioning can be conceptualized as consisting of three basic types of observable quantities: (1) the *conditioned stimuli*; (2) the *unconditioned stimuli*; and (3) the response[3]. Learning may be viewed as the animal's attempt to recover the basic statistical regularities between these basic quantities. However within the conditioning literature there exists a spectrum of perspectives on how learning is manifest in the mental processes of subjects undergoing conditioning. With regard to the question of what is actually learned, there are a number of distinct theories. These break down along the lines of the relationships between the stimuli, reinforcement and response. Many of the points of distinction between these various theories are principally algorithmic and will be discussed in more detail in section 2.4. Here we are concerned with the fundamental computational distinctions that exist.

2.3.1. Stimulus-response theory

One of the most fundamental distinctions that exists between theories of conditioning concerns the role behavior plays in learning. The debate centers around the question of

[3]Classical conditioning theory typically distinguishes between the unconditioned response elicited by the unconditioned stimulus and the conditioned response elicited by the conditioned stimulus after training. I do not make such a distinction here, though the response to which I refer should be clear from the context.

whether learning is manifest as the development of an association between the conditioned stimulus and the reinforcement or between the conditioned stimulus and the response.

Popular during the behaviorist period (approx. 1930–1960), SR theories contend that through the course of conditioning the subject learns an explicit association between stimulus and response (or between their respective mental representations). According to many SR proponents, the presence of reinforcement simply modulates the connection between the conditioned stimulus and the corresponding response. Learning is envisioned as the development of a response reflex triggered by the stimulus. Thus SR theories do not distinguish between the learning mechanisms responsible for classical conditioning and those responsible for operant or instrumental conditioning. Both are seen as a kind of response *shaping*, with reinforcement providing an incentive for behavior modification.

2.3.2. Stimulus-stimulus and stimulus-value theories

The prominence of the SR perspective has diminished substantially over recent decades. The reasons for this decline in popularity will be discussed in greater detail further in this chapter, when I discuss these theories in more detail. It suffices to say that the weight of empirical evidence did not support SR theory's presumed role of reinforcement as a facilitator of a reflexive connection between stimulus and response. The alternative that has largely displaced SR theory holds that the conditioned response triggered by the conditioned stimulus (after training) is the natural behavioral manifestation of the subject's *prediction* of impending reinforcement. A subject that learns through experience that an electric shock to the foot closely follows the sound of a tone, will respond to future presentations of the tone in a manner appropriate to an expectation of shock. According to this perspective, prediction is viewed as the central computational goal of conditioning and behavior is assigned a secondary role of merely providing a window into the predictions made by the subject.

Further distinctions exist within this prediction-based perspective. One such distinction, between stimulus-stimulus (SS) theories and stimulus-value (SV) theories, concerns the treatment of reinforcement.

Stimulus-stimulus (SS) theories treat the reinforcer similar to other stimuli, having distinct and salient perceptual characteristics. According to SS theory, conditioning forms a direct association between the mental representations of conditioned stimulus (e.g. tone, bell or light) and the unconditioned stimulus (e.g. food, nausea or foot-shock). After training, presentations of the conditioned stimulus stimulate the mental representation of the

unconditioned stimulus and thereby cause the subject to respond in a manner appropriate for the particular invoked unconditioned stimulus.

SV theories, on the other hand, reduce the unconditioned stimulus to a scalar *reinforcement value*. Thus SV theory considers conditioning to be the assignment of a reinforcement value to conditioned stimuli. The famous model of Rescorla and Wagner (1972), the Pearce-Hall model (Pearce & Hall, 1980) and the configural conditioning model of Pearce (1994) are all examples of SV theories.

Empirically evaluating the relative claims of the SS and SV theories of classical conditioning continues to be an active enterprise. There are experimental findings that seem to support each of these two perspectives over the other. A key experimental phenomenon in support of the SV view is transreinforcer blocking. As previously described, a subject exposed to repeated presentations of $A+$ followed by presentations of $AB+$ training will tend to show a significant *blocking* effect, with test presentations of B failing to elicit as strong a response as in control groups who lack the initial $A+$ training phase. In one variation of this classic experiment, the reinforcer in the second phase of training is replaced by a qualitatively distinct reinforcer (with $AB+'$ in place of $AB+$). In certain circumstances A may continue to act as a blocking stimulus and inhibit the acquisition of a robust response to B (relative to controls lacking $A+$ training) (Williams, 1994; Ganesan & Pearce, 1988). Transreinforcer blocking supports the SV perspective on reinforcement because it appears to demonstrate that the association between A and reinforcement is not tied to a specific reinforcer, but rather to reinforcement in general. Also, as the relative difference between the intrinsic values of the reinforcers $+$ and $+'$ increases (for example, with differing quantities of food), the transreinforcer effect diminishes (Dickinson, Hall, & Mackintosh, 1976) as would be predicted by a SV theory.

Empirical evidence in support of SS theory over SV theory takes the form of an experiment that appears to directly contradict the transreinforcer blocking effect. Blaisdell et al. (1997) report that by making a qualitative change in reinforcement between the two phases of a typical blocking paradigm (for example from $+$ to $+'$ above) actually results in a significant reduction of the blocking effect. The resulting *unblocking* effect appears to support the SS theory based notion that subjects learn more about the reinforcer than simply its affective value. This experiment suggests that they also learn perceptual qualities of the reinforcer that render it distinct and not simply an interchangeable measure of affective value. At this point, it seems that some compromise or hybrid of the SS and SV treatment of reinforcement is necessary to account for all the available data. I will return to some of

these issues and experiments in Chapter 4 where they will be addressed in the context of the latent cause theory of classical conditioning.

2.3.3. A statistical perspective

A statistical formalism offers a precise and unambiguous description of SS and SV theories at the computational level of analysis. However it also opens the door to other questions relating to the process of learning that are not directly reflected in the distinctions between SR, SS and SV theories. SS and SV theories both describe learning as the formation of associations between conditioned stimuli and reinforcement. In the language of probability theory, many of these models broadly correspond to attempting to recover the probability (or probability density) of reinforcement conditioned on the observed stimuli, $P(Reinforcement \mid CSs)$. Thus learning may be seen as estimating the model that best reflected the subject's experience of reinforcement in the presence of the conditioned stimuli. This type of model is often referred to as a *discriminative* or *conditional* model because it is solely concerned with recovering information directly related to reinforcement. An important consequence of this strict interpretation of the discriminative view of conditioning is that it there is no mechanism for the subject to predict impending conditioned stimuli. The stimuli are of no interest save for their use in predicting reinforcement.

A *generative* model is an alternative view that has received relatively little attention as a model of classical conditioning. According to the generative perspective of conditioning, animals are engaged in an attempt to recover the full joint probability distribution of all observed stimuli, including both the conditioned stimuli and the reinforcer, $P(CSs, Reinforcement)$. While most formal models of conditioning focus on prediction of reinforcement and thus are best described as discriminative models, some informal (or descriptive) models of conditioning (including the comparator hypothesis of Miller and Matzel (1988) which will be described in section 2.4) contend that subjects learn direct and explicit associations between reliably co-occurring conditioned stimuli. As I will discuss at length in this and subsequent chapters, there is considerable empirical evidence in that animals do learn to predict conditioned stimuli.

2.3.4. Contiguity versus contingency

The relationship between contiguity, contingency and learning is one of the most controversial issues in classical conditioning. Before the 1960s, the dominant behaviorism perspective was that temporal proximity or contiguity between stimulus, reinforcement

and the evoked response was necessary and sufficient for the establishment of conditioning (Mowrer & Klein, 2001). Tolman was the exception. He was an early adherent to the perspective that contingency was critical to learning—that learning was a matter of discovering "what leads to what" (Wasserman & Miller, 1997; Tolman, 1932). [4] Yet, regardless of Tolman's influence and despite Pavlov's extensive empirical investigations into the role of contingency in inhibitory conditioning effects (Pavlov, 1927); the behaviorists of the early to mid-1900s remained set in sufficiency of contiguity as a determinant of conditioning.

It was not until the mid-1960's that experimental evidence such as Kamin's (1969) blocking experiments as well as his investigations into overshadowing (Kamin, 1967, 1969) and the predictive reliability of stimuli (Wagner et al., 1968) that the role of contingency was taken seriously as a determinant of conditioning phenomena. These experiments demonstrated that temporal contiguity was not sufficient for conditioning: a stimulus could be repeatedly paired with reinforcement in close temporal contiguity and still fail to elicit a response. Also, contrary to the predictions of earlier S-R theories, these experiments seemed to demonstrate that the relationship between one stimulus and the reinforcer was dependent upon other stimuli present during training.

2.4. Representational/Algorithmic theories of conditioning

While Marr's computational level of analysis is useful to delineate gross theoretical distinctions between the various schools of thought, the conditioning theories themselves are typically expressed at Marr's algorithmic level of analysis. By specifying learning rules, associative learning theorists explore the consequences of these choices and gauge how well they explain behavioral data. In this section, I review the key theories that have shaped our understanding of classical conditioning.

2.4.1. Stimulus-Response theories of conditioning

Thorndike: A contemporary of Pavlov, Thorndike formulated his influential law of effect (Thorndike, 1911) without knowing of Pavlov's earlier investigations with dogs. Thorndike's theory was developed to explain how hungry cats learned to operate an arbitrary release mechanism in order to escape from the "puzzle box" and gain access to the food outside. Both Pavlov and Thorndike used food reinforcement, but in Thorndike's experiments its delivery was dependent on the subject performing the correct response—i.e. the action that activated the release mechanism on the door of the box. Thorndike's law

[4] Some theorists, most notably Guthrie, took the extreme position that simple contiguity between the stimulus and the response was sufficient for conditioning (Guthrie, 1935, 1959). According to Guthrie, the role of reinforcement was simply to reset the stimulus context, thereby preventing further responses becoming associated with the stimulus. This perspective has not held up to experimental verification (Seward, 1942).

of effect is an early example of an SR theory. In establishing the SR association, the reinforcement plays a special role in Thorndike's theory. In essence, if a response or behavior is followed by a "satisfying" state of affairs then the S-R connection or association is strengthened (meaning that the response is more likely next time the stimulus is experienced) and if a response is followed by an "annoying" state of affairs then the S-R connection is weakened. Thorndike (1913) provided the following operational definition:

> By a satisfying state of affairs is meant one which the animal does nothing to avoid, often doing things which maintain or renew it. By an annoying state of affairs is meant one which the animal does nothing to preserve, often doing things which put an end to it. (p.2)

Hull: An extremely influential figure in the world of psychology of the 1930s, Hull's most significant work, *Principles of Behavior* (1943), elaborates on Thorndike's law of effect. This is, in part, through the inclusion of the concept of drive. According to Hull, drive may be either innate (unconditioned) or acquired, and it acts to automatically motivate behavior. Innate drives include internal stimuli such as hunger or thirst, but can also include intense environmental events such as electric shock or a loud noise. Besides innate sources of drive, Hull suggested that environmental stimuli can acquire the ability to produce a drive through classical conditioning.

When an organism is aroused by a drive, it is motivated to reduce this drive. Drive reduction is rewarding and any behavior that reduces the drive is likely to be associated with that drive state. If a behavior consistently functions to reduce a drive, the behavior's habit strength increases. Habit strengths are similar to Thorndike's S-R connections in that they form an association between a behavior and a stimulus. However in Hull's theory the drive takes the place of the stimulus.

Hull largely dismissed Pavlov's stimulus substitution theory and viewed the effects Pavlov observed as simply more elementary versions of the same phenomena as those studied by himself and Thorndike before him. He suggested that Pavlov's "mistaken induction was presumably due in part to the exceedingly limited type of experiment which he employed" (Hull, 1943).

Skinner: Of the early behaviorists, it was Skinner who first acknowledged the value of Pavlov's stimulus substitution theory (Skinner, 1938). He recognized the distinction between Pavlovian (or classical) conditioning and instrumental conditioning. Like Pavlov, he understood classical conditioning as the process of a conditioned stimulus coming to substitute for the reinforcing stimulus. He contrasted this with instrumental conditioning

where the animal "selects" from a large repertoire of actions to actively produce (or avoid) reinforcement.

While Skinner's proposal—that classical conditioning is a legitimate phenomenon distinct from instrumental conditioning—did allow for a greater role for Pavlov and his theoretical ideas, it was a tempered endorsement as Skinner himself insisted that instrumental conditioning 'played a more important role' (Skinner, 1938, p.111). Skinner thought that classical conditioning was a relatively unimportant phenomenon because he believed the conditioned responses it gives rise to accounted for a tiny fraction of the animal's potential behavioral repertoire and above all because these conditioned responses do not achieve anything (Mackintosh, 1983).

Skinner's approach to the study of learning differed significantly from that of Pavlov or of even Thorndike and Hull. Skinner believed that the goal of the scientific study of learning is to identify and isolate the environmental factors that govern behavior (Mowrer & Klein, 2001). He was an extreme empiricist, who claimed that a true understanding of behavior could only be achieved if it can be predicted and controlled, which in turn can only be accomplished through an understanding of the environmental factors governing the occurrence of behavior. Consequently Skinner undervalued the role of theory in advancing an understanding of learning and behavior.

2.4.2. Evidence against the SR perspective

The 1960's saw a substantial resurgence of interest in classical conditioning. This is largely due to a number of key empirical developments that demonstrated that classical conditioning was far more pervasive than previously believed. One of the most significant developments was the discovery of autoshaping in pigeons (Brown & Jenkins, 1968). Skinner used pigeons in his operant or instrumental conditioning experiments. In his experiments, pigeons were trained to peck at an illuminated key by means of food reinforcement. The pecking response was established through a procedure known as "shaping." In the case of key pecking the shaping process involved the experimenter providing a food reward whenever the subject was standing near the key, then when the subject looked at the key, and then eventually when the subject pecked the key. According to Skinner, shaping involved the instrumental reinforcement of successive approximations to the final desired behavior (Mackintosh, 1983). Brown and Jenkins discovered that pigeons could be trained to peck at an illuminated response key simply by illuminating the key briefly before delivering food reinforcement. No tedious shaping procedure was necessary.

The experiment of Brown and Jenkins, while difficult to explain through an operant conditioning mechanism, was readily interpretable in Pavlovian terms. The illuminated key served as a conditioned stimulus (CS) and the delivery of food as the unconditioned stimulus (US). The conditioned response to the key was pecking – a pigeon's natural response to a food stimulus. To use Pavlov's terminology, the key comes to substitute for the food reinforcement and hence elicits the characteristic pecking response.

2.4.3. Stimulus-Stimulus theories

Pavlov: Throughout his long career investigating conditioning phenomena, Pavlov continued to view the adaptive phenomena he was studying as belonging more to physiology than psychology. Accordingly he labeling them conditioned reflexes. Yet despite this, Pavlov's (1927) stimulus substitution theory was in some sense much more cognitive than those offered by Thorndike and later by Hull. According to Pavlov, his dogs came to salivate at the ticking of the metronome because the metronome had come to substitute for the food, as a consequence of the pairing of metronome and food. Pavlov saw the basis of conditioning as simple temporal contiguity: if two stimuli are presented together and one of those stimuli innately elicits some reaction, then the other stimulus will eventually come to elicit a similar reaction.

Tolman: During the heyday of behaviorism in the 1930s-50s and amidst the development of the more prominent SR theories of Hull and others, there were relatively few learning theorists taking a more cognitive perspective. One notable exception is E. C. Tolman. While Tolman considered himself a behaviorist (eschewing the introspection of Wundt and Titchener), today he is recognized as more of a cognitive behaviorist (Mowrer & Klein, 1989). He rejected the kind of direct or simplistic behaviorism introduced by Watson and later espoused by Thorndike, Hull and Skinner.

The SR theories of Thorndike and Hull fit nicely within classical reflex theory, a dominant neurophysiology theory of the day. Classical reflex theory views brain function—ultimately cognition—as a series of responses to stimuli. According to this perspective, conditioning and more generally learning is the process of establishing new stimulus-response connections or rearranging old ones. Though simplistic, the theory seemed to be supported by the available experimental evidence. The behavioral adaptations apparent in Thorndike's cats as they learn to escape from the puzzle-box, or in Pavlov's dogs as they come to salivate to the sound of a metronome, or even in Skinner's rats as they learn to lever-press for food, are all readily explained as the emergence of a new stimulus-response reflex.

Tolman understood the connection between stimulus and response as something more sophisticated than a simple reflex. Rather than viewing learning as the acquisition of new reflexes, Tolman saw learning as the acquisition of new information (or as he put it, knowledge) about the world. Consequently, Pavlov's dogs do not learn to salivate *per se*; instead they are learning that the sound of the metronome signals an increase in the probability of food delivery. From this perspective, conditioning involves the encoding of relationships between events in the environment (Mowrer & Klein, 2001). Tolman believed the function of conditioning is "to enable animals to discover the causal structure of their world" (Mackintosh, 1983, p. 11) (based on (Tolman & Brunswick, 1935)).

One consequence of Tolman's cognitive interpretation of conditioning is that it introduced a gap between learning and behavior or, as it is more commonly stated, a distinction between learning and performance. If learning is, as Tolman contends, a process of information acquisition and processing, then no behavioral manifestation is required. With this, Tolman introduced the notion of *latent* learning (Tolman, 1955), a process of acquiring knowledge that has no immediate impact on behavior. In a classic experiment, Tolman and Honzik (1930) allowed hungry rats to run through a complex maze, but without receiving any food reward. Following these free running sessions, food was introduced into the maze goal box and the rats were again permitted to run through the maze. Tolman observed that on the second reinforced trial, the rats dramatically changed their behavior, running quickly and making few errors on their way to the goal box. In fact, these rats would make on average the same number of error as rats that had been reinforced for all the "free running" trials. Experiments such as this one convincingly demonstrated that learning took place both without reinforcement, but also without apparent behavioral modification. Hull (1943) seemed to have accepted this distinction between learning and performance, as did Skinner (1938) with his distinction between a *reflex reserve* (learning) and a *reflex strength* (performance potential).

Comparator Hypothesis: A more recent example of an SS theory is offered by Miller and Matzel's (1988) comparator hypothesis. According to the comparator hypothesis, at least three associations are learned during classical conditioning. One is the traditional association between the target stimulus and the reinforcement (as in the Rescorla-Wagner model). The second is an association between the target stimulus and a *comparator stimulus* which could be the context or another salient stimulus present during training with the target. The third is an association between the comparator stimulus and reinforcement. Responding to the target stimulus during testing is determined by a competition between the strength of the association between it and reinforcement and the product of the strength

of the two associations involving the comparator stimulus. Thus if either of the two comparator associations are weak, the target will elicit a response, otherwise the response will be suppressed.

The comparator hypothesis has been used to account for various experimental phenomena and provides a novel performance deficit explanation of cue competition effects such as blocking. As described in Section 2.1.3, training with $A+$ followed by training with $AB+$ classically results in decreased responding to B (relative to controls). The comparator hypothesis accounts for blocking by considering the initial $A+$ training as having established a significant association between A and reinforcement. Subsequent training with $AB+$ further strengthens the A-reinforcement association, but also establishes associations between B and reinforcement and between A and B. In the context of testing with B, A plays the role of the comparator stimulus and, due to the establishment of a strong A-reinforcement association through $A+$ training, is able to significantly inhibit responding to B.

One of the more compelling features of the comparator hypothesis is its ability to explain recovery from blocking. Blaisdell, Gunther, and Miller (1999) showed that subjects who receive unpaired presentations of the blocking stimulus (A) subsequent to a typical blocking paradigm achieve significant recovery of responding to the otherwise blocked stimulus (B). According to the Comparator Hypothesis, the unpaired presentations of A act to weaken the A-reinforcer association and thus impair A's ability to compete against the B-reinforcement association. As a result, further testing with B reveals a recovery of responding.

2.4.4. Stimulus-Value theories

The SV theories, which include most of the currently popular models of conditioning, have dominated classical conditioning theory since the late 1960s. In this section I review the most successful theories.

Rescorla-Wagner theory: The Rescorla-Wagner model casts classical conditioning as a process of learning an association between the initially neutral stimuli and a reinforcer (US). The nature of association is determined by the contingent relationship between the stimuli and the reinforcer. If there exists no contingency, that is, if a particular stimulus is not useful in predicting the presence (or absence) of the reinforcer, then no association is formed.

According to the Rescorla-Wagner model, conditioning proceeds as a process of apportioning *associative strength* or responsibility for reinforcement to the various stimuli

present in the environment. The associative strength of a given stimulus is a measure of its reliability as a predictor of reinforcement. More precisely, each stimulus i possesses an associative strength, V_i. The associative strength of stimuli present on a trial sum to form the total predicted reinforcement intensity V_Σ:

$$V_\Sigma = \sum_i V_i X_i, \tag{2.1}$$

where the X_i assumes a value of 1 if stimulus i is present and 0 otherwise. Through the summation of associative strengths, the stimuli present on a trial combine their individual predictions of reinforcement. Learning, in the Rescorla-Wagner model, amounts to gradient descent on the squared error between the observed reinforcer intensity, R, and its predicted value,

$$\Delta V_i = \alpha_i \beta (R - V_\Sigma) X_i \tag{2.2}$$

Here α_i is known as the associability of conditioned stimulus i — a stimulus specific learning rate. The parameter β is the associability of the reinforcer. The Rescorla-Wagner model successfully accounts for an impressive array of conditioning phenomena, including acquisition, extinction, conditioned inhibition, overshadowing and blocking (Miller, Barnet, & Grahame, 1995). I will briefly describe the Rescorla-Wagner model's explanation of blocking to illustrate how the model deals with cue competition effects.

As described above, a blocking protocol typically begins with a phase of $A+$ training. Early in this phase of learning, the presentation of the reinforcer causes a large error as both V_A and consequently V_Σ are initially zero. This non-zero error causes the associative strength of A to grow. Eventually $V_A = R$ and A comes to fully account for the occurrence of the reinforcer. By the beginning of the second phase of learning, with the $AB+$ trials, there is only residual error resulting in little associative strength accumulation for stimulus B. Thus on testing with B, the small associative strength, V_B, translates (via equation 2.1) into the weak prediction of reinforcement characteristic of blocking.

Over the past three decades the Rescorla-Wagner model has remained the dominant theory of classical conditioning. It has influenced a generation of learning theorists and researchers to such an extent that relatively few question the basic tenets of the associative learning approach. Despite the success of Rescorla-Wagner theory, there are a number of classic phenomena that remain beyond its ability to predict. For the most part, these can be broken down into a few distinct lines of active research (for a much more exhaustive treatment of the successes and failures of the Rescorla-Wagner theory of conditioning (Miller et al., 1995)) and I will delve into some of these in greater detail in subsequent chapters of this thesis.

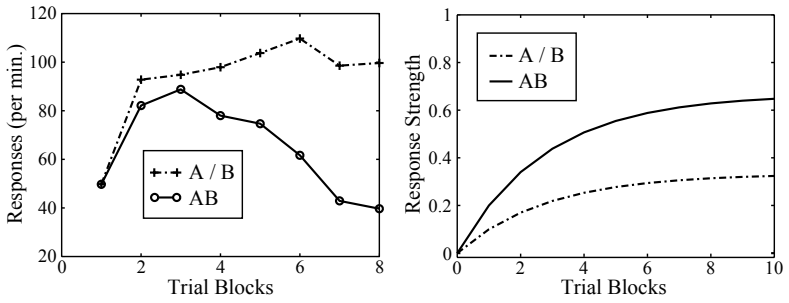
One of the classic failures of the Rescorla-Wagner model is the inability to account for second-order conditioning effects (as discussed in Section 2.1.5). According to the Rescorla-Wagner model, the first phase of learning ($A \rightarrow +$) will result in A receiving a positive associative strength. On subsequent pairings of $B \rightarrow A \rightarrow -$ (with no reinforcement), the positive associative strength of A causes a prediction of reinforcement. Since no reinforcement arrives on these trials, the incorrect prediction results in a negative error which accumulates as the negative associative strength V_B. Thus, according to the model B should become a conditioned inhibitor rather than a second-order excitor.

The Rescorla-Wagner model also fails to explain many so-called attentional effects such as latent inhibition. According to the model, nothing is learned without a reinforcement prediction error. Hence, the large number of presentations of $A-$, that form the latent inhibition pre-training phase, have no effect on the model. As a result, subsequent acquisition proceeds as it would if the unpaired stimulus presentations never occurred and the rate of acquisition is entirely unaffected.

Configural conditioning effects are another set of empirical phenomena for which the Rescorla-Wagner model offers no explanation. As described in section 2.1.6, these are phenomena where the reinforcement contingency is a nonlinear function of the observed stimuli. Since the RW model represents reinforcement prediction of the compound as equal to the sum of the reinforcement predictions of the constituent elements, the model cannot adequately represent the nonlinear contingencies characteristic of configural phenomena such as negative patterning discrimination. Figure 2.2 illustrates the contrast between the empirical results and the predictions of the Rescorla-Wagner model as they pertain to the negative patterning (XOR) discrimination task (see section 2.1.6).

Pearce-Hall theory: Shortly after the emergence of the Rescorla-Wagner model, the question of how to best handle the associabilities (learning rates) arose within the conditioning community. Rescorla-Wagner assumes a constant associability with the update of the associative strengths being the focus of learning. However, there is considerable empirical evidence – in the form of phenomena such as latent inhibition (Lubow & Moore, 1959) and unblocking effects (Dickinson et al., 1976) – that appear to support the notion of changing associabilities. There have been some number of attempts to reflect these empirical observations in a theory, with the Pearce-Hall model being generally regarded as the most successful.

The Pearce-Hall model (Pearce & Hall, 1980) was originally developed as an adaptation of an earlier model by Mackintosh (1975) and offers an alternative account of conditioning phenomena to that of Rescorla-Wagner. Like its predecessor, the Pearce-Hall model



(a) Experimental Data (b) Rescorla–Wagner Simulation

Figure 2.2. Negative patterning: $A+$, $B+$ and $AB-$. (a) The empirical results (Pearce, 1994) showing that animals are capable of learning negative patterning discrimination. (b) A simulation of the Rescorla-Wagner model illustrates the model's failure to capture nonlinear discriminations such as this negative patterning experiment.

focuses on the adaptation of stimulus associability; however in direct contrast to the model of Mackintosh, the Pearce-Hall theory posits that the associability of a stimulus is proportional to the extent that it is *not* an accurate predictor of reinforcement on the last trial:

$$\alpha_{i,t} = |R_{t-1} - V_{\Sigma,t-1}|, \tag{2.3}$$

where $\alpha_{i,t}$ is the associability of the ith stimulus on trial t, R_{t-1} is the value of reinforcer on the last trial $(t-1)$, and $V_{\Sigma,t-1}$ is the total predicted reinforcement from all stimuli at trial $t-1$. $V_{\Sigma,t}$ is given by the same equation as in the RW model:

$$V_{\Sigma,t} = \sum_i V_{i,t} X_{i,t}. \tag{2.4}$$

where $X_{i,t}$ represents the presence ($X_{i,t} = 1$) or absence ($X_{i,t} = 0$) of the ith stimulus at time t. The associative strength update rule is proportional to the current value of the associability, weighted by the values of the stimulus and reinforcer intensity ($S_{i,t}$ and R_t respectively),

$$\Delta V_{i,t} = S_{i,t} \alpha_{i,t} R_t. \tag{2.5}$$

The Pearce-Hall model is able to account for a number of the standard conditioning phenomena including blocking and overshadowing. A more elaborate version of the model (Pearce & Hall, 1980) is needed to account for "inhibition effects" such as extinction and conditioned inhibition. Where the Pearce-Hall model distinguishes itself from the Rescorla-Wagner model is its prediction of attentional effects such as latent inhibition. Unlike Rescorla-Wagner theory, the Pearce-Hall model is able to offer a satisfactory explanation of latent inhibition. According to this perspective, initial presentations of $A-$ establish a low associability for A. With the associative strength of A, $V_{A,t}$ initially low, unpaired

presentations of A result in an associability near zero, $\alpha_{A,t} \approx 0$ (from Equation 2.3). The reduced associability resulting from $A-$ pre-training leads to a reduction in the associative strength learning rate (through Equation 2.5) which in turn results in the empirically observed retarded rate of acquisition characteristic of latent inhibition.

The temporal difference model: One important line of recent experimental and theoretical advancement addresses the Rescorla-Wagner model's lack of a representation of time within a trial. As a *trial-level* model, stimuli and reinforcements are represented as simply being present or absent on a trial. This representational deficiency of Rescorla-Wagner was addressed by the introduction of a number of so called *real-time* models. Of these the most popular is the temporal difference model.

According to the temporal difference (TD) model of conditioning (Sutton & Barto, 1990), learning can be understood as the attempt to predict expected future reward. Similar to the Rescorla-Wagner model, the predicted future reward, $V_\Sigma(t)$, is given by a weighted sum of stimuli present on the current time-step,

$$V_\Sigma(t) = \sum_i V_i X_i(t) \tag{2.6}$$

$X_i(t)$ represents the presence of stimulus i at time t. V_i plays an analogous role as it did in the Rescorla-Wagner model. Each V_i is updated at each time step by the learning rule, $V_i(t) = V_i(t-1) + \Delta V_i(t)$, where

$$\Delta V_i(t) = \alpha_i \beta (r + \gamma V_\Sigma(t+1) - V_\Sigma(t)) \overline{X}_i(t) \tag{2.7}$$

The parameter $\gamma \in (0, 1)$ is the discount factor. The eligibility trace, $\overline{X}_i(t)$, is a memory representation of stimulus i that persists after the stimulus disappears. It has dynamics given by $\overline{X}_i(t+1) = \overline{X}_i(t) + \delta(X_i(t) - \overline{X}_i(t))$, where the parameter δ determines the decay rates of stimulus memories. Comparing this learning rule with that of Rescorla-Wagner, we see that TD acts to minimize the difference between $r + \gamma \hat{V}_{t+1}$ and $\hat{V}_t$.

As a real-time model, the TD model can account for a number of stimulus timing effects. It can explain the empirically-observed overriding of prior training by temporal primacy (Kehoe, Schreurs, & Graham, 1987). The model can also account for the observation that rabbit eyeblinks are specifically timed to occur at the time of expected reinforcement (Kehoe, Graham-Clarke, & Schreurs, 1989). However, in order to make this prediction the TD model requires an additional representational device known as the "complete serial compound representation" (Sutton & Barto, 1990). The complete serial compound incorporates timing information about past stimuli by having them march through a cascade

of delayed internal stimulus representations that can be individually associated with reinforcement.

The TD model of conditioning is able to account for second-order conditioning (as it is described here). The first phase of $A \rightarrow +$ training causes A's associative strength to increase as it learns to predict future reinforcement. With the first few trials of $B \rightarrow A \rightarrow -$, the future reinforcement prediction backs up from A to B imparting it with significant associative strength. TD also correctly predicts the transient nature of second order conditioning with it giving way to conditioned inhibition when information regarding the absence of reinforcement eventually reaches B (through the backups). Unfortunately the TD account of this effect is quite brittle. It does not account, for instance, for the observation that second-order conditioning may be established via simultaneous stimulus presentations (Rescorla, 1982; Yin et al., 1994).

Theories of configural conditioning: As described in section 2.1.6, configural conditioning effects are phenomena where the reinforcement contingency is a nonlinear function of the observed stimuli. There is currently considerable debate regarding the theoretical account of configural conditioning phenomena. The two dominant perspectives are Pearce's configural theory (Pearce, 1994) and the *added elements* Rescorla-Wagner model (Wagner & Rescorla, 1972) (name due to Wagner and Brandon (2001)). Both theories augment the stimulus representation X_i of the original Rescorla-Wagner model to include "configural" units corresponding to conjunctions of stimuli. In particular, both theories assume that a unique configural unit is added for each stimulus compound observed, such as AB.

The theories differ as to how they apportion activation over the stimulus representation X_i (for all i) and how they apportion learning over the associative weights V_i. In the added elements Rescorla-Wagner model, an input unit is active ($X_i = 1$) if the stimulus pattern it encodes appears within the observed stimulus pattern ($X_i = 0$ otherwise). For learning, the weight corresponding to each active input is updated according to the typical Rescorla-Wagner delta rule. Alternatively, the Pearce model spreads graded activation over all X_i, based on a measure of similarity between the observed stimulus compound (or element) and the compounds represented by the model's representational units. In particular, denoting the number of stimulus elements present in an observed stimulus pattern a as $\mathrm{size}(a)$, and in the pattern represented by the ith configural unit as $\mathrm{size}(i)$, then the activation of unit i by pattern a is given by $X_i = \mathrm{size}(\mathrm{overlap}(a, i))^2/(\mathrm{size}(a) \cdot \mathrm{size}(i))$. The learning phase only updates the weight corresponding to the configural unit that exactly matches the observed stimulus configuration.

Both the configural theory and the added elements Rescorla-Wagner model handle negative patterning with ease. The inclusion of the AB configural unit renders the discrimination linearly solvable in the associative weights. Yet neither model is able to completely account for all known configural conditioning effects, with some experiments seeming to support Pearce's configural theory while others seem to support the added elements Rescorla-Wagner model. In Chapter 3, I explore the relationship between these models and the empirical data in greater depth.

2.5. Physiologically constrained theories of conditioning

The few theoretical successes notwithstanding, the advancement in our understanding of conditioning phenomena has been relatively modest since the development of the Rescorla-Wagner model. The challenge we face is that, if we are to capture some aspect the richness and sophistication of animal behavior, the models we develop almost certainly have to become more complex. It seems exceedingly unlikely that animals (or humans) process information using the elementary representation and adaptation described by Equations 2.1 and 2.2. Yet, within the classical conditioning research community, there is significant reluctance to explore more sophisticated models. Some of this reluctance is justified. Without powerful constraints guiding theory development, searching for the "right" model among the infinite possibilities is a daunting proposition.

The behavior of an organism is constrained both from the top down and from the bottom up. Top-down constraints are those that recognize that the organism's behavior cannot be arbitrary if it is to continue to survive. These are the constraints imposed on the system from analysis at the computational level. Bottom-up constraints are those that acknowledge the neurophysiological underpinning of any behaving organism. In this section, I review the major developments in imposing bottom-up constraints on theories of conditioning.

Of particular relevance to the theory advanced in this thesis are the recently developed connectionist models of hippocampal function. Among the most influential of the early computational theories of the role of hippocampus in classical conditioning was Sutherland and Rudy's (1989) configural association theory. While the hippocampus was already recognized to be involved in spatial localization and navigation (O'Keefe & Dostrovsky, 1971; O'Keefe & Nadel, 1978) and suspected to play a role representing stimulus configuration (Wickelgren, 1979), the proposal of Sutherland and Rudy was perhaps the strongest claim to date regarding the hippocampal role in the conjunctive representation of stimuli. They argued that the hippocampus was responsible for representing conjunctions of

stimuli and, since configural conditioning phenomena require conjunctive represention for their solution, they predicted that damage to the hippocampus would impair performance on configural discrimination problems (at least those that lack a solution linear in the individual stimulus elements). The idea that the hippocampus is necessary for configural learning grew very popular and several computational theories of cortical-hippocampal interaction were advanced that cast the hippocampus in this role.

The "real-time" connectionist model of (Schmajuk & DiCarlo, 1992) describes the relationship between the hippocampus, conjunctive representations of stimuli and classical conditioning. The nodes in their model were mapped to regions of the brain including the hippocampus, the cortex and the cerebellum (which has been demonstrated to be involved in conditioned response timing (Medina, Nores, Ohyama, & Mauk, 2000)). The model ascribes to the hippocampus the computation of an aggregate reinforcement prediction error. The Schmajuk and DiCarlo model accounts for a considerable number of conditioning phenomena, but it fails to address several theoretically important phenomena involving stimulus-stimulus learning including sensory preconditioning, in which the hippocampus is thought to play a role (Gluck & Myers, 1993). Sensory preconditioning is similar to second-order conditioning, but with the order of the training phases reversed ($AB-$ followed by $A+$ elicits responding to B during testing).

The anatomy of the hippocampus, specifically the recurrent connectivity of CA3 neurons led, in part, to Marr (1971) proposing his *auto-associator* theory of hippocampal function. An auto-associator is a network capable of learning associations between the individual stimulus inputs. Building on the computational principle of an auto-associator network, Gluck and Myers (1993) proposed an *auto-encoder* theory of hippocampus. An auto-encoder is a network that learns to reproduce its input at the output layer, while possessing a constricted internal layer that forces the network to develop an internal representation that compresses redundancies in the input stimulus pattern. Like the Schmajuk and DiCarlo model, the Gluck and Myers model is also more than simply a model of hippocampus. It is a model of the interaction between the hippocampus and the cortex (cerebellum included), with the hippocampus auto-encoder driving the internal representation of the cortical network. In this way the cortex is envisioned as a rather impoverished learning structure, reliant on the hippocampus for learning anything more than simple linear associations between stimuli and reinforcement. Unlike the Schmajuk and DiCarlo model, the auto-encoder of Gluck and Myers permits an accounting of stimulus-stimulus learning effects such as sensory preconditioning and other phenomena such as latent inhibition.

Both the model of Schmajuk and DiCarlo (1992) and that of Gluck and Myers (1993) share the Sutherland and Rudy (1989) prediction that hippocampal damage would seriously impair configural conditioning discriminations. They both assume a key role for the hippocampus in learning and in particular in learning configurations of stimuli[5]. While at the time the available empirical evidence supported the hypothesis of a significant role for the hippocampus in configural conditioning phenomena (Rudy & Sutherland, 1989), further experimentation did not bear this out (Rudy & Sutherland, 1995). There are today clear examples in which extensive damage to the hippocampus did not impair subjects' performance on configural conditioning tasks (Whishaw & Tomie, 1991; Gallagher & Holland, 1992).

In light of these revelations regarding the role (or perhaps lack of a role) of the hippocampus in configural conditioning, O'Reilly and Rudy (2001) proposed a new theory of the interaction between the hippocampus and the cortex where the hippocampus plays a much less crucial role. While they still maintain the idea that the hippocampus is involved in conjunctive stimulus representations, its role is reduced to being necessary for incidental configurations, such as the formation of context associations (such a role appears to have empirical support (Fanselow, 1990; Philips & LeDoux, 1992)) and not necessary for the establishment of configurations that are driven by prediction error, as found in nonlinear discrimination tasks. Following the ideas of McClelland, McNaughton, and O'Reilly (1995), O'Reilly and Rudy advanced a theory of the division of labor between the hippocampus and the cortex. The hippocampus is claimed to be involved in fast conjunctive learning and maintaining distinct representations for similar stimulus patterns. To compliment hippocampal function, the cortex is responsible for slower learning, integrating similar stimulus experience to resolve general trends. Based on this division of labor, O'Reilly and Rudy proposed a neural network model that demonstrates performance on incidental and non-incidental configural tasks with simulated hippocampal lesions that is in accord with the empirical findings.

Beyond the various theories concerning the roles of hippocampus, cortex and cerebellum in learning, a recent avenue of research relating temporal difference (TD) theory with the activity of dopaminergic neurons has drawn considerable attention throughout the sciences of behavior and brain. Dopamine-containing neurons are found in the ventral tegmental area (VTA) and substantia nigra pars compacta (SNc) and project widely to

[5]The Schmajuk and DiCarlo model hypothesized a stronger role for the hippocampus in conditioning phenomena with the property that even phenomena such as blocking would be impaired by hippocampal damage.

the striatum, nucleus accumbens and cortex. As described above, the TD theory of conditioning claims that animals learn and encode within the associative strengths a prediction of future reward. A series of neural recording studies (see (Schultz, Dayan, & Montague, 1997) for a review) showed that a considerable proportion of dopamine neurons respond in bursts of activity to unexpected rewards and to stimuli that are predictive of reward; however these neurons do not respond to predicted or expected rewards (Mirenowicz & Schultz, 1994). In addition, when expected rewards fail to arrive the dopamine neurons momentarily pause their background activity roughly at the time the reward is expected. Together these empirical results strongly suggest a role for the dopamine system in the prediction of future reward. More specifically, a number of researchers hypothesize that dopamine neurons encode the TD error signal (Equation 2.7) and have developed computational models to test this prediction (Friston, Tononi, Reeke, Sporns, & Edelman, 1994; Houk, Adams, & Barto, 1995; Montague, Dayan, & Sejnowski, 1996; Schultz et al., 1997). Although the details of the models differ and none completely accounts for all of the available data, further empirical study (Schultz, 1998; Bao, Chan, & Merzenich, 2001) has largely supported the TD-dopamine hypothesis.

2.6. Statistical theories of conditioning

Consideration of the physiology of the brain provides a means of constraining the algorithms that might be considered plausible theories of learning in animals and humans. However these sorts of implementational concerns are not the only source of constraints to be exploited. As stated in Section 2.2, both Marr (1982) and Anderson (1990) believed that a coherent computational or *rational*[6] theory of behavior offers a means of guidance in our search for successful theories of conditioning.

In recent years, there has been a concerted effort to advance computational theories of conditioning. These theories tend to view learning as the attempt to recover statistical regularities present in the environment. In this setting the learning problem is usually framed as involving the estimation of the probability or rate of reinforcement in the presence of some pattern of observed stimuli. This line of inquiry has led to the development of novel

[6] A rational theory of behavior is in some sense the antithesis of the reflex theory that dominated the theoretical landscape of psychology in the early part of the century. Embraced by Thorndike and later by Hull, reflex theory interprets cognition simply as a network of stimulus-response connections, with more complex responses requiring more network connections. The brain and its function were analogized to the great technological innovation of the day: the telephone switchboard. Where reflex theory takes a mechanistic approach, rational theory takes a computational approach. A rational theory of conditioning would have the animal reasoning about the stimulus information and arriving at a response that is optimal under some set of environmental and evolutionary constraints. Thorndike, who was emphatically opposed to the notion that animals reason (Thorndike, 1911), would likely have rejected a rational theory of conditioning. Rational theory is much closer to the more cognitive expectancy theory of Tolman (Tolman & Brunswick, 1935).

theories of conditioning as well as new analyses' of the statistical properties of existing theories. In particular, the Rescorla-Wagner model has be shown to correspond to an optimal sequential estimator of the expectation of reinforcement assuming a linear relationship between reinforcement value and the combination of available conditioned stimuli (Sutton, 1988).

In the remainder of this section I review some of the more influential statistical models of conditioning: rate estimation theory (Gallistel & Gibbon, 2000) and the statistical learning models of Dayan and colleagues (Dayan & Kakade, 2001; Dayan & Long, 1998; Kakade & Dayan, 2002).

2.6.1. SET and RET

Across many conditioning paradigms and many measures of responding, variability in animals' response timing is consistently proportional to the mean of the timed interval. Scalar expectancy theory (SET) (Gibbon, 1977) was developed as a statistical model of response timing to provide a theoretical account of this "scalar timing" aspect of animal behavior. Originally a fairly abstract theory, SET has grown more mechanistic, incorporating a timing mechanism, a memory mechanism and sources of noise in the decision variable. Eventually SET was fused with rate estimation theory (RET), becoming the response timing module of the extended theory (Gallistel & Gibbon, 2000).

Gallistel and Gibbon's (2000) Rate Estimation Theory (RET) offers a frequentist account of some striking quantitative regularities emerging from autoshaping experiments with pigeons. Compiling data from a large number of separate experiments (following a *delay* protocol where the stimulus is temporally extended and reinforcement delivery coincides with stimulus offset), Gallistel and Gibbon (2000) argue that the speed of acquisition depends on three critical variables: the duration of the stimulus, T; the duration of the intertrial interval, I; and the training schedule, S (the fractional number of reinforcement deliveries per stimulus presentation). In fact, over several orders of magnitude the median number of trials to acquisition depends only on the ratio $(I + T)/T$ and not on I and T independently. According to RET, animals undergoing learning are engaged in estimating the rates of reinforcement attributable to each stimulus including the background context. Specifically, the estimated instantaneous rate of reinforcement is determined by the sum of the rates associated with each stimulus present.

Despite RET's billing as a normative, statistical account of autoshaping, the interpretation of RET in rational terms is not without its troubles. As Kakade and Dayan persuasively argue, in order for it to quantitatively match the empirical data, RET is required to

make unreasonable assumptions regarding an animal's tolerance for uncertainty (Kakade & Dayan, 2000, 2002). With more reasonable tolerances, RET dramatically overestimates the speed of acquisition. Notwithstanding these deficiencies, RET stands as one of the most widely recognized normative accounts of conditioning phenomena.

2.6.2. Uncertainty and conditioning

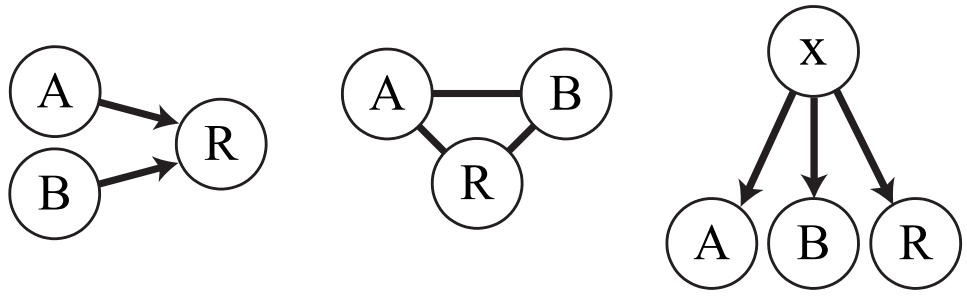
Building on their analysis of RET, Dayan and Kakade went on to propose a statistical approach to conditioning phenomena involving multiple stimuli (Dayan & Kakade, 2001; Dayan, Kakade, & Montague, 2000). They formulated a Bayesian version of the Rescorla-Wagner model that formally incorporated uncertainty over the associative strengths. By encoding the uncertainty information with Gaussian distributions, they were able to use Kalman filter updates to determine optimal values of the associabilities (or learning rates) with respect to the uncertainty.

Dayan and Kakade applied the model to explain the phenomenon of backward blocking, where two stimuli are initially paired with reinforcement, $AB+$, and then in a second phase of training one of the two is individually paired with reinforcement, $A+$. In experiments with people and animals[7], the second phase of $A+$ training results in significant reduction in responding to stimulus B (when compared to appropriate controls) The Kalman filter model of Dayan and Kakade successfully captures this effect by maintaining uncertainty between the relative contributions of A and B to the prediction of reinforcement through the first phase of training. The second phase reveals that A alone is predictive of reinforcement and consequently the uncertainty collapses to a point where A and B have large and small associative strengths respectively. Thus, subsequent testing with B reveals it predicts little reinforcement intensity. I will explore this model further in Chapter 5 when I discuss trial order effects and generative models of change.

2.6.3. Discriminative versus generative models revisited

One commonality between RET and the models of Dayan and colleagues is that they are all conditional or *discriminitive* models. In fact, almost all theories of conditioning reviewed in this chapter are discriminative in nature. In statistical terminology, each attempts to recover the probability or expected rate of reinforcement conditional on the observed stimuli: $P(R \mid A, B, ...)$. As discussed in section 2.3, the only information discriminative models encode regarding the stimulus–stimulus associations is how they bear on reinforcement prediction. This style of model has been the foundation of computational

[7]Observing the phenomenon in animals requires further experimental manipulation (Arcediano, Escobar, & Miller, 2004).



(a) A discriminative model (b) A generative model (c) A latent cause model

Figure 2.3. Graphical depictions of three different probability models: (a) a discriminative model, (b) a generative model and (c) a latent cause model.

models of classical conditioning, and most associative theories, including the Rescorla-Wagner model, may be interpreted as models of this type. Figure 2.3(a) shows a graphical depiction of a discriminative model describing the relationship between stimuli A, B and the reinforcer R.

While it is clear from the success of associative learning theory that many conditioning phenomena may be satisfactorily explained without appeal to explicit inter-stimulus associations, it is appropriate to question whether this is really the right strategy in the context of classical conditioning. Are animals capable of learning associations between neutral stimuli? The evidence appears to be strongly in the affirmative. The clearest example of this kind of learning is the sensory preconditioning effect (Brogden, 1939; Prewit, 1967; Rizley & Rescorla, 1972; Barnet, Grahame, & Miller, 1991b), where two neutral stimuli, A and B are initially paired in the absence of reinforcement: $AB-$. After a second phase of training, where one of the stimuli is paired with reinforcement, $A+$, significant responding is observed in B. Presumably an association formed between A and B in the initial pairing of the two, thus allowing for subsequent excitatory transfer from A to B. Discriminative models would not predict excitatory transfer because they learn nothing during the initial pairing of A and B. Without some sort of an association between A and B there is no mechanism to support excitatory transfer.

The alternative to discriminative models are *generative* models, so called because they explicitly model the generative process that gives rise to both the input and the output. In the context of classical conditioning this means a process generating both stimulus patterns and reinforcements. In some sense the simplest generative model of stimulus configurations would encode pairwise correlations among the stimuli. If stimuli A and B reliably

co-occur, the correlation would be represented explicitly with a parameter encoding the strength of the correlation. Figure 2.3(b) gives a graphical depiction of this kind of model.

While such a generative model encodes relationships between stimuli, it actually fails to reproduce some of the fundamental patterns of behavior we wish to use it to explain. Consider once again the sensory preconditioning effect. Training with trials of $AB-$ and $A+$ would presumably result in associations being formed between A and B and A and reinforcement ($+$). There is no direct association between B and reinforcement. Consequently when testing with B alone, we are required to condition on the event of A being absent. That is we are asking: What is $P(R \mid B, notA)$? Following formal inference procedures, conditioning on the absence of A renders B and reinforcement statistically independent. Such a model would predict that the presentation of B alone results in no increase in expectation of reinforcement over that of the background context.

Figure 2.3(c) shows an alternative generative model structure with each of A, B and the reinforcer R being independently associated with a fourth variable x. If we assume that x represents an unobserved or *latent* variable whose value must be inferred from the values of the stimulus variables, then such a model provides an interpretation of the phenomenon of sensory preconditioning. Training trials of $AB-$ and $A+$ will form positive correlations between each of the stimuli (A, B and R) with the latent variable x because, in both cases, the only means of correlating stimuli is through simultaneous correlations to the latent variable. One consequence of these positive correlations is the emergence of a imputed positive correlation between B and reinforcement. I refer to it as spurious because it arises from a structural constraint on the model rather than from the natural statistics of the training schedule (B and reinforcement are never observed together). Nevertheless this spurious correlation captures the behavior of animals subject to sensory preconditioning training: conditioned responding with presentations of B. Generative models with this structure form the basis of the *latent cause theory* of conditioning advanced in this thesis. In the following section I consider how such an associative structure of events might naturally arise in the context of classical conditioning.

2.7. A latent cause theory of conditioning

In this section I propose a new rational perspective on animal learning. The ideas developed here form the central theoretical construct of this thesis: the latent cause theory of conditioning.

In accord with the rational perspective (Anderson, 1990), a central tenet of the theory developed in this thesis is that animals are approximating rationality, in the sense that their

behavior corresponds to some normative standard of behavior (Schooler, 2001). Unfortunately since the appropriate normative standard depends on the context, the definition of rationality necessarily must also be context-dependent. In the context of economics and decision making, rational behavior is often defined as maximizing a personal utility function. In the context of classical conditioning such a definition is inappropriate since no action the animal performs has a significant impact on the environment. Thus an alternative notion of optimality must be adopted, leading to the obvious question: *What is an appropriate normative standard to adopt in the context of classical conditioning?*

Consider the original justification motivating the rational approach to conditioning: behavior is optimized as a result of evolutionary pressure. Animals are presumably optimized to behave appropriately in the natural world. There they are not passive observers, but must act and interact with their environment to survive. To the animal seeking to manipulate events in its favor, statistical correlations in the absence of any information bearing on the causal flow of events is of limited value. Recent empirical investigations with human subjects have demonstrated the importance of considering causal inferences (Glymour, 2001; Tenenbaum & Griffiths, 2001b, 2003).

This line of reasoning leads to the hypothesis that, in the context of classical conditioning, animal cognition is optimized to recover what Tolman and Brunswick (1935) called "the causal texture of the environment." Associative learning theories such as the Rescorla-Wagner model are often interpreted as capturing the subject's inference regarding the causal relationship between stimulus and reinforcement (A. G. Baker & Mehta, 2001). In certain circumstances such an approach may even be justified from a rational perspective (Yuille, 2005). However, in the context of classical conditioning, interpreting Rescorla-Wagner causally leads to unlikely and unreasonable inferences regarding causal mechanisms in the world.

Consider a subject undergoing simple acquisition training. The subject hears a tone and then immediately feels a jolting shock. How is the subject to interpret the relationship between the tone and the shock? Is it a reasonable interpretation that the sound of the tone caused the shock? What in the animal's experience would lead it to believe that sounds *cause* anything? Do sounds cause shocks? It seems an unlikely inference. A more plausible inference would be that there is a common unobserved or latent cause for both the tone and the shock. In this case, this would be a correct inference with the experimenter (or some experimental control mechanism) being that common cause.

Considerations such as this lead to a very different model than the discriminative structure shown in Figure 2.3(a) on which most models of conditioning are based. Throughout this thesis, I take the position that all stimuli, including reinforcement, are conditionally independent given the value of some latent or unobserved cause: $P(R, A, B, \cdots \mid x) = P(R \mid x)P(A \mid x)P(B \mid x)$, where x is the latent cause. The model is illustrated in Figure 2.3(c) as a directed graphical model.

It is important to note that in some experimental situations inferring a direct causal relationship between the stimulus and the reinforcer is the most reasonable inference. For example, in studies where food is paired with a noxious agent (such as lithium chloride), interpreting the food as having *caused* the illness is a plausible and parsimonious inference.[8] I will focus on the latent causal structure both because I believe it to be the most common inferred structure in conditioning paradigms (a consequence of the commonality of symptomatic stimuli such as sounds and lights) and because it is the more novel and hence interesting of the two causal structures described above.

In this thesis I do not explicitly consider the development of a complete model of causal inference in animals. Such a project would incorporate the effect of interventions[9] on the model of inference. While such an endeavor would be enlightening, the extension is not a simple one. It requires prior knowledge about causal mechanism - for example, that food can make one sick, or that tones are unlikely to cause shock. The theory-based inference scheme (Tenenbaum & Griffiths, 2003) is a formal framework developed to address just these issues in problems of human causal induction. The framework stipulates that humans combine prior assumptions about the plausibility of causal relationships between elements of their environment with experiential data in a manner consistent with Bayesian inference. It appears that the theory-based framework would readily apply to the problem of animal causal induction, although it seems an entirely open question as to what the set of prior plausible causal relationships might be (beyond what I have already suggested).

The theory-based framework has important implications for analysis at the rational level. Following a Bayesian formalism, rationality is defined as standard statistical inference with respect to the prior information available to the subject. As part of the prior information, the theory-based framework stipulates that the subject has prior assumptions or "theories" regarding the possible causal structure of the world. Such assumptions could include the concept of stimulus-reinforcer "preparedness" introduced by Seligman (1970) to account for stimulus-reinforcer specific acquisition speeds (Garcia & Koelling, 1966) (see

[8] I will return to this point in Chapter 4 where this distinction is manifest in the behavior of subjects.

[9] A causal intervention is an act of setting one of the variables to some value. It has different semantics than the probabilistic notion of conditioning.

section 2.1.7). Indeed such stimulus-reinforcer specificity has recently been traced down to specific neural pathways in the rat (Balleine & Dickinson, 2000). This use of these theories as priors permits biological constraints to inform analysis at the rational level without breaking with the tenets of rationality. Such use of priors as a means of interaction between Anderson's various levels of analysis is a promising avenue of future theory development.

Following the theory-based framework, I adopt Bayesian statistical inference over the latent cause structure as the basis of rational behavior. I will not require the additional machinery that has recently been developed to handle interventions (Pearl, 2000) because in the classical conditioning setting no such intervention is possible. It may seem strange that I appeal to the necessity of causal considerations over the purely statistical when modeling conditioning phenomena, yet simultaneously reject the need for anything other than standard Bayesian inference once the model structure is established. There is no inconsistency in this position. I am explicitly following the causal inference program: with a known causal structure and without the ability to execute interventions, causal inference, as described by Pearl, collapses to standard Bayesian inference.

The theory proposed in this thesis is a theory of classical conditioning, not a theory of causal induction in animals. I offer the causal interpretation of the theory simply as a motivation for the particular choice of generative model structure (as shown in Figure 2.3(c)). Despite its name, the relevance of the latent cause theory of classical conditioning to learning phenomena does not hinge on the veracity of its causal interpretation. Alternative justifications of the latent cause model structure exist that do not involve the ambiguities inherent in consideration of causality. Specifically, the latent cause approach to conditioning may be motivated by considerations of the statistical properties of the natural environment in which animals evolved.

2.8. Latent cause: an efficient representation

Both in the laboratory and in their natural environment, stimuli are abundant and highly redundant. In the laboratory for example, experiments manipulating contexts (Justin A. Harris & Westbrook, 2000; Bouton, 1993) and experiments such as the sensory preconditioning experiments discussed above all demonstrate that subjects are learning more about their surroundings than a simple association between the target stimuli and reinforcement. These experiments suggest that they learn to correlate contexts with punctate stimuli and stimuli with each other.

The potential proliferation of stimuli presents a problem if we take to heart the data that suggest subjects routinely learn associations between congruent neutral stimuli. Can

we realistically expect subjects to generate associations between all possible combinations of stimuli? This would result in an explosion of parameters (associative strengths) exponential in the number of stimuli. Even if we presume that subjects are only keeping track of stimulus pair frequencies, the number of parameters to learn would still grow quadratically with the number of stimuli.

There are a number of ways we can get out of this predicament. It's possible that certain stimuli – presumably those with low saliency – are ignored and are not associated with other stimuli. This still leaves the problem of selecting which stimuli are to be ignored. An alternative to limiting the number of stimuli to include in an all pairwise correlation memory structure is to focus on how all the stimuli are associated with a single cue. In such a structure, the number of parameters grows linearly with the number of stimuli. Using an observable stimulus is not feasible because any predictions would be exceedingly brittle if the stimulus is noisy. Instead, if we employ a latent cause linked to all stimuli, then we are left to infer its value from the presence of the various observable stimuli. The use of a latent variable to which all observable stimuli are associated significantly limits the kinds of associations that are expressible. However, in cases where the stimuli are highly redundant this may be an efficient representation scheme, where the notion of "everything is correlated with everything" is simply encoded as "everything is correlated with the latent variable".

CHAPTER 3

Similarity and Discrimination

Animals routinely demonstrate the ability to represent complex relationships between stimuli and the capability to distinguish subtle differences in patterns of stimulation. At the same time, they display an impressive ability to generalize from previous experience to novel situations. But most impressive is their ability to negotiate between these competing and often contradictory requirements of discrimination and generalization. Within the classical conditioning literature, an area of recent empirical and theoretical development known as configural conditioning explores issues of generalization and discrimination through experiments involving configurations of stimuli.

Many configural conditioning phenomena highlight the ability of animals to solve nonlinear discrimination tasks. Because the Rescorla-Wagner model assumes that the associative strength of a stimulus compound is the sum of its elements' weights, it famously cannot solve nonlinear discriminations. The best known example of these is the negative patterning task, in which for example, a light or tone predict reinforcement when presented individually, but not together. Associative learning theorists have offered a number of models that improve upon the Rescorla-Wagner account. The two most popular of these are the *added elements* Rescorla-Wagner theory (Wagner & Brandon, 2001) and Pearce's configural theory (Pearce, 1994). Both these models solve nonlinear discriminations by augmenting the Rescorla-Wagner model with *configural units* that act as detectors of conjunctions of stimuli and independently enter into association with reinforcement. The inclusion of a configural unit detecting the light-tone compound renders the above negative patterning example linearly solvable.

The introduction of the configural unit device raises a number of questions. For instance, under what circumstances should a configural unit be created? If a particular stimulus compound is observed, should activation generalize to partially overlapping configural

units? How should learning be distributed over the configural units? Though the elemental and configural perspectives suggest different answers to these questions, there seems to be no obvious theoretical reason to favor either. Pearce's model differs from Rescorla and Wagner's by positing different, and more "configural," answers to these questions. Both theories match an impressive range of experimental data, but each is refuted by some experiments that the other explains. Without unambiguous experimental or theoretical guidance, it is unclear how to move beyond this stalemate.

In this chapter I consider configural conditioning effects and related phenomena from the perspective of the latent cause theory of conditioning. While it is generally recognized that configural conditioning experiments probe how animals generalize and discriminate between patterns of stimuli, here I demonstrate that generalization and discrimination actually represent opposing sides of a tradeoff fundamental to the theory of adaptive systems: the tradeoff between increasing data fidelity and decreasing model complexity. I suggest that animals are approximating rational inference with respect to the uncertain *structure* of the latent cause model. A characteristic feature of rational structural inference is the tradeoff between the model complexity and the amount of evidence needed to support it. By considering model simulations of a number of conditioning paradigms (including some not previously viewed as "configural"), I reveal behavioral signs that animals are employing these kinds of tradeoffs attributed to them.

Relating latent cause theory to the previous models of configural conditioning, I show how the latent causes play a role analogous to the configural unit in Pearce's configural theory. Questions regarding the formation of configural units, how their activations generalize to overlapping stimulus patterns, and how learning should be apportioned over them are answerable using standard Bayesian inference.

3.1. Previous models of configural conditioning

As discussed in Chapter 2, the Rescorla-Wagner model's prediction of reinforcement is taken to be the sum of the individual predictions of reinforcement of all the stimuli present on that trial. This linear relationship prevents the model from explaining many phenomena recognized as configural conditioning. There have been a number of attempts at fixing this deficiency in the Rescorla-Wagner model by augmenting the representation structure while maintaining the basic associative mechanism for learning. In this section I review theories of configural conditioning and discuss the supporting data.

3.1.1. Pearce's Configural Model

As an alternative to the elemental Rescorla-Wagner perspective, Pearce (1994) proposed a theory of learning with configurations of stimuli. According to the theory, conditioning with a pattern of stimuli results in a single association between a representation of the pattern and the reinforcer. This is in contrast to the Rescorla-Wagner elemental model which claims that associations may also be formed with the individual elements of the stimulus pattern. The Pearce configural model states that, during learning, a pattern of stimuli is presented to a network that spreads activation from an input stimulus layer to a layer of configural units. The associative strength E_i of the configural unit i is then updated according to

$$\Delta E_i = \alpha_i \beta (R - V_i). \tag{3.1}$$

The parameters α_i and β play similar roles here as they do for the Rescorla-Wagner model. The term V_i represents the total reinforcement prediction in the presence of stimulus pattern i, and is given by

$$V_i = \sum_j a_{ij} E_j \tag{3.2}$$

where the sum is over all configural units and a_{ij} is the activation of the configural unit j in the presence of stimulus pattern i. It is determined by

$$a_{ij} = \frac{n_{ij}^2}{n_i n_j}. \tag{3.3}$$

n_i is the number of stimulus elements in the pattern represented by configural unit i, n_j is the number of elements present in stimulus pattern j and n_{ij} is the number of member stimulus elements of configural unit i that occur in stimulus pattern j. While any configural unit with a nonzero number of elements in common with the current stimulus pattern j will be *active* in the sense of having a nonzero contribution to the reinforcement prediction V_i, only the weight corresponding to the configural unit j that exactly matches the stimulus pattern i (with $a_{ij} = 1$) will be updated according to equation 3.1.

3.1.2. Other models of configural conditioning

A number of other models offer alternative accounts of configural conditioning phenomena than those provided by either the added elements Rescorla-Wagner model or

Pearce's configural theory. One recent model of Brandon et al. (2000), the *replaced elements* model, takes a compromise position between the Rescorla-Wagner and Pearce models. Their model includes unique compound cues similar to the added elements Rescorla-Wagner model, but also contains specific inhibitory connections between stimulus representations which collectively have an effect similar to Pearce's generalization rule. The model is able to account for data that neither the added elements Rescorla-Wagner model or Pearce's model can explain (Wagner & Brandon, 2001). An additional set of models stemming from the connectionist tradition offer viable competing accounts for configural conditioning phenomena. In particular the model of hippocampus by Gluck and Myers (1993) seems capable of accounting for a wide range of configural conditioning phenomena—though to the best of my knowledge it has never been offered as an explanation for these phenomena.

While these alternative models provide interesting avenues for progress in explaining a widening set of experiments, they are not yet established contenders within the configural conditioning literature and so I will not consider them further. Instead I will continue to compare the two models that have dominated the theoretical development of configural conditioning phenomena: the added elements Rescorla-Wagner model and Pearce's configural conditioning model.

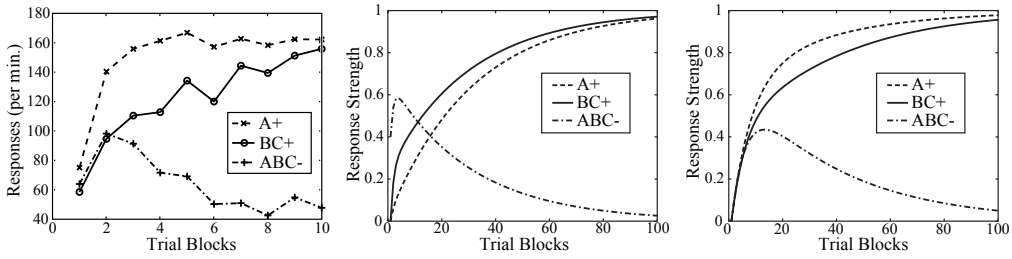
3.2. Assessing competing accounts of learning patterns

Attempts to assess the predictive accuracy of these two competing theories has been inconclusive. In this section I present a selective review of the evidence in support of each theory. Some of these experiments will be revisited in Section 3.6 when I present the predictions made by latent cause theory.

3.2.1. Evidence in support of Pearce's configural theory

Much of the data supporting the Pearce model over more elemental theories (such as the added elements Rescorla-Wagner model) originates from experiments relating the relative discriminability of various stimulus compounds during training. The core idea linking these experiments is the notion that compounds with more stimuli in common are harder to discriminate than those with fewer stimuli in common and that this generalization across stimulus compounds results in slower discrimination.

A variant on negative patterning highlights this effect. Redhead and Pearce (1995a) presented subjects with reinforced stimulus patterns $A+$, $BC+$ and the unreinforced pattern $ABC-$. They found that discrimination between A and ABC was acquired more



(a) Experimental Data (b) AERW Simulation (c) Pearce Simulation

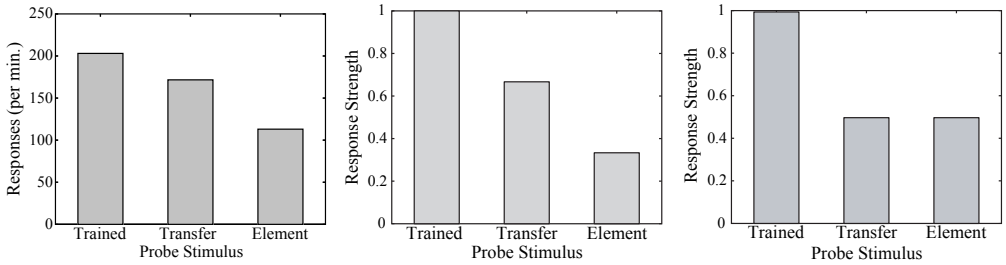
Figure 3.1. Asymmetric negative patterning experiment, trained with $A+$, $BC+$ and $ABC-$. (a) The empirical results (adapted from (Redhead & Pearce, 1995a)). Comparison of (b) the added elements Rescorla-Wagner model (AERW) and (c) Pearce's configural theory applied to the asymmetric negative patterning experiment. In the simulations a trial block consists of one presentation of each of $A+$ and $BC+$ and two presentations of $ABC-$.

quickly (achieved in fewer trials) than the discrimination between BC and ABC. Figure 3.1 illustrates the empirical results as well as the predictions of both the added elements Rescorla-Wagner model and Pearce's configural model.

The compounds BC and ABC share two stimuli while A and ABC only share one. Thus according to the configural theory of Pearce, BC is more similar to ABC than is A. This difference in similarity results in more activation of the ABC compound during testing of the BC compound than with the A stimulus. Increased activation of the ABC compound results in a reduced prediction of reinforcement because ABC is associated with an absence of reinforcement. The resulting generalization asymmetry leads to faster learning with A than with BC as shown in Figure 3.1 (c).

The elemental approach predicts the opposite result: that discrimination between BC and ABC should proceed more quickly than between A and ABC. This is a result of the greater number of elements in the BC compound (there are 3 elements B, C and BC) compared to the stimulus A (with only one element). On each trial, the weight associated with each element is updated by an amount equal to $\mathit{prediction\ error} \times \mathit{learning\ rate}$. With more stimuli the combined effect of updating each element results in faster learning.

A second experiment, along a similar theme as the last, was performed by Rescorla (1972) (see also Pearce and Redhead (1993)). In this experiment, one group, the control, is presented with traditional negative patterning training: $A+$, $B+$ and $AB-$. The other group is also presented with a negative patterning discrimination but with the addition of a common stimulus, C, to all the trial types: $AC+$, $BC+$ and $ABC-$. Rescorla found that discrimination was slower to learn when the subjects were trained in the presence of C, the common stimulus.



(a) Experimental Data (b) AERW Simulation (c) Pearce Simulation

Figure 3.2. The trained compounds experiment with training: $AB+$, $CD+$. (a) The empirical results (adapted from (Rescorla, 2003a)) testing the response to the trained compounds (AB, CD), the transfer compounds (AD, BC) and the individual elements (A,B,C and D). Compare the predictions offered by (b) the added elements Rescorla-Wagner model (AERW) and (c) Pearce's configural theory.

Pearce's configural theory is able to account for this result as another consequence of stimulus generalization. A greater number of stimuli in common causes a greater activation of the reinforced configural units (AC and BC) during presentations of the unreinforced compound ($ABC-$). The activation of the unreinforced configural unit ABC is similarly increased, compared to the control discrimination, during presentations of $AC+$ and $BC+$. Increased activation of the configural units on the opposite side of the discrimination reduces the speed with which the learning rule, equation 3.1, distinguishes the stimulus compounds.

As with the previous experiment, the added elements Rescorla-Wagner model predicts an increase in the rate of discrimination with an increase in the number of stimuli, even when those stimuli are common to both sides of the discrimination. This occurs because the addition of a redundant stimulus causes there to be more units available for learning (A, B, C, AC, BC and ABC) compared to the non-redundant case (with A, B and AB). As before, the extra units effectively increase the learning rate and lead to faster discrimination with the redundant stimulus. Thus the model prediction is in direct opposition to the experimental findings (Rescorla, 1972).

3.2.2. Evidence in support of the elemental theory

When two stimuli are individually paired with a reinforcer and then presented together, the result is often a greater response than either stimulus would elicit in isolation. This effect is appropriately termed summation and has been observed repeatedly and in a variety of conditioning preparations (Pavlov, 1927; Kehoe, 1986; Rescorla & Coldwell, 1995;

Rescorla, 1997). However there have also been specific failures to observe summation in pigeon autoshaping experiments (Aydin & Pearce, 1997; Rescorla & Coldwell, 1995).

The added elements Rescorla-Wagner theory of conditioning accounts for summation phenomena very naturally. Individual reinforcement of the stimuli, $A+$ and $B+$ results in significant associative strength for each. When paired together the associative strengths add and result in a linear increase in the expectation of reinforcement.

The configural theory of Pearce does not explain summation effects as easily as the elemental models. A direct application of the stimulus generalization rule, Equation 3.3, would result in a 0.5 contribution from the A unit and a 0.5 contribution from the B unit resulting in the same predicted response for the compound as for the individual stimuli. Pearce (1994) suggests that one can account for summation effects through the inclusion of a context stimulus, X. In this formulation, the subject is trained with $AX+$ trials and $BX+$ trials and the response to the probe ABX is measured. With the inclusion of the context stimulus X, the generalization rule results in the compound ABX getting a factor of $(0.667 + 0.667) = 1.333$ increase in the expectation of reinforcement over that elicited from the presentation of individual stimuli.

While the inclusion of a context stimulus does allow Pearce's model to explain some forms of summation, Rescorla (1997) notes that in some cases the salience of the context stimulus would have to be greater than that of the elements. Taken as a whole, summation phenomena seem more parsimoniously explained with an elemental account.[1]

From the point of view of comparing these two theories of conditioning with compound stimuli, a particularly interesting series of experiments were recently performed by Rescorla (2003a). The basic experiment consisted of training the subject on a pair of compounds $AB+$ and $CD+$, then measuring the response to the originally trained compounds, the individual elements and a pair of transfer compounds AD and BC. The investigation involved both pigeons and rats with consistent results across a number of different conditioning paradigms. After training, the strongest responding was elicited when the animal was presented with the originally trained compounds (AB and CD). Presentations of the transfer compounds (AD and BC) elicited a moderate amount of responding while presentations of the individual elements resulted in the weakest conditioned responding. Figure 3.2 illustrates the behavioral results for this experiment together with the predictions of the added elements Rescorla-Wagner model and Pearce's configural theory.

[1] A recent line of experimentation (Pearce & George, 2002) has supported the claims that the salience of the background stimuli can affect the occurrence of summation. While these experiments do not address the specific problems with the Pearce account of conditioning (raised in (Rescorla, 1997)), they do indicate that more experimentation is needed to further elucidate this vexing issue.

The added elements Rescorla-Wagner theory, with the inclusion of the "unique cue" assumption, provides a clear accounting of the experiment. During training, the elemental units A, B, C, D and the compound units AB and CD all acquire positive associative strength. When testing with the trained compounds, the elements and the compound units are activated together and their associative weights sum to the full expectation of reinforcement acquired during training. Testing on the individual elements results in a small expectation of reinforcement since only one of the three units simultaneously active during training is activated with the presentation of a single element. Testing on the transfer compound reveals a moderate expectation of reinforcement. Presenting AD, for example, results in the A and D units both activating and summing their individual reinforcement expectations. Thus elemental theory predictions are consistent with the experimental findings of Rescorla (2003a).

Configural theory can account for the finding of this experiment only by again appealing to the inclusion of an explicit background stimulus. Such a maneuver causes the total generalization from the trained compound to the transfer compound to be greater that the generalization from the trained compound to the elements, resulting in a prediction of the experimentally observed order of response magnitudes to the three test conditions. Without the appeal to the common background stimulus, the configural theory predicts that generalization should result in equal responding to individual elements and the transfer compound.

3.2.3. Contradictory Evidence

Consider a discrimination experiment with trials of $A+$, $B+$, $C+$, $AB+$, $AC+$, $BC+$, and finally $ABC-$. Pearce's configural model would predict that discriminating between $A/B/C$ and ABC would be faster than discriminating between $AB/AC/BC$ and ABC. In contrast to this prediction, the added elements Rescorla-Wagner model would predict that the compounds $AB/AC/BC$, having more units active during learning, effectively possess a larger learning rate and consequently would discriminate faster from ABC than would the elements.

One might be tempted to suggest that this would be a means to choose among these two alternatives, but unfortunately this experiment seems to be of little help. Redhead and Pearce (1995a) report that pigeons are faster to discriminate between $A+$ and $ABC-$ than between $AB+$ and $ABC-$ — in support of the configural model; while Myers et al. (2001) report the opposite result in rat experiments — in support of the elemental model. Other attempts to discriminate between these two alternative models are divided, with some

contributing evidence in favor of Pearce's configural model (Nakajima, 1997; Nakajima & Urushihara, 1999; Pearce, Aydin, & Redhead, 1997; Redhead & Pearce, 1995b); while others provide results supporting the elemental perspective (Rescorla, 1997, 1999). Numerous researchers have suggested that a more parsimonious explanation may lie somewhere between these two models, incorporating features of each (Rescorla, 1999; Delamater, Sosa, & Katz, 1999).

3.2.4. Where do we go from here?

Both the configural model of Pearce and the added elements Rescorla-Wagner model offer superior accounts of configural conditioning phenomena to the original Rescorla-Wagner model, but comparing the two has proved challenging. Experiments performed over the past decade have failed to produce a clear victory for either model. The two models seem to possess complementary predictive powers that naturally lead one to consider whether some sort of compromise model that incorporates features of each might offer some way out of the present stalemate. Such an approach has been pursued by Brandon et al. (2000).

As an alternative to these associative theories, I approach the phenomena of configural conditioning from the perspective of latent cause theory. I advance a theory of animal behavior consistent with the notion that animals synthesize their previous experiences in a manner equivalent to Bayesian inference over latent causes. By reconsidering the wider issues of similarity and discrimination in the context of a general learning system, one gains a powerful vantage point on the principles behind animal behavior.

3.3. A theoretical framework for generalization and discrimination

The tradeoff between generalization and discrimination might best be understood by studying models at opposite extremes of the tradeoff. First, consider the model M_1 where the predicted reinforcement intensity in the presence of a given stimulus compound is taken to be average reinforcement intensity across all stimulus presentations. Such a scheme is depicted graphically in Figure 3.3(a). While this model would readily generalize to any stimulus compound, its impoverished representational capacity renders it unable to discriminate between any stimulus compounds irrespective of the degree of similarity.

Alternatively, consider the model M_2 where perceptually distinguishable events are treated as entirely independent. Experiences with light and tone individually paired with reinforcement reveals nothing about reinforcement in the presences of both light and tone.

Such a model, illustrated in Figure 3.3(b), possesses the ability to discriminate between arbitrarily similar stimulus patterns but no capacity to generalize. M_2 is arbitrarily complex in the sense that any pattern of reinforcement contingency may be represented.

Finally consider the compromise model, M_3, shown in Figure 3.3(c). M_3 displays a degree of generalization with "nearby" or similar stimulus patterns being similarly predictive of reinforcement; while still maintaining the potential to discriminate between stimulus patterns that are differentially predictive of reinforcement.

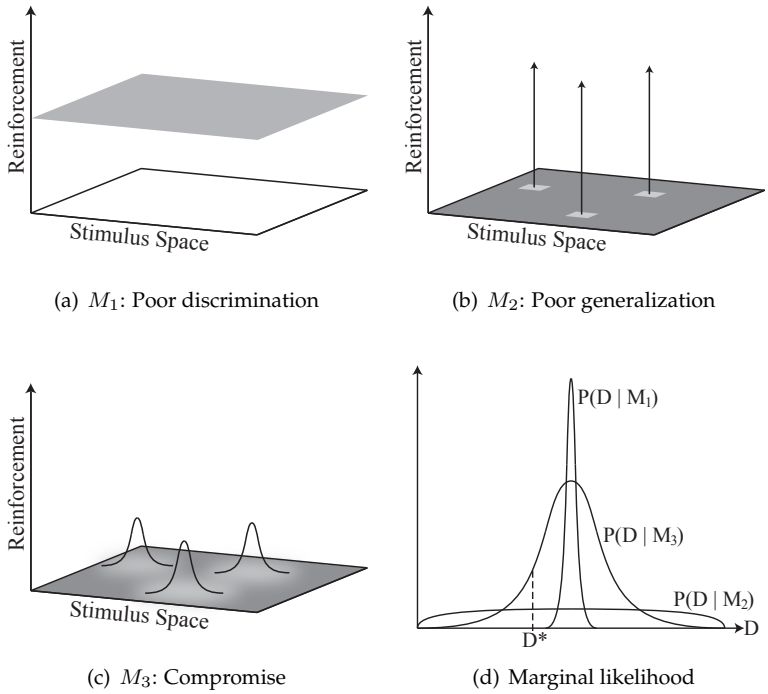
Each of M_1, M_2 and M_3 may be understood as a version of the Pearce configural model with a particular choice of the generalization rule. M_1 confers complete generalization between all stimulus compounds without regard to the degree of overlap; M_2 confers no generalization between any stimulus compounds; and M_3 corresponds to a generalization rule similar to that advanced by Pearce, allowing for generalization between similar stimulus patterns while still permitting discrimination between perceptually distinguishable stimulus compounds.

Of course, neither M_1 nor M_2 would fare particularly well interacting with a complex and dynamic world. Consequently, neither scheme would be particularly successful as a model of animal behavior. Like the compromise model M_3, successful models of configural conditioning exist as points on the continuum between the extremes of M_1 and M_2.

Configural conditioning experiments reveal how animals negotiate the tradeoff between models of sufficient complexity to discriminate between different outcomes and models simple enough to afford useful generalizations. Thus an important challenge for the animal is to choose an appropriate level of representational capacity: flexible enough to allow discrimination when required, but restrictive enough to allow for useful generalizations. In building a predictive model of the world, the animal must balance these competing constraints. In this effort the relevance of the centuries old principle of Occam's razor becomes apparent:

> *Pluralitas non est ponenda sine necessitate*—Plurality should not be posited without necessity — William of Occam (1349).

Occam's razor states that one should choose the simplest model consistent with the data. Such a choice would allow a model sufficiently flexible to discriminate stimulus patterns that are reliably differentially predictive of reinforcement while not being so complicated that it hinders generalization. While a compelling guiding principle, Occam's razor does not itself suggest how it might be cast in computational terms. However, by formulating an explicitly probabilistic model, one can exploit a property of Bayesian inference that is in essence an implementation of Occam's razor (MacKay, 1991).



(a) M_1: Poor discrimination (b) M_2: Poor generalization

(c) M_3: Compromise (d) Marginal likelihood

Figure 3.3. An illustration of the effect of model complexity on the ability to generalize and discriminate. (a) A simple hypothetical model that averages the predictions of all stimuli represented in the *Stimulus Space*. The model cannot make distinct predictions of reinforcement for distinct stimulus configurations (b) A hypothetical model of extreme complexity with sufficient degrees of freedom to represent every point in stimulus space and assign them unique associations with reinforcement. Such a model has no ability to generalize experience with certain stimulus patterns to situation with similar but distinct stimulus patterns. (c) A compromise model with some ability to generalize reinforcement predictions to similar patterns of stimuli and some ability to discriminate between less similar stimulus patterns. (d) A cartoon of the marginal likelihoods of M_1, M_2 and M_3.

To understand how Bayesian inference implements Occam's razor, it is necessary to wade into the mathematical formalism of probability. Consider again the set of models introduced above: M_1, M_2, and M_3. Cast in the formalism of probability theory, these models require a likelihood function that specifies the probability of the training data $\mathcal{D}$, $P(\mathcal{D} \mid \theta_i, M_i)$, conditioned on the value of the model parameter θ_i for each $i \in \{1, 2, 3\}$. In selecting a model from among these three to represent a given data set, the goal is to recover $P(M_i \mid \mathcal{D})$, the posterior probability distribution over the models. Following Bayes's rule, $P(M_i \mid \mathcal{D}) \propto P(\mathcal{D} \mid M_i)P(M_i)$. The model prior $P(M_i)$ expresses a prior preference over the models and the term $P(\mathcal{D} \mid M_i)$ is known as the marginal likelihood. The marginal likelihood is the Bayesian mechanism implementing Occam's razor. Even with equal

prior probability mass assigned to each model, the marginal likelihood has the effect of conferring greater probability mass to the simplest model that is capable of describing the data.

The marginal likelihood is computed by integrating the likelihood over all possible values of the model weights:

$$P(\mathcal{D} \mid M_i) = \int P(\mathcal{D} \mid \theta_i, M_i) p(\theta_i \mid M_i)\, d\theta_i. \tag{3.4}$$

Owing to the requirement that the marginal likelihood be normalized, complex models that are flexible enough to describe a great variety of data sets $\mathcal{D}$ necessarily must spread the probability mass over each possible $\mathcal{D}$ resulting in less mass per data set. Figure 3.3(d) shows an illustration of the marginal likelihood for the models M_1, M_2 and M_3. The complex model M_2 can emulate any arbitrary function mapping stimulus patterns to reinforcement predictions and so must assign relatively little probability mass to any one set. The simple model M_1 can emulate relatively few functions mapping the stimulus pattern to reinforcement prediction. However for the few data sets for which the simple offers a good description—where all stimulus patterns lead to the same prediction of reinforcement—the marginal likelihood assigns relatively large probability mass. The marginal likelihood of the compromise model M_3 is found between the extremes of M_1 and M_2. For moderately complicated data sets such as $\mathcal{D}^*$ that are not well described by M_1 but are not so complicated that the exaggerated flexibility of M_2 is required, M_3 possesses the greatest marginal likelihood.

An interesting consequence of Bayesian inference over models is that as the amount of data increases, Bayesian inference will tend to place more probability mass over more complicated models. In the context of a conditioning paradigm, this implies that as the number of observed trials grows, probability amasses over more complex models. I suggest that evidence in favor of such a growth in complexity with data is commonplace in the classical conditioning literature and I explore some of this evidence in the context of the latent cause theory in Sections 3.6 and 3.7.

3.4. A latent cause model of configural conditioning

In this section I present a latent cause model of configural conditioning and related phenomena. To render model inference more manageable, time is explicitly removed from the model. Trials are treated as singular events without any sense of a temporal order of stimuli within a trial. Temporal order of trials themselves is also removed, as the trials are assumed to be independent, identically distributed (IID) samples from an unknown, but

stationary distribution. The stimuli are represented as simple binary variables encoding their presence or absence on a trial.

The specific formalism chosen to represent the relationship between the stimuli and the latent cause factors is the sigmoid belief network. Sigmoid belief networks are an attractive framework for models of conditioning. They offer a simple additive relationship between multiple latent causes on the probability of observing a stimulus. This allows us to capture summation effects often observed in classical conditioning (Rescorla, 2003a; Rescorla & Coldwell, 1995; Rescorla, 1997) (but also see (Pearce & George, 2002)). The sigmoid belief networks also admit negative weights between the latent causes and the stimuli, permitting us to encode negative associations (or anti-correlations) between stimuli including reinforcement.

Note that the theory is not necessarily tied to the use of sigmoid belief networks. They are by no means the only family of models that would suffice for these purposes. For example, the noisy-OR model (Saul, Jaakkola, & Jordan, 1996) shares many characteristics in common with the sigmoid belief network (though correlations or associations are not expressible with the noisy-OR model).

3.4.1. Sigmoid Belief Networks

Consider a binary random variable, y_j, representing the jth observable stimulus. Here, reinforcement is included as a possible observable stimulus. The value $y_j = 1$ is interpreted as the jth stimulus being present on a given trial, and $y_j = 0$ to represents its absence. I model the latent causes with an additional set of binary random variables, x_i. The ith cause is present (or active) on a given trial when $x_i = 1$, and absent (or inactive) when $x_i = 0$.

A simple and expressive way to model the influence of causes on the stimuli is as a log-linear model, where the log odds of the value of y_j is modeled as a linear combination of the causes:

$$\log\left(\frac{P(y_j = 1 \mid x)}{P(y_j = 0 \mid x)}\right) = \sum_i w_{ij} x_i + w_{0j}. \tag{3.5}$$

The weight, w_{ij}, represents the degree of influence cause x_i has over stimulus y_j. The term w_{0j} is the bias and encodes the log odds in the absence of all causes.

Setting $P(y_j = 0 \mid x) = 1 - P(y_j = 1 \mid x)$ results in the following expression for the probability of observing the jth stimulus,

$$P(y_j = 1 \mid x, w) = \frac{1}{1 + \exp(-\sum_i w_{ij} x_i - w_{0j})} \tag{3.6}$$

Equation 3.6 describes the conditional probability of $y_j = 1$ conditioned on values of all causes, represented by the vector x. The sigmoid belief networks I consider possess two

layers of units, with a layer of observable stimuli and a single layer of latent causes. Within this family of latent cause structures, I consider all models up to some maximum number of latent causes, depending on the experiment; each cause has links potentially to any and all stimulus variables.

Conditioned on the set of causes, the stimuli are all independent. Therefore, the probability of a pattern $\mathcal{S}$ of multiple stimuli on a trial, given the cause activations, x, is simply the product of the probabilities of the individual stimuli,

$$P(\mathcal{S} \mid x, w, m) = \prod_{j \in \mathcal{S}} P(y_j \mid x, w, m) \prod_{j \notin \mathcal{S}} (1 - P(y_j \mid x, w, m)), \tag{3.7}$$

where the notation $j \in \mathcal{S}$ indicates all stimuli that are present in the stimulus pattern $\mathcal{S}$.

According to the sigmoid belief network model, the prior probability of a cause being active is similarly described by a sigmoid function.

$$P(x_i = 1) = \frac{1}{1 + \exp(-w_{i0})},$$

where the weight w_{i0} is another bias term and encodes the propensity of x_i to be active. Since the causes are *a priori* independent, the probability of the vector x is given by a product of the probabilities of the activations of the individual causes: $P(x \mid w, m) = \prod_i P(x_i \mid w, m)$. I model the latent causes as independent.

3.4.2. Inference over the latent causes

Generalization between observed stimulus patterns is a key aspect of both the added elements Rescorla-Wagner model and Pearce's configural model. I now describe how generalization arises in latent cause theory. Given a particular belief net structure m, weight vector w, and the presentation of a stimulus pattern $\mathcal{S}$, one can recover a probability distribution over the activations of the latent causes as a straight forward application of Bayes' rule:

$$P(x \mid \mathcal{S}, w, m) \propto P(\mathcal{S} \mid x, w, m) P(x \mid w, m). \tag{3.8}$$

The effect of Bayesian inference is to weight particular settings of the latent causes proportionally to the likelihood that they would give rise to the observed stimuli.

According to the theory, the degree of conditioned responding is an observable measure of the inferred probability of imminent reinforcement. Therefore the goal is to determine the probability of reinforcement R conditioned on both the training set $\mathcal{D}$ and the observed probe stimulus pattern $\mathcal{S}$ (excluding R). For a given model structure m and setting of the weights w, this quantity can be computed by summing over the possible settings

of the latent causes:

$$P(R \mid \mathcal{S}, w, m) = \sum_x P(R \mid x, w, m) P(x \mid \mathcal{S}, w, m) \tag{3.9}$$

Here I am averaging over the possible activities of the latent causes, taking their full distribution into account. If two settings of the causes are equally probable, they will have equal influence on determining $P(R \mid \mathcal{S}, w, m)$.

As previously mentioned, in the interest of computational tractability I make the rather strong assumption that the training data $\mathcal{D}$ consists of a set of independent trials, with the kth trial consisting of stimulus pattern d_k. For a fixed model, the resulting probability of $\mathcal{D}$ is given by

$$P(\mathcal{D} \mid w, m) = \prod_k P(d_k \mid w, m), \tag{3.10}$$

where the elements of the product are, in turn, given by

$$P(d_k \mid w, m) = \sum_x P(d_k \mid x, w, m). \tag{3.11}$$

Here I have again marginalized over the possible values of the latent causes.

The process of inference over the latent causes is a counterpart to Pearce's generalization rule for configural units. Unlike Pearce's rule, inference over x considers settings of the individual causes x_i jointly (allowing for *explaining away* effects (Dayan & Kakade, 2001)) and incorporates prior probabilities over each cause's activation. Nevertheless, the new rule broadly resembles its predecessor in that a cause is judged likely to be active (and contributes to predicting R) if the constellation of stimuli it predicts is similar to what is observed. This process may also be loosely interpreted as implementing a computational version of the comparator hypothesis (Miller & Matzel, 1988) as reviewed in Chapter 2. Philosophically, the two approaches are similar. Both theories suggest that in many conditioning phenomena, diminished responding resulting from cue competition effects (overshadowing, blocking and relative validity effects) can be the result of performance deficits rather than acquisition deficits (Arcediano et al., 2004).

3.4.3. Implementing Occam's razor: inference over models

The process of conditioning is modeled as the subject's attempt to recover, through Bayesian inference, the generative process that gave rise to the training data. In section 3.3 I discussed how Bayesian inference over the parameters of multiple models implements a preference for simpler models (a mechanism known as the automatic Occam's razor). I now consider what is involved in performing Bayesian inference over a family of sigmoid belief networks.

The model structure m and corresponding weights that gave rise to the training data are not known. Consequently these quantities are treated as uncertain and subject to standard Bayesian inference. I assume that, given a model structure, the weights are *a priori* mutually independent, with each distributed according to a Laplace distribution. The Laplace prior distribution is given by $p(w_{ij}|m) = \frac{1}{2b}e^{-|w_{ij}|/b}$ (in the simulations $b = 2$). As a prior, it encodes a bias for sparseness consistent with a preference for simpler model structures. The bias weights are fixed. For the simulations presented in Section 3.6 were simulated with $w_{0j} = -6$. The simulations presented in Section 3.7 were simulated with $w_{0j} = -12$. In many cases, the alternative values were used to render the phenomena more readily apparent. To the extent that such an abstract model may be interpreted quantitatively, the change in the bias could reflect a change in some underlying property of the stimuli that might impact on its salience, such as its duration or intensity.

As I discussed in the previous section, one can integrate over the values of the weights and thereby compute the marginal likelihood,

$$P(\mathcal{D} \mid m) = \int P(\mathcal{D} \mid w, m)p(w \mid m)\, dw, \tag{3.12}$$

which is then used in conjunction with Bayes' rule, once again, to recover the posterior probability over the family of sigmoid belief network model structures under consideration, $P(m \mid \mathcal{D}) \propto P(\mathcal{D} \mid m)P(m)$. The prior over models, $P(m)$ is expressed as a distribution over n_x, the number of latent variables, and over l_i, the number of links between the stimuli and the ith latent variable: $P(m) = P(n_x) \prod_{i=1}^{n_x} P(l_i)$. I assume that $P(n_x)$ and each $P(l_i)$ are given by geometric distributions ($\mathrm{param.} = 0.1$), renormalized to sum to unity over a maximum of eight latents and eight stimuli. This prior reflects a bias against complex models.

To produce a final conditioned response prediction, one can again average over all models, weights and structures. The probability of R given a test stimulus pattern $\mathcal{T}$ is given by:

$$P(R \mid \mathcal{T}, \mathcal{D}) = \sum_m \int P(R \mid \mathcal{T}, w, m)p(w, m \mid \mathcal{D})\, dw. \tag{3.13}$$

Progressively conditioning on experience to resolve prior uncertainty in the weights and model structure produces a gradual change in predictions akin to the incremental learning rules of previous models. The extent to which a particular model structure m participates in predicting R in Equation 3.13 is, by Bayes' rule, proportional to its prior probability, $P(m)$, and to the extent that it explains the data, $P(\mathcal{D} \mid m)$. Thus a prior preference for simpler models competes against better data fidelity for more complex models.

As data accumulate, the balance shifts toward the latter, and predictions become more accurate. Analogously, weights are small *a priori* but can grow with experience.

Together with the generalization effects discussed above, these inference effects explain why animals can learn more readily to discriminate stimulus compounds that have less overlap. Key to the discrimination is inferring that different compounds are produced by separate latent causes; the more the compounds overlap, the more accurately will the data be approximated by a model with a single latent cause (preferred *a priori*), which biases the complexity-fidelity tradeoff toward simplicity and delays acquisition.

3.5. Monte Carlo integration

In order to determine the predictive probability of reinforcement, Bayesian model averaging necessitates the evaluation of Equation 3.13. Unfortunately, this integral is not amenable to analytic solution. However, if it were possible to sample from the posterior $p(w, m \mid D)$, with samples $\{w^{(k)}, m^{(k)}\}$, then by the law of large numbers:

$$\frac{1}{K}\sum_{k=1}^{K} P(R \mid \mathcal{T}, w^{(k)}, m^{(k)}) \xrightarrow{p} P(R \mid \mathcal{T}, \mathcal{D}), \tag{3.14}$$

that is, the sample mean of $P(R \mid \mathcal{T}, w, m)$ computed from K samples of $P(w, m \mid \mathcal{D})$ converges in probability to $P(R \mid \mathcal{T}, \mathcal{D})$.

Acquiring samples from $p(w, m \mid D)$ is complicated by the need to sample over a union of parameter spaces of different dimensions representing the various latent cause structures under consideration. A solution to this problem is provided by the reversible jump Markov chain Monte Carlo (MCMC) method (Green, 1995). It is important to point out that I am not suggesting that reversible jump MCMC is implemented in the brains of animals. Consistent with the principle of rational analysis, I am making a claim regarding the nature of the computation that animals perform. I claim that they are engaged in recovering an unknown generative model of events in a manner consistent (or approximately so) with Bayesian inference. I am making no claims here as to how they carry out such a computation, though I do include some speculation on this subject in Chapter 6.

3.5.1. Reversible jump Markov chain Monte Carlo

Markov chain Monte Carlo methods[2] are efficient sampling schemes, well adapted to sampling from distributions in relatively large spaces, but usually fixed-dimensional spaces. The problem with applying MCMC methods to sampling over a union of parameter spaces of varying dimension was addressed by Green (1995). As a solution, he

[2]See (Robert & Casella, 1999) for a good reference on MCMC methods.

introduced the reversible jump MCMC method where in addition to the more traditional intradimensional MCMC updates, he proposed transdimensional updates that take the Markov chain from a model of one dimension to a model of a different dimension. To ensure that the samples are indeed drawn from $p(w, m \mid \mathcal{D})$ as desired, it is sufficient to ensure that detailed balance[3] (with respect to $p(w, m \mid \mathcal{D})$) is satisfied by both intradimensional updates and transdimensional updates. As noted by Green, the transdimensional updates require some special machinery to be assured of detailed balance.

Consider a jump from model $\{w, m\}$ to model $\{w', m'\}$. Green imposes a dimension matching condition, where the spaces of $\{w, m\}$ and $\{w', m'\}$ are augmented to $\{w, m, u\}$ and $\{w', m', u'\}$ respectively, such that the mapping between $\{w, m, u\}$ and $\{w', m', u'\}$ is a bijection:

$$(w', m', u') = T(w, m, u). \tag{3.15}$$

The random variables u and u' are designed to bring the two models to a comparable dimensional space, with either u or u' often being null. With the mapping thus defined, the probability of accepting the jump from $\{w, m\}$ to $\{w', m'\}$ is $r(\{w, m\}, \{w', m'\})$ with

$$r(\{w, m\}, \{w', m'\}) = \min\left(1, \frac{f(\{w', m'\})q(\{w', m', u'\}, \{w, m, u\})}{f(\{w, m\})q(\{w, m, u\}, \{w', m', u\})} \left|\frac{\partial T(w, m, u)}{\partial(w, m, u)}\right|\right), \tag{3.16}$$

where $q(a, b)$ is the probability of proposing the jump from a to b, $f(\dot{;})$ is the target (posterior) distribution, and the last term is the determinant of the Jacobian of the transdimensional mapping. The ratio may be given a rigorous measure theoretic definition as the ratio of Radon-Nikodym derivatives with respect to a dominating measure (a measure spanning the joint space of the origin and target models).

Applying the reversible jump MCMC method to the problem of sampling from $p(w, m \mid \mathcal{D})$ for a family of sigmoid belief networks, the algorithm proceeds by iteratively and stochastically choosing from among three possible updates. These are: add or remove a link from cause to stimulus, add or remove an entire cause including all its links, and finally sample a new value of w, leaving m unchanged (the intradimensional update).

1. Add or remove a link: A latent cause, x_i, and a stimulus, y_j, are chosen uniformly at random and an addition or removal of a link between them is proposed, each with probability 0.5. If an addition is proposed, the weight w'_{ij} is sampled from $g(w'_{ij}) = \mathcal{N}(0, 9)$. If the link already exists, then this update does not change the dimension of the parameter space and the weight is resampled from $g(\cdot)$. The probability of accepting the link addition

[3] A Markov chain with transition kernel (proposal distribution) K satisfies the detailed balance condition if there exists a function f satisfying $K(y, x)f(y) = K(x, y)f(x)$. If such an f exists, it is the invariant distribution of the chain (Robert & Casella, 1999).

update with the number of proposed links $l_i' = l_i + 1$ (one plus the number of existing links). is $\min\{1, r_l(\{w,m\},\{w',m'\})\}$ with

$$r_l(\{w,m\},\{w',m'\}) = \frac{P(\mathcal{D} \mid w',m')}{p(\mathcal{D} \mid w,m)} \times \frac{P(l_i')p(w_{ij}')}{P(l_i)g(w_{ij}')} \tag{3.17}$$

The acceptance probability of a removal is simply provided by $\min\{1, r_l(\{w,m\},\{w',m'\})^{-1}\}$ with the substitution $l_i' = l_i - 1$.

2. Add or remove a latent cause: Let c be the current number of latent causes with $c_{\max}$ being an upper bound on c. A choice is made to propose the addition of a latent cause with probability q_c or to propose the removal of an existing latent cause with probability $1 - q_c$ ($q_c = 0.5$ for $0 < c < c_{\max}$, $q_0 = 1$ and $q_{c_{\max}} = 0$). Links are added to the proposed latent cause with probability 0.2 for each link, with weights sampled from $g(\cdot)$. The latent cause bias, w_{c_i}, is sampled from the prior. The probability of accepting an addition move such that $c' = c + 1$, is $\min\{1, r_c(\{w,m\},\{w',m'\})\}$ with

$$r_c(\{w,m\},\{w',m'\}) = \frac{P(\mathcal{D} \mid w',m')}{P(\mathcal{D} \mid w,m)} \times \frac{P(c+1)P(l_i')\prod_{j=1}^{l_i'} p(w_{ij}')(1-q_{c+1})}{P(c)(c+1)q_c\prod_{j=1}^{l_i'} 0.2g(w_{ij}')} \tag{3.18}$$

The acceptance probability of a latent cause removal is $\min\{1, r_l(\{w,m\},\{\mathbf{w}'_{m'},m'\}))^{-1}\}$, with appropriate substitutions.

3. Update weights: A candidate parameter vector, w, of the same dimension as the current parameter vector is sampled from a Gaussian with isometric covariance structure $\mathcal{N}(w, \sigma_{w_{ij}})$. For all $j > 0$, $\sigma_{w_{ij}} = 0.3$, and $\sigma_{w_{i0}} = 0.05$. The probability of accepting the move is simply $\min\{1, r_w(\{w,m\},\{w',m'\})\}$ with r_w equal to the ratio of posteriors, $p(w',m' \mid \mathcal{D})/p(w,m \mid \mathcal{D})$. This is simply a standard Metropolis-Hastings update.

3.5.2. Exchange MCMC

One serious drawback of the reversible jump MCMC method is that it is subject to poor mixing properties due to the typically low acceptance rates for the transdimensional jumps. In an effort to alleviate this problem, I turned to exchange MCMC, a strategy known to boost MCMC mixing in applications with multi-modal target distributions. Inspired by simulated annealing techniques, exchange MCMC enables fast mixing between modes of a target distribution through the coupling of parallel Markov chains (Iba, 2001). See (Iba, 2001) for an excellent exposition of exchange MCMC and other extended ensemble MCMC methods.

Consider the extended system,

$$\hat{f}(w,m) = \prod_{n=1}^{N} f_n(w,m) \qquad \text{with} \qquad f_n(w,m) = p(w,m \mid \mathcal{D})^{\frac{1}{\lambda_n}}, \tag{3.19}$$

with temperature parameters λ_n given by $\lambda_1 = 1 \leq \lambda_2 \leq \cdots \leq \lambda_N$. In the problem at hand, each $f_n(w, m)$ is the target distribution ($p(w, m \mid \mathcal{D})$) of a reversible jump Markov chain, each at a different temperature. One can sample from the extended system, $\hat{f}(w, m)$, by simultaneously sampling from the N Markov chains. In addition to the intra-chain reversible jump updates already described, I introduce exchange moves between neighboring chains n and $n+1$. Exchange moves involve swapping states between neighboring chains, $\{w', m'\}_n = \{w, m\}_{n+1}$ and $\{w', m'\}_{n+1} = \{w, m\}_n$. The probability of accepting the exchange move is given by $\min\{1, r_e\}$,

$$r_e = \frac{f_n(\{w', m'\}_n) f_{n+1}(\{w', m'\}_{n+1})}{f_n(\{w, m\}_n) f_{n+1}(\{w, m\}_{n+1})} = \frac{f_n(\{w, m\}_{n+1}) f_{n+1}(\{w, m\}_n)}{f_n(\{w, m\}_n) f_{n+1}(\{w, m\}_{n+1})}. \tag{3.20}$$

The resulting extended Markov chain satisfies detailed balance with respect to $\hat{f}(w, m)$.

The samples drawn from chain n correspond to the distribution $f_n(w, m)$. So to recover the original distribution of interest, $p(w, m \mid \mathcal{D})$, one simply needs to consider the samples from the chain corresponding to $f_1(w, m)$. The exchange moves permit the propagation of states from high temperatures, where fast relaxation can facilitate mixing to lower temperatures where we are sampling from the distribution of interest. In the present context the exchange MCMC method greatly increases the mixing rate of the chain.

3.6. Simulations I: configural conditioning

In this section I point to experiments that support my claim that animals approximate Bayesian inferences over latent cause model structures. My goal is to highlight the role of the evidence/complexity tradeoff in the behavior. To this end, I consider a range of experimental phenomena including configural conditioning but also extending beyond it to show how these same considerations provide a parsimonious explanation of a wide range of animal behavior.

Each point in the learning curves presented in this section represents an entire run of the MCMC sampler with the training data consisting of the trials up to that point. For example, the simulation at the 10th trial consists of a simulation of the first 10 trials considered as a batch. The next point (at the 11th trial) is an independent MCMC simulation run on the first 11 trials, again considered as a batch. The simulations all involved MCMC runs of at least 1×10^6 samples from which 1×10^4 were used in Equation 3.14.[4] The number of parallel exchange MCMC chains that were used depended on the simulation, with larger latent cause models requiring more chains. There were never less than 10 chains and never more than 40.

[4]Subsampling in this way is common to reduce the correlations between the samples.

Before delving into the latent cause account of traditional configural conditioning phenomena, I briefly present the model's explanation of two cue competition effects: overshadowing and blocking. Inference over model structure plays a minimal role in the account of these phenomena, allowing for focus on the effects of inference over the values of latent causes.

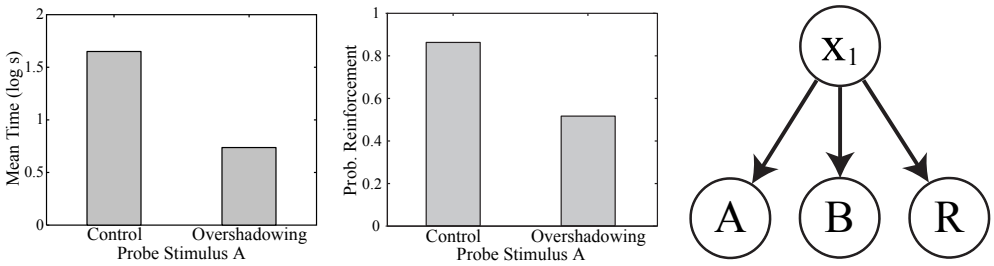
3.6.1. Overshadowing

In the standard overshadowing procedure (Kamin, 1967), two stimuli are repeatedly presented together followed by reinforcement, $AB+$. Overshadowing is said to occur if conditioned responding to one of the stimuli, A is significantly less than if it were paired alone with reinforcement, $A+$. In this case B is said to have overshadowed A.

Figure 3.4(a) illustrates the experimental effect of overshadowing. Responding to A in the overshadowing condition (trained with $AB+$) is reduced relative to the control condition (trained with $A+$). Figures 3.4(b) and 3.4(c) show the results of simulating overshadowing and the corresponding inferred maximum *a posteriori* (MAP) model structure. The MAP model structure corresponds to the model structure that is most often sampled from the posterior distribution over models.

The reduced responding effect is captured via inference over the values of the latent causes (Equation 3.11). After training, the model structure shown in Figure 3.4(c) possesses the majority of the probability mass, with links emanating from a single cause. Here the latent cause is simply encoding the observed co-occurrence of A, B and the reinforcer R. On testing, presenting the test stimulus B in isolation is interpreted as $\bar{A}B$ (where $\bar{A}$ indicates the absence of A) and the absence of A constitutes evidence against activation of the latent cause by equation 3.9. Relative to control, this results in an overall decrease in the probability of activation of latent cause x_1 and hence a decrease in the probability of reinforcement.

Often the saliencies of the two stimuli differ (Pavlov, 1927), but this need not be the case. While the role of saliency in overshadowing has not been considered here, it may be represented in the model as a different setting of the bias term, w_{0j}. As mentioned previously, this bias term encodes the prior belief of observing that stimulus in the absence of any latent cause. As the prior probability of observing the overshadowed stimulus B in the absence of a cause increases, the probability of activation of the x_1 unit (shown in Figure 3.4(c)) decreases through the inference mechanism of equation 3.9. An important advantage of the Bayesian approach to models of conditioning is that it relates somewhat



(a) Experiment (b) Simulation (c) Model structure

Figure 3.4. Overshadowing experiment training with $AB+$ and testing with the probe stimulus A. (a) The experimental phenomenon of overshadowing (adapted from (Blaisdell et al., 1998)). (b) The simulation of the latent cause theory demonstrating overshadowing. (c) The maximum *a posteriori* (MAP) model structure resulting from training with $AB+$.

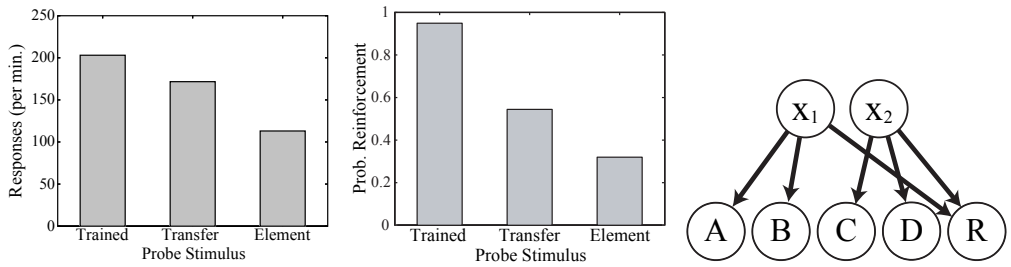
vague notions such as saliency with quantities—such as the probability of observing a stimulus—that are subject to direct experimental manipulation.

3.6.2. Blocking

The typical procedure for obtaining blocking is divided into two phases. The first phase is a simple acquisition procedure, $A+$. In the second phase the *blocking* stimulus A is once again paired with reinforcement, but this time with an additional *blocked* stimulus B: $AB+$. Typically A and B are presented simultaneously and immediately preceding reinforcement. After training the subject demonstrates significantly reduced response to B compared to control conditions that do not include the $A+$ pre-training (Kamin, 1967).

The theory's account of blocking is similar to the account of overshadowing. Without excessive training, a single latent cause is inferred to have given rise to both phases of training. The MAP model structure is identical to that in the overshadowing case (Figure 3.4(c)). The weight associating the cause and stimulus A is greater, on average, than the weight associating the target B with the cause. Due to the greater affinity between the cause and A relative to B, inference leads to a reduced probability of activation of the cause on a probe trial of B (in the absence of A) and consequently leads to a reduction in the predicted probability of reinforcement.

Because of the assumption that all trials are independent, the model is unable to capture trial order effects. In the case of blocking, this amounts to assuming that subjects experience the trial types interleaved rather than in two distinct phases of training. While this is likely not capturing all aspects of phenomena such as blocking, our basic account would not change if trial order effects were incorporated (see e.g. Dayan and Kakade (2001)).



(a) Experiment (b) Simulation (c) Model structure

Figure 3.5. The trained compounds experiment of Rescorla (2003a), where subjects were trained with trials of $AB+$ and $CD+$ and tested with the trained compounds (AB and CD), the transfer compounds (AD and BC) and the stimulus elements (A,B,C and D). (a) The experimental data (adapted from Rescorla (2003a)). (b) The latent cause model simulation of the trained compounds experiment. (c) The latent cause model MAP model structure as measured from simulation.

3.6.3. Summation

I now turn to a set of configural conditioning experiments. These are the principal phenomena that the added elements Rescorla-Wagner and Pearce's configural conditioning model set out to explain. I begin the exploration of the Bayesian latent cause account of configural conditioning phenomena with the experiment of Rescorla (2003a). The simulation involves 10 trial blocks, each with one trial of $AB+$ and one of $CD+$. Figure 3.5 shows the results of the simulation together with the experimental data. The theory captures the basic trend apparent in the empirical data, with greatest responding to the original trained compound, moderate responding to the transfer compounds (AD or BC) and least responding to the individual elements.

Figure 3.5(c) shows the inferred MAP model structure for this experiment. The MAP model possesses two latent causes, one associated with each training trial type. The pattern of predicted probability of reinforcement as shown in Figure 3.5(b) is understood as another instance of generalization through inference over the latent variables. On presentation of a trained compound, the latent cause associated with that compound is inferred to be active with high probability. Presentations of a single element result in a reduced probability of reinforcement by the same inference mechanism that brought about overshadowing and blocking. On presentation of the transfer compound, with one element from each of the trained compounds, both latent causes are inferred to have moderate probabilities of activation. The two causes interact additively to produce an increased probability of reinforcement relative to the individual elements. In this case the additive interaction

of two latent causes was insufficient to completely overcome the generalization decrement resulting from inference over the latent causes.

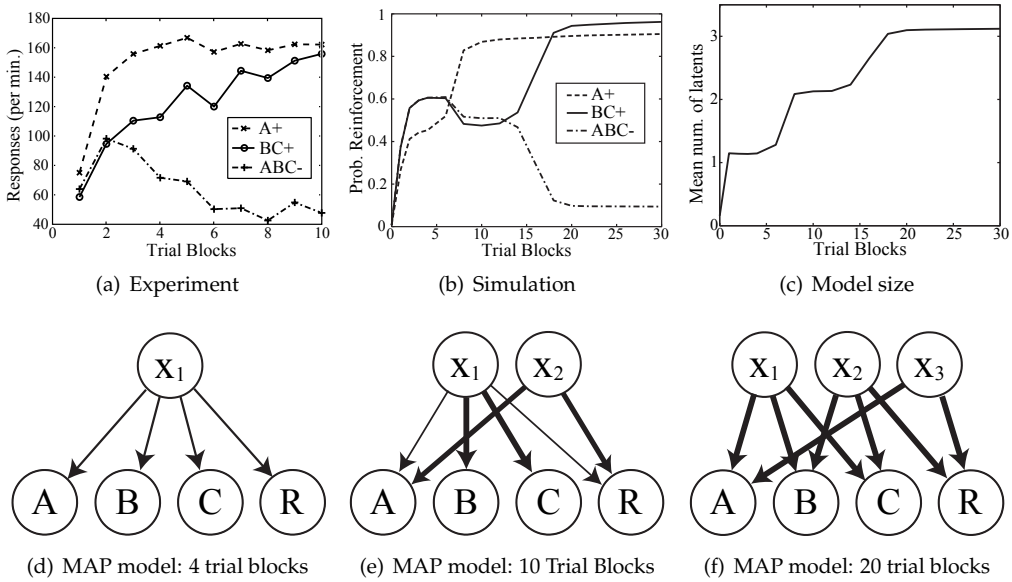
3.6.4. Asymmetric negative patterning

The second configural conditioning experiment I consider is the asymmetric negative patterning experiment. Redhead and Pearce (1995a) used this experiment to demonstrate that Pearce's configural model correctly predicted that discrimination between A and ABC proceeded more quickly than did discrimination between BC and ABC. Figure 3.6 illustrates the simulation results as a function of number of trial blocks. Each trial block consists of one trial each of $A+$ and $BC+$ and two trials of $ABC-$. As shown in Figure 3.6(b), the theory captures the general order of discrimination, with discrimination between $A+$ and $ABC-$ proceeding more quickly than that between $BC+$ and $ABC-$.

A critical feature of the account of this experiment is the tradeoff between generalization and discrimination negotiated through inference over the model structures and parameters. As the number of training trials increases, posterior probability grows for larger models that are better able to discriminate between the reinforced and the non-reinforced trial types. Figure 3.6(c) shows that the average model size increases as a function of training data. Figures 3.6(d), 3.6(e) and 3.6(f) show the MAP model structures with training on 4, 10 and 20 trial blocks respectively. With few (4) trial blocks, most probability mass is assigned to the single latent cause model, shown in Figure 3.6(d), that is unable to distinguish between any of the trial types. With more training, a model of moderate size emerges as the MAP model structure, as shown in 3.6(e). Here, the $A+$ trial type is visibly encoded in a second latent cause, thus giving rise to the observed discrimination of $A+$ trials with moderate training. After a large amount of training, posterior probability mass principally falls on the three cause model given in Figure 3.6(f), with each encoding one of the three training trial types, thus allowing for discrimination between the reinforced trial types $A+$ and $BC+$ and the non-reinforced trial type, $ABC-$.

The expected probability of reinforcement conditional on observing A is noticeably smaller than on observing either BC or ABC (Figure 3.6(b)). This is a consequence of inference over the latent variable. With a single latent variable with associations to A, B, C and reinforcement, the unit is much more likely to be active on seeing either BC or ABC than on seeing A alone.

Why, after moderate training, did the structure shown in Figure 3.6(e) emerge as the MAP model? Why not a model with a unit representing $BC+$ rather than $A+$? The answer lies at the heart of the account of configural conditioning. With a moderate amount of



(a) Experiment (b) Simulation (c) Model size

(d) MAP model: 4 trial blocks (e) MAP model: 10 Trial Blocks (f) MAP model: 20 trial blocks

Figure 3.6. Asymmetric negative patterning experiment, trained with interspersed trials: $A+$, $BC+$ and $ABC-$. (a) The empirical phenomenon (modified from Redhead and Pearce (1995a)). (b) The latent cause model simulation of the experiment. (c) A plot of the average model size as a function of amount of training. (d)–(f) The MAP model structure for the model simulated with 4, 10 and 20 trial blocks respectively.

training there is still insufficient data to fully counter the combined complexity-penalizing effect of the prior and marginal likelihood of the more descriptive three cause model of Figure 3.6(f). However there is sufficient data to warrant a more detailed account than that offered by the single cause model of Figure 3.6(d). The model of 3.6(e) is inferred to be the most probable because the $A+$ trials share only two elements with the single $ABC+$ cause while the other two trial types, $BC+$ and $ABC-$, each share three elements with the cause encoding $ABC+$. The $A+$ trials types are the least well explained under the single cause model and consequently are the first to be separately explained by the parsimonious model of Figure 3.6(e).

3.6.5. Irrelevant stimulus negative patterning

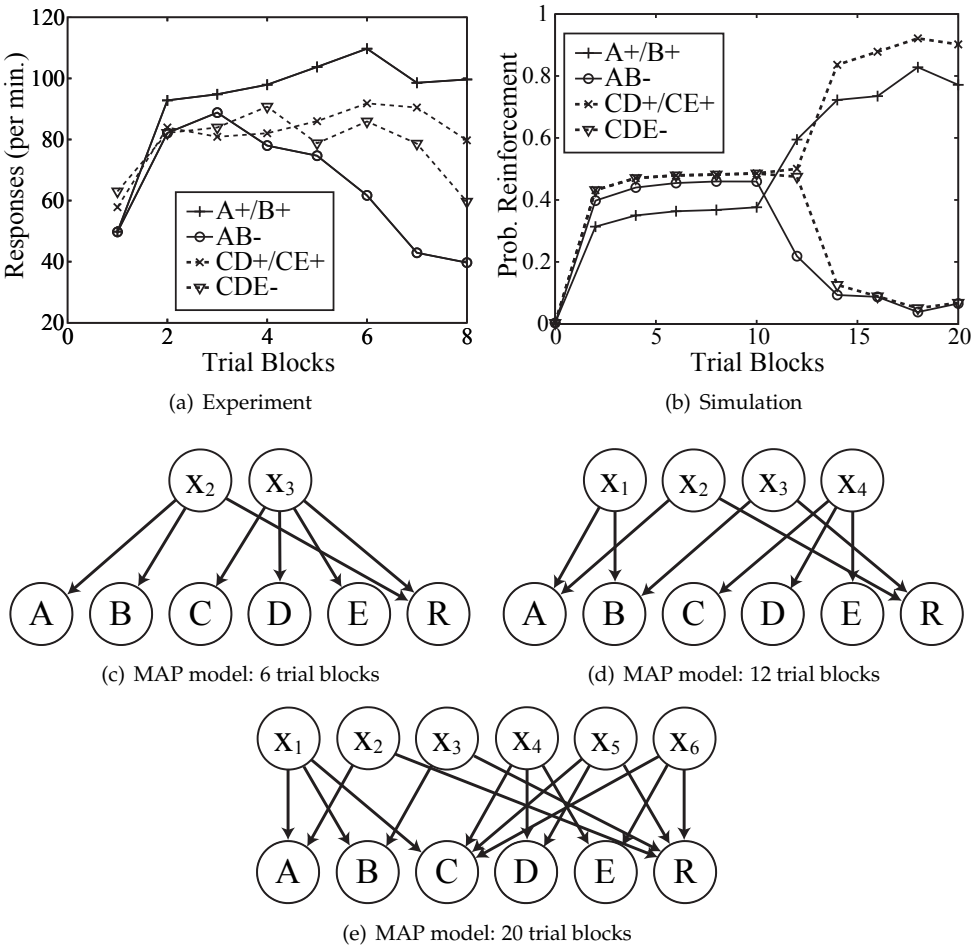
A further investigation of the relative order of discrimination with learning was done by Pearce and Redhead (1993) by adding an irrelevant stimulus to a negative patterning protocol (also see Rescorla (1972) for a variant). This experiment involves training subjects on 6 different trial types: $A+$, $B+$, $AB-$, $CD+$, $CE+$ and $CDE-$. In these simulations the model was trained on a variable number of trial blocks with each block consisting of $[1, 1, 2, 1, 1, 2]$ trials of the respective types.

The empirical results, shown in Figure 3.7(a), reveal the discrimination of $A+$, $B+$ and $AB-$ proceeding more quickly than the discrimination of $CD+$, $CE+$ and $CDE-$. Figure 3.7(b) shows the latent cause model account of this experiment. While the observed pattern of discrimination is captured, with the discrimination of $A+$, $B+$ and $AB-$ occurring in fewer trials than the discrimination between $CD+$, $CE+$ and $CDE-$, the discriminations both occur much more abruptly. As apparent in Figure 3.7(b), there is much less of a distinction between the two discriminations tasks in simulation than is apparent in the experimental data. According to the model, because of how probabilities are combined, the redundant stimulus adds no increased measure of similarity (as measured by inference over the latent causes) among the trials $CD+$, $CE+$ and $CDE-$. More specifically, there is no discriminative advantage, measured in terms of the probability of the observations, to first discriminate the set of non-redundant XOR stimuli. Thus the only difference between these trial types is the addition of a single link between the latent causes and the redundant stimulus and the extra model complexity (relative to the non-redundant trials) is penalized with slower discrimination. As shown in Figure 3.7(d), the MAP model structure after a moderate amount of training encodes the discrimination between $A+$, $B+$ and $AB-$, but does not represent the discrimination between $CD+$, $CE+$ and $CDE-$. The former discrimination arises first because it is the simplest to describe, requiring the fewest number of links.

This is a fairly weak result and it is difficult to know whether this experiment constitutes evidence in favor of the latent cause model or against it. The discrepancy between the animal subjects and the model in the significance of the redundant stimulus suggests one of two things: either animals do not weight the presence and absence of stimuli in the same way as the model, or animals possess priors that give much greater weight to smaller models than that used in these simulations.

3.7. Simulations II: beyond configural conditioning

Latent cause theory effectively redefines the configural unit as encoding a parsimonious representation of the observed pattern of stimulus presentations. The implications for our understanding of classical conditioning are far-reaching. The next few sections explore experimental paradigms that are not traditionally regarded as configural conditioning phenomena. I offer these experiments as further evidence that animals subjected to classical conditioning paradigms are routinely engaged in negotiating the conflict between model complexity and data fidelity.



(a) Experiment (b) Simulation

(c) MAP model: 6 trial blocks (d) MAP model: 12 trial blocks

(e) MAP model: 20 trial blocks

Figure 3.7. Negative patterning experiment with an irrelevant stimulus, trained with interspersed trials: $A+$, $B+$, $AB-$, $CD+$, $CE+$ and $CDE-$. (a) The empirical phenomenon (adapted from Pearce and Redhead (1993)). (b) The corresponding simulation of the latent cause model. (c)–(e) The MAP model structure for the model simulated with 6, 12 and 20 trial blocks respectively.

3.7.1. Second-order conditioning and conditioned inhibition

Second-order conditioning describes the transfer of an excitatory association from one stimulus to another. Typically the procedure consists of two phases. In the first phase, a stimulus A is repeatedly paired with reinforcement, $A+$, in a standard acquisition paradigm. The second phase of training involves a small number of pairings of A with the target stimulus B in the absence of reinforcement, $AB-$. After training, presentation of B alone elicits conditioned responding from the subject—despite B never having been paired

Group	$A+$	$AB-$	$C+$	Test $\rightsquigarrow$ Result	Test $\rightsquigarrow$ Result	
No $AB-$	96	0	8	$B \rightsquigarrow -$	$BC \rightsquigarrow Resp.$	Control
Few $AB-$	96	4	8	$B \rightsquigarrow Resp.$	$BC \rightsquigarrow Resp.$	2nd Order
Many $AB-$	96	48	8	$B \rightsquigarrow -$	$BC \rightsquigarrow -$	Cond. Inhib.

Table 3.1. A summary of some of the experiments of Yin et al. (1994). In this experiment, the reinforcer (+) is a footshock, $A =$ white noise or buzzer sound, $B =$ tone, $C =$ click train is the transfer excitor. The term $Resp.$ indicates a conditioned response to the stimulus

with the reinforcer. Presumably an excitatory association between A and reinforcement is transferred to B through their being paired together.[5]

Conditioned inhibition is an extensively studied phenomenon involving a experimental paradigm strikingly similar to second-order conditioning. While there are numerous procedures that have been reported to lead to conditioned inhibition effects (Savastano, Cole, Barnet, & Miller, 1999), one standard paradigm consists of interleaving the same two trial types used in second order conditioning: $A+$ and $AB-$. Yet in the case of conditioned inhibition training, B comes to suppress responding to a separately trained excitor C and is retarded in acquiring an excitatory associations with further $B+$ training (compared to controls that did not receive the previous phase of training).

Considering the similarities of these two conditioning paradigms, the difference in effect on the target stimulus B is surprising. They both contain the same two basic trial types: $A+$ and $AB-$. In the case of second-order conditioning B acquires a positive association with reinforcement, while in the case of conditioned inhibition training, B acquires an negative association with reinforcement. Yin et al. (1994) explored the dimensions of this two trial type procedure in an effort to distill the essential requirements for each learning phenomenon. While interspersing the two trial types, they were able to shift from second order conditioning to conditioned inhibition simply by increasing the number of $AB-$ presentations. Table 3.1 summarizes the details of the experiment.[6]

In seeking to explain this phenomenon, Yin et al. (1994) suggest that Pearce's configural model offers a possible account. I believe this to be inaccurate. While Pearce's model readily accounts for conditioned inhibition (a negative association forms between the AB configural unit and reinforcement), the model offers no account of second-order conditioning. On presentation of the $AB-$ trials, the positive association of A with reinforcement would force the AB configural unit to learn a negative association with reinforcement. On

[5] In the classical paradigm of second-order conditioning the second-order stimulus B briefly precedes the transfer stimulus A.

[6] Yin et al. (1994) also looked at the influence of sequential versus interspersed trials. We do not present these results.

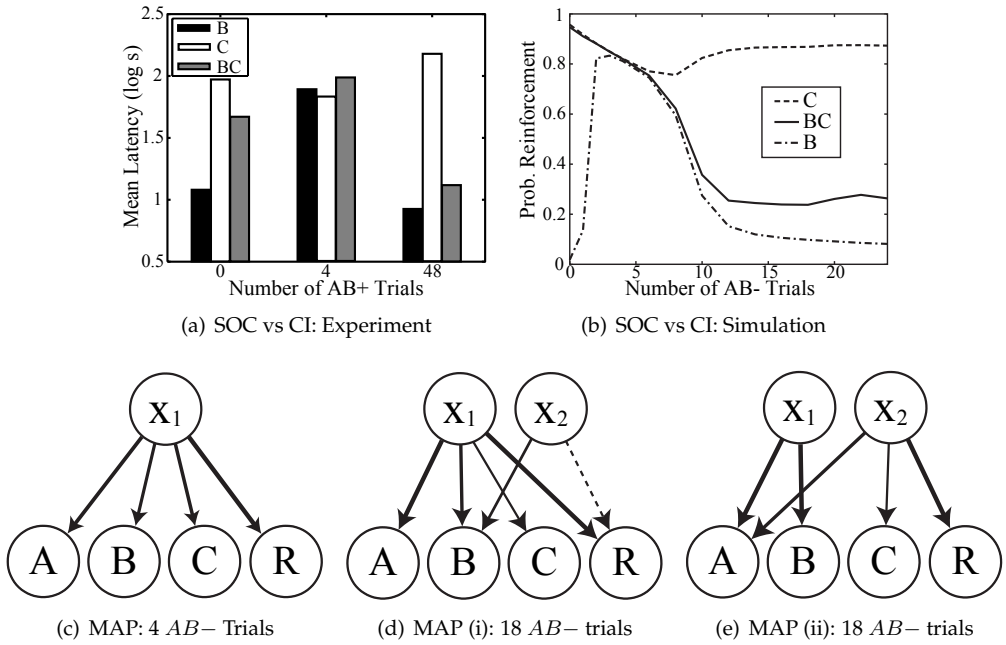
probe presentations of B, the only unit to activate would be the AB configural unit. Thus the Pearce model predicts that conditioned inhibition would occur with any number of $AB-$ trials.

One of the very few models of conditioning that has demonstrated the ability to capture second-order conditioning effects is the temporal difference (TD) model (Sutton, 1988). As discussed in the previous chapter, the TD model apportions associative strength according to relative temporal contiguity with reinforcement. TD accounts for second-order conditioning in the classical examples where the second-order stimulus precedes the excitor in their pairing: $B \rightarrow A$. The positive association of A is transfered to B because B predicts A. However in the experiment of Yin et al., the two stimuli are presented simultaneously and still demonstrate second-order conditioning. TD does not predict second-order conditioning in this situation. Instead it predicts that a negative association between B and the reinforcer (characteristic of conditioned inhibition) begins to form from the first presentation of $AB-$.

Latent cause theory offers an alternative account of the relationship between second-order conditioning and conditioned inhibition. Figure 3.8 illustrates the results of simulating the experiment of Yin et al.. As shown in Figure 3.8(b), simulations from the model demonstrate the key characteristics present in the experimental data: 1) with few $AB-$ trials, B is imparted a positive association with reinforcement; 2) with further training, this association disappears and B begins to demonstrate the summation characteristics of a conditioned inhibitor. Both in the experiment of Yin et al. (1994) and in the simulations, further acquisition training with B also revealed a retardation of acquisition relative to controls reminiscent of a sensory pre-conditioning (latent inhibition) effect and understood to be characteristic of conditioned inhibitors (Savastano et al., 1999).

Figure 3.8(c) gives the MAP model structure with 4 $AB-$ trials and shows how latent cause theory accounts for second-order conditioning. With few $AB-$ trials, the most parsimonious inference is all trials, of both $A+$ and $AB-$ trial type, arise from a single latent cause. A spurious correlation between B and reinforcement is created by the pressure to find a simple account of the data and manifests itself as second-order conditioning.

With more $AB-$ trials, there is greater support for a larger, more complex description of the data and the spurious correlation between B and reinforcement disappears. Figures 3.8(d) and 3.8(e) show two model structures that have significant posterior probability mass with 20 trials of $AB-$. They each offer an alternative account of the training data. The model in Figure 3.8(d) uses a second latent cause to encode a negative correlation between B and reinforcement: whenever B is observed, reinforcement is not observed. The model



(a) SOC vs CI: Experiment (b) SOC vs CI: Simulation

(c) MAP: 4 $AB-$ Trials (d) MAP (i): 18 $AB-$ trials (e) MAP (ii): 18 $AB-$ trials

Figure 3.8. An experiment exploring second-order conditioning and conditioned inhibition, with the training schedule described by Table 3.1. (a) The experimental results (adapted from Yin et al. (1994)). The response is measured as the mean cumulative time (latency) to take 5 licks of water when expecting footshock. It is a measure of a characteristic freezing response. (b) The latent cause model simulations of the experiment. (c)–(e) MAP model structures after (c) 4, (d) and (e) 18 $AB-$ trials. Figures (d) and (e) show the two dominant modes of the posterior probability density, each holding roughly equal proportions of the simulation samples. Dashed lines represent links with negative weights

in Figure 3.8(e) offers a different interpretation: there are two types of independent events, one where $AB-$ is observed and one where $A+$ is observed. Note that the second cause also includes a weak link to the transfer excitor C due to pressures of parsimony similar to that resulting in second-order conditioning. While the second model structure offers an account that is perhaps a more faithful representation of the observed trial types, both have approximately equal posterior probability.

The model of Figure 3.8(d) illustrates perhaps the more orthodox perspective on conditioned inhibition, where a negative association (or correlation) is formed between the conditioned inhibitor B and reinforcement. When the probe trial of BC is presented to the model, both causes are inferred to be (mostly) active. Through the additivity of causes (Equation 3.6), the positive weight emanating from cause x_1 to the reinforcer R is partially canceled by the negative weight emanating from x_2 to R.

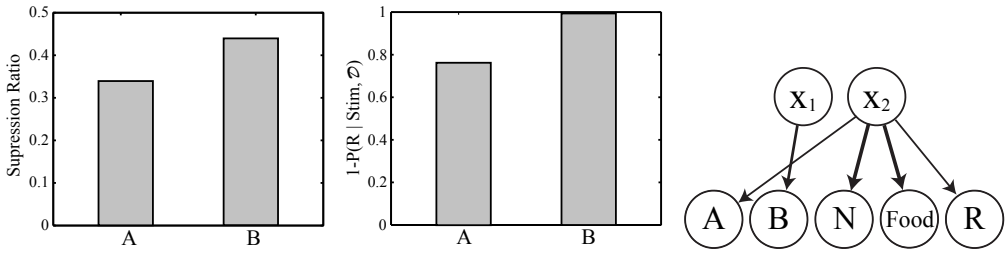
Interestingly, the model shown in Figure 3.8(e) also demonstrates the conditioned inhibition summation effect, albeit an unorthodox one. On probe presentation of BC, inference over the activations of the causes must weigh the presence of B and C against the absence of A. The presence of B and C are each predicted by a separate cause, but both predict the presence of A, and together they strongly predict the presence of A. The absence of A on the probe trial BC constitutes overwhelming evidence against the simultaneous activation of both causes and since B is better explained by cause x_1 than is C by cause x_2 (due to its stronger association), the most probable activation pattern is x_1 is active and x_2 is inactive, resulting in a minimal predicted probability of reinforcement.

3.7.2. Reinterpreting acquired equivalence

Thus far I have been concerned with discrimination and generalization between stimuli that have been paired together. However, the conditioning literature contains many demonstrations of generalization and discrimination between stimuli that have never been explicitly paired. The possibility that discrimination and generalization between stimuli may be mediated by the similarity or distinctiveness of their associative history has been explored by many authors including (Miller & Dollard, 1941; James, 1890; Gibson, 1959; Grice, 1965).

In a particularly thorough set of experiments, Honey and Hall (1989) convincingly demonstrated that the similarity of previous training experience is capable of mediating the degree of generalization between stimuli. In one of their experiments (experiment 3), they presented subjects with one of two training procedures: 1) $A - food$, $B - no\ food$, $N - food$; or 2) $A - no\ food$, $B - food$, $N - no\ food$. After repeated presentations of 1 or 2, the subjects were exposed to revaluation training pairing N with $shock$. As illustrated in Figure 3.9(a), subjects demonstrated greater generalization from N to A, which shared N's training history, than to B.

Figure 3.9(b) shows the simulation results for the acquired equivalence (AE) experiment with the first training procedure (simulation results for the second procedure are comparable). Consistent with the experimental evidence, the simulations demonstrate preferential generalization between stimuli with similar training experience. According to latent cause theory, acquired equivalence is simply another consequence of the pressure to simplify the model resulting from standard Bayesian inference. Stimuli with similar training history are lumped together through a single latent cause because they are most simply explained as having been caused by the same latent factor. Figure 3.9(c) illustrates



(a) AE Experiment (b) AE Simulation Results (c) MAP model structure

Figure 3.9. The acquired equivalence experiment of Honey and Hall (1989). (a) Experimental results showing greater generalization from the $N - shock$ association to A than to B (modified from Honey and Hall (1989)). The suppression ratio is an measure of responding where smaller values (closer to zero) represent greater responding. (b) Simulation results of the acquired equivalence effect, plotted as 1 - Probability of responding to match the suppression ratio measure. Similar to the empirical finding, the simulations demonstrate greater prediction of $shock$ on presentation of A rather than B. (c) MAP model structure shows the mechanism of superior generalization between N and A. The simulations consisted of 4 presentations of each trial type: $A - food$, $B - no\ food$, $N - food$ and $N - shock$.

the MAP model structure resulting from the inference procedure. The model demonstrates the merging of the A and N related trial types due to the overlap of the $food$ stimulus.

In accounting for the acquired equivalence effect, inference over model structures resulted in A and N being associated through a single latent cause, despite the two never co-occurring. Similar to the case of second-order conditioning, they are associated by transitivity. In this case both N and A appear with the unconditioned stimulus $food$ and without excessive training the most parsimonious explanation of these two trial types is that they emerged from a single latent cause. Subsequent revaluation training with $N - shock$ is likewise integrated into the same latent cause with the effect of A becoming associated with $shock$ (as shown in the MAP model structure in Figure 3.9(c)).

More recently Honey and colleagues have extended their investigations of acquired equivalence to study phenomena they refer to as acquired relational equivalence (Honey & Watt, 1998, 1999). They demonstrate that the pattern of generalization results that emerge in basic acquired equivalence effects extend to cases where the common training history is considerably more complex. In their experiment, Honey et al. presented trials of either stimulus A or B together with a stimulus Y and $food$ interspersed with trials of A or B together with Z and $no\ food$. In addition to these four trial types, they also presented four additional trial types with either C or D together with Y and $no\ food$ as well as trials with either C or D together with Z and $food$. Table 3.2 summarizes the experimental procedure. The procedure is designed to maintain equivalent training experience (with respect to Y, Z

Biconditional training phase		A revaluation phase	Test
$AY \rightarrow food$	$AZ \rightarrow no\ food$		
$BY \rightarrow food$	$BZ \rightarrow no\ food$	$A \rightarrow shock$	B vs. D
$CY \rightarrow no\ food$	$CZ \rightarrow food$	$C \rightarrow no\ shock$	
$DY \rightarrow no\ food$	$DZ \rightarrow food$		

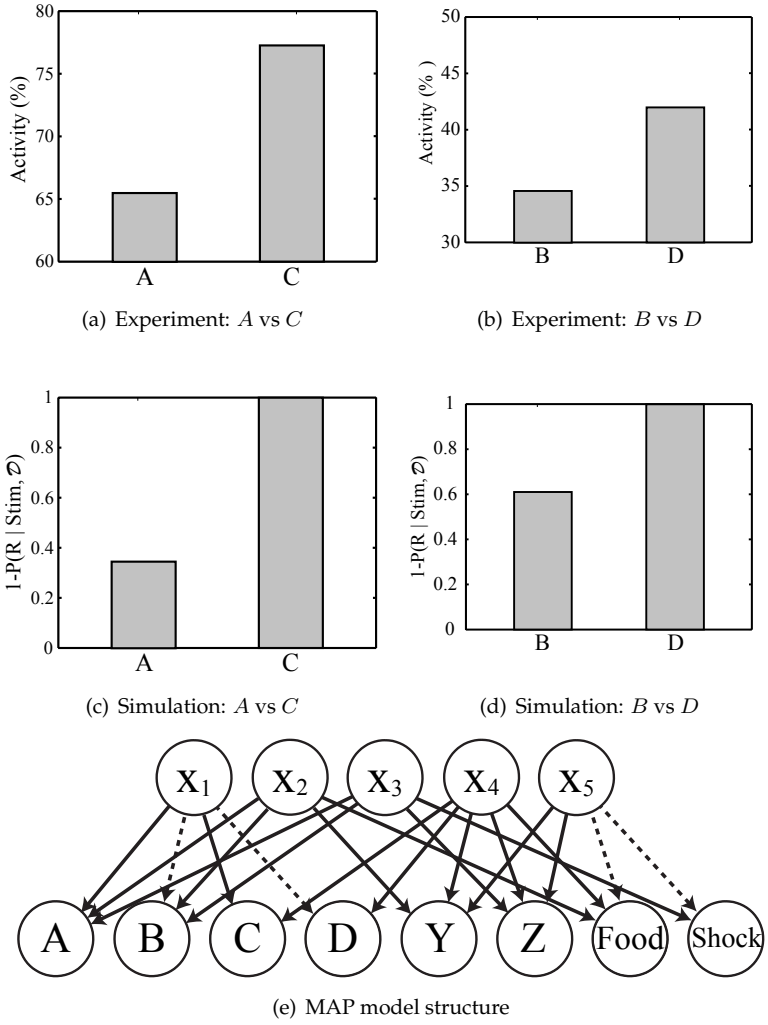
Table 3.2. Summary of the acquired relational equivalence experiment from Honey and Watt (1998) and Honey and Watt (1999).

and $food$) between A and B and similarly between C and D. Finally, a phase of revaluation training pairs A with $shock$ and C with $no\ shock$. Testing probed the question: how does revaluation training with A generalize to B and D? Predictably, generalization from A to B was more pronounced than generalization from A to D in the sense that B generated a greater degree of behavior consistent with an expectation of $shock$. This basic pattern of results has been observed across a number of experimental variations including procedures involving context cues for A, B, C, and D (Honey & Watt, 1999).

Figure 3.10 depicts the simulation results for the acquired relational conditioning experiment of Honey and Watt (1999). The simulations demonstrate that the theory is able to capture the general pattern of behavior with B showing itself to be a better predictor of $shock$ than is D. The mechanism of generalization is apparent in Figure 3.10(e). Latent cause x_3 has merged the separate associations between A and Z and between B and Z as well as the revaluation association between A and $shock$ resulting in the spurious correlation between B and $shock$.

3.7.3. Representation and reinforcement

Holland (1998) investigated the effects of variations in the amount of training on the representations of events. In Table 3.3, I present simplified versions of his experiments that ignore certain control conditions. In one experiment, subjects each received 16 pairings of an auditory stimulus A with a flavored sucrose solution $food_1$. The subjects also received either 0, 16, 28 or 40 pairings of auditory stimulus B and a different flavored sucrose solution $food_2$. In a second phase of training, 6 presentations of B preceded injections of the nausea inducing substance LiCl. Subject response was then probed by measuring the rate of consumption of $food_1$ or $food_2$ in the absence of any stimuli. As shown in Figure 3.11(a), subjects with a small amount of B-$food_2$ training significantly reduced their consumption of $food_2$; however consumption of $food_2$ increased, across groups of subjects, with the number of B-$food_2$ trials they experienced, until it matched (or exceeded) consumption of $food_1$.



(a) Experiment: A vs C

(b) Experiment: B vs D

(c) Simulation: A vs C

(d) Simulation: B vs D

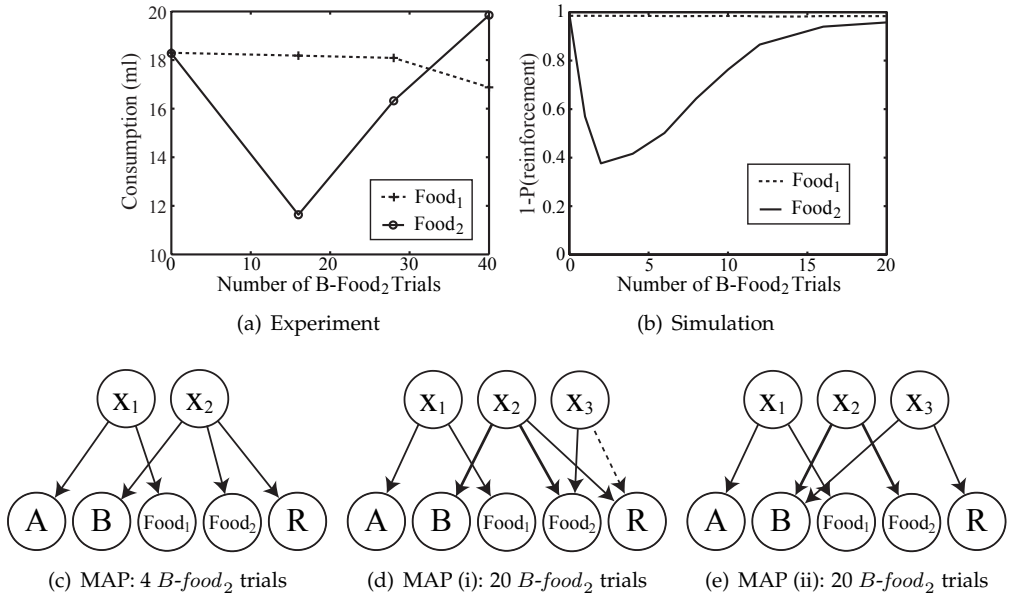
(e) MAP model structure

Figure 3.10. Acquired relational equivalence (ARE) experiment of Honey and Watt (1999). Experimental results comparing responding to (a) A with C and to (b) B with D. Testing revealed no significant difference between Y and Z (not shown). (c)-(d) Simulations demonstrating the acquired relational equivalence experiment with the probability of shock generalizing more readily from A to B than to D. Simulation results are plotted as $1 - P(shock)$ to match the empirical suppression ratio measure. A typical model structure resulting from the simulations (dashed lines represent links with negative weights). Latent cause x_3 encodes the association between A, B and $shock$. The simulations used 10 of each biconditional trial type and 8 of each revaluation trial type.

Associative transfer between stimulus and reinforcement presents a challenge for associative learning theories; however, since latent cause theory makes no distinction between stimuli and reinforcers it naturally accounts for these results as another example of the effect of forming spurious correlations through inference over the latent causes. In addition, the theory correctly predicts that such an effect should be alleviated by increasing

Phase 1		Phase 2		
$A - food_1$	$B - food_2$	$B - LiCl$	Test $\rightsquigarrow$ Result	Test $\rightsquigarrow$ Result
16	0	6	$food_1 \rightsquigarrow -$	$food_2 \rightsquigarrow -$
16	16	6	$food_1 \rightsquigarrow -$	$food_2 \rightsquigarrow Resp.$
16	28	6	$food_1 \rightsquigarrow -$	$food_2 \rightsquigarrow -$
16	40	6	$food_1 \rightsquigarrow -$	$food_2 \rightsquigarrow -$

Table 3.3. Summary of Experiment 2 from Holland (1998) with $Resp.$ indicating a conditioned response to the stimulus.



(a) Experiment (b) Simulation

(c) MAP: 4 B-$food_2$ trials (d) MAP (i): 20 B-$food_2$ trials (e) MAP (ii): 20 B-$food_2$ trials

Figure 3.11. The empirical results and simulations of the experiment of Holland (1998). (a) The experimental finding (adapted from Holland (1998) Experiment 2.). (b) Simulations from the latent cause model. (c)–(e) MAP model structure with 4, 20 and 20 B-$food_2$ trials respectively. As is the case in Figure 3.8, Figures (d) and (e) represent two roughly equal modes of the simulated posterior distribution.

the number of training trials. Figure 3.11 presents the results of the model simulations on a training regimen similar to that used in Holland (1998).

3.8. Discussion

In this chapter I advance the idea that issues of generalization and discrimination in animal learning may be understood as an instance of Bayesian inference over latent cause model structures. In support of this notion I demonstrate how such a theory accounts for a range of conditioning phenomena, some of which have not previously been considered as originating from the balance between the competing requirements of generalization and discrimination.

It is noteworthy that the configural model of Pearce also captures the relative order of discrimination in both the asymmetric and irrelevant stimulus variations of negative patterning experiments presented earlier. It may be possible to interpret the Pearce model as approximating the Occam's razor effect of Bayesian inference. Configural units with a higher degree of stimulus overlap are similarly activated. As a result discrimination between them is slow to evolve relative to configural units with little stimulus overlap. However the correspondence is not complete.

There are some important differences between the role the latent causes in our theory and the role of configural units in the models of Pearce and Rescorla-Wagner. Most notably, the latent causes are not generated according to heuristic rules based on single observations of novel stimulus configurations. Instead they are inferred, following established inference procedures, with more probability mass going to better choices of model structures. The chief consequence of this distinction is that the latent cause "configural units" are capable of grouping stimuli (sharing common latent causes) that have never been observed together as was seen in the latent cause account of acquired equivalence. This leads to the prediction that competition effects typically arising from compound conditioning procedures (such as blocking and overshadowing) should be observable in settings where the stimuli are trained apart. Empirical evidence seems to support such a claim (Escobar, Matute, & Miller, 2001; Matute & Pineño, 1998; Underwood, 1966).

3.8.1. A connectionist model of configural conditioning

The latent cause theory of discrimination and generalization advanced in this chapter bares strong similarities to the model of (Gluck & Myers, 1993) introduced in Chapter 2. Their connectionist model considered aspects of discrimination and generalization between stimulus patterns and hypothesized that a key site of stimulus representation is the hippocampus. The model demonstrates a similar kind of tradeoff between generalization and discrimination as demonstrated by the latent cause model, with discrimination gaining influence with further training. While the latent cause model uses the Bayesian (or automatic) Occam's razor effect to achieve this tradeoff, the Gluck-Myers model exploits the well known neural network phenomenon of growth of model complexity with increasing weight magnitudes (Bishop, 1995). Since these weight magnitudes tend to increase with training, the network is able to emulate the Bayesian model averaging effect of model complexity growth with accumulated training. To the extent that the Gluck-Myers model

Phase 1		Phase 2	
$A - food_1$	$B - food_2$	$food_i - LiCl$	Test $\rightsquigarrow$ Result
16	16	6	$B \rightsquigarrow Resp.$
16	40	6	$B \rightsquigarrow Resp.$
16	160	6	$B \rightsquigarrow Resp.$

Table 3.4. Summary of Experiment 3 from Holland (1998). This experiment is similar to the previous experiment (shown in Table 3.3) with the crucial exception that the roles of the stimulus B and $food_i$ are reversed. In phase 2 training, half of the subjects receive $food_1 - LiCl$ pairings while the other half receive $food_2 - LiCl$ pairings. The test results are reported for those subjects that received $food_2 - LiCl$ pairings in phase 2.

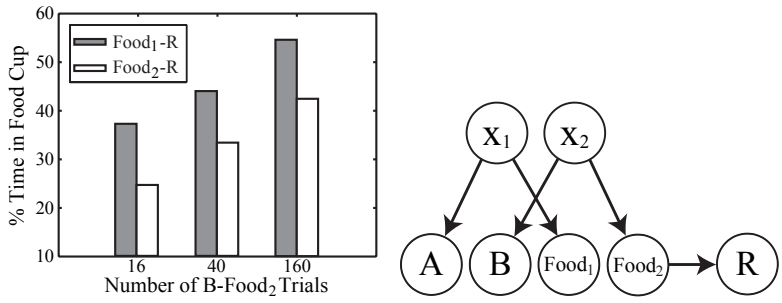
approximates Bayesian model inference, it can be interpreted as an algorithmic implementation of the computational level theory offered by the latent cause model. None of the experiments considered in this chapter were simulated with the Gluck-Myers model, but for the most part it seems that the model would offer a reasonable account of discrimination phenomena, such as the various negative patterning phenomena and even the acquired equivalence effect.

3.8.2. Representation and causality

As discussed at the end of Chapter 2, the latent cause framework is motivated by considerations of realistic causal relationships that arise within classical conditioning paradigms. There is an interesting result of Holland (1998), closely related to the one considered in section 3.7.3 and beyond the scope of the latent cause model, that offers additional insight into the causal interpretation of the model as described in Chapter 2.

The Holland experiment, summarized in Table 3.4, involves a similar initial training regimen to that previously described in section 3.7.3: subjects received 16 pairings of A with a flavored sucrose $food_1$ and either 16, 40 or 160 pairings of B with flavored sucrose $food_2$. However this time, the second phase of training consisted of pairings of either $food_1$ or $food_2$ with $LiCl$ reinforcement instead of A or B. As shown in Figure 3.12(a), responding to B was significantly lower in those subjects who experienced $food_2$ paired with $LiCl$. Interestingly, this effect did not disappear with further training as it did in the earlier experiment. Comparing the two paradigms, it seems the only appreciable difference is that the stimuli B and $food_2$ change roles. Since latent cause theory makes no distinction between stimuli and reinforcers, it incorrectly predicts identical results for the two experiments.

Reconsidering the latent cause assumption in light of the current experiment, an interesting picture emerges. As stated in Chapter 2, the theory is based on an assumption that in the context of classical conditioning experiments, animals infer a generative model



(a) Experiment, testing B (b) Possible model structure

Figure 3.12. Experiment 3 of Holland (1998). (a) The experimental results show that unlike the previous experiment of section 3.7.3, the second-order association between B and the reinforcer R does not disappear with further training, (b) a proposed model structure that includes a direct link between food and nausea (R). Such a model would account for the lasting association between stimulus B and the nausea inducing reinforcer R.

of the world where observables such as stimuli and reinforcers are generated from underlying and unobserved causes. This interpretation was motivated with the example of an experimental subject experiencing a tone followed immediately by the stinging feeling of a footshock. I suggested that the most reasonable causal inference was that *something* caused both the tone and the footshock, not that the tone caused the footshock. On this understanding I proceeded to define a computational model that reflected this assumption.

In the present experiment, however, an auditory stimulus B is paired with a flavored sucrose solution $food_2$ and then that sucrose solution is paired with $LiCl$, a nausea inducing agent. While it may be the most plausible causal inference to suppose that a common cause gave rise to B and $food_2$, it certainly seems that the most parsimonious causal interpretation of the co-occurrence of consuming a sucrose solution and the experience of nausea is that ingestion of the solution itself *caused* the nausea. This is a situation where the standard Rescorla-Wagner causal interpretation of the relationship between stimulus and reinforcer is the most plausible and the assumption of a latent cause is distinctly less plausible. Thus this experiment is in some sense beyond the scope of the latent cause theory of conditioning.

Ironically, the model's failure to account for this experimental result may be interpreted as support for the basic assumption of causal inference underlying latent cause theory. To fully account for both this experiment and the experiment found in section 3.7.3 would require a kind of theory-based inference scheme (Tenenbaum & Griffiths, 2001b) as

discussed in Chapter 2. A possible model structure that might be inferred from a more general cause model inference theory is given in Figure 3.12(b). Such a model would capture the lasting association between B and nausea.

CHAPTER 4

Modeling Change

4.1. Introduction

Animals constantly face the challenge of negotiating a dynamic and complex world where the ability to detect and adapt to changing circumstances is paramount to survival. Yet despite the importance of this quality, many rational theories of learning, including the one advanced in the previous chapter, explicitly assume that subjects treat their environment as if it did not change—as if the order in which events are experienced is irrelevant to predictions of future events.

In the previous chapter I developed a normative account of configural conditioning phenomena. Learning was framed as a process of balancing the ability to account for the observed data with the complexity of the model used to describe the data. I claimed that animals approximate rationality, in the sense that their reasoning about uncertainty in their environment roughly follows the calculus of probability theory. I assumed that the trials were sampled independently from some uncertain but stationary process. The consequence of this assumption was that the order of the trials was irrelevant to the inferential process under study. This assumption was largely valid as the majority of the experiments involved interleaved trial types, creating a more or less homogeneous pattern of events over time. However, there were other experiments—including the acquired equivalence experiments—where there were distinct phases of trial presentations. In these cases, the assumption of stationarity appears to be a particularly poor assumption. There are still other conditioning experiments, such as simple extinction, for which the independence assumption renders the model useless in making predictions. Correcting this failure is the subject of the current chapter.

It is hardly controversial to suggest that animals are able to adapt to changing contingencies in the environment. Ever since the early explorations of Pavlov, the adaptive character of behavior has been a central focus of study with experimental manipulations

explicitly changing established contingencies. It is not surprising, therefore, to find that many computational models of conditioning possess mechanisms of adaptation to changing contingencies. Even the Rescorla-Wagner model is capable of accounting for some simple forms of adaptation apparent in conditioning experiments. However, of the established associative learning theories, it is the model of Pearce and Hall (1980) that has most successfully accounted for how animals respond to change. The theory of Pearce-Hall states that surprising changes in reinforcement contingencies increase the learning rate. Over the past few decades, there have been a significant number of experiments, some of which are reviewed below, investigating the sometimes surprising predictions of the Pearce-Hall model. By and large, the empirical data seem to support the claim that unexpected reinforcement drives faster learning.

In contrast to most rational theories of cognition, the conditioning models of Dayan and Kakade (2001) are able to account for some aspects of the effect of a non-stationary environment. They explicitly define a generative *model of change*—a formal device meant to reflect the subject's belief regarding changing contingencies within their environment. With the model of change, Dayan-Kakade construct a probabilistic version of Rescorla-Wagner that offers an improved account of animal behavior. Dayan and Yu (2003) furthered the generative model account to include two important ideas: first, that the environment can change by a variable amount from trial to trial; and second, that the amount of change experienced by the environment is not known to the animal and must be inferred from data. These insights allow the Dayan-Yu model to provide a normative account of many of the empirical phenomena described by Pearce-Hall.

In this chapter I explore the consequence of merging latent variable theory with the notion of a generative model of change. Incorporating the insights of Dayan and Yu, the resulting non-stationary latent cause model negotiates model change with the same Bayesian inference mechanism that gave rise to the model complexity / data fidelity tradeoff in the previous chapter.

The unique generative structure of latent cause theory, when combined with a generative model of changing contingencies, generates a surprising prediction. Unlike the models of Pearce-Hall and Dayan-Yu that predict a learning rate increase with surprising reinforcement events, the non-stationary latent cause model predicts that *any* surprising stimulus event should precipitate an increase in the learning rate. Interestingly, the experimental data seem to support such a prediction.

4.2. Previous work

In this section I review the dominant theoretical perspectives on how animals perceive, detect and track change and discuss some of the experimental evidence supporting the various theories.

4.2.1. The Rescorla-Wagner model

While the model of Rescorla and Wagner (1972) possesses an almost trivial mechanism to track change, nonetheless the model is able to capture fundamental conditioning phenomena that the Bayesian model of the previous chapter is woefully ill-equipped to explain.

The original Rescorla-Wagner model possesses constant learning rates referred to as *associabilities*. Recall, from Chapter 2, the equation for the weight update:

$$\Delta V_{i,t} = \alpha_i \beta (R_t - V_{\Sigma,t}) X_{i,t}, \tag{4.1}$$

where I have explicitly represented the dependency on time. There are two associabilities: α_i and β, specific to the ith stimulus and the reinforcer respectively. The product $\alpha_i \beta$ amounts to a constant learning rate per stimulus. A constant learning rate has the effect of weighting more recent trials more heavily than previous trials. A given reinforcement error $(R_t - V_{\Sigma,t})$ would have the same impact on the associative strength update if it were the tenth trial or the ten-thousandth trial, meaning that the last 9 trials or the last 9,999 trials would be weighted equivalently. As a result, the Rescorla-Wagner model is able to track slowly changing associations (or contingencies) between observed stimuli and reinforcement. It is also able to follow, over a number of trials, more dramatic changes in contingencies such as the transition from reinforcement to non-reinforcement found in extinction paradigms.

Extinction is perhaps the simplest experimental paradigm where the issue of the non-stationarity of the environment is clearly apparent. Empirically, extinction is the loss of conditioned responding to an established conditioned excitor (a stimulus that has previously been paired with reinforcement and is able to elicit a conditioned response). Extinction is typically accomplished through repeated presentations of the stimulus in isolation (Pavlov, 1927). Experimentation has revealed that the number of training trials required for extinction is not correlated with the number of acquisition trials (assuming this is enough to provoke a strong conditioned response). At the very least, this suggests that either implicitly or explicitly animals are not using the full quantity of evidence in favor of a positive

association between stimulus and reinforcement. As a consequence, animals extinguish faster than would be expected if observations were emitted from a stationary process.

The Rescorla-Wagner model explains extinction as a loss of associative strength. During acquisition with $A+$ the associative strength V_A is increased, eventually reaching an asymptotic level where the predicted reinforcement intensity matches the observed reinforcement intensity. With no prediction error the associative strength remains constant until there is a change in reinforcement contingency. Once the extinction phase begins, the stimulus presentations are no longer reinforced, causing a negative prediction error and a decrease in the associative strength, in accordance with Equation 4.1. Because nothing is learned once the asymptotic value of associative strength is achieved, the number of extinction trials required to reach some extinction criterion is independent of the number of reinforced trials, assuming the asymptote had been reached.

While the model of Rescorla-Wagner is able to capture some basic experiments that require that changing contingencies be tracked, it falls short of explaining many phenomena designed to study how animals judge changes in reinforcement contingencies. The basic issue is that animals do not seem to be assuming a constant learning rate. Instead, they seem capable of adapting the learning rate in certain circumstances. A number of models have stepped forward with proposals regarding the mechanism controlling the learning rate. None of these have met with more success than the model of Pearce and Hall (1980).

4.2.2. The Pearce-Hall model

The Pearce-Hall model (Pearce & Hall, 1980) was originally developed as an adaptation of an earlier model by Mackintosh (1975) and offers an alternative account of conditioning phenomena to that of Rescorla-Wagner. Like its predecessor, the Pearce-Hall model focuses on the adaptation of stimulus associability; however in direct contrast to the model of Mackintosh, the Pearce-Hall theory posits that the associability of a stimulus is proportional to the extent that it is *not* an accurate predictor of reinforcement on the last trial:

$$\alpha_{i,t} = |R_{t-1} - V_{\Sigma,t-1}|, \tag{4.2}$$

where $\alpha_{i,t}$ is the associability of the ith stimulus on trial t, R_{t-1} is the value of reinforcer on the last trial ($t-1$), and $V_{\Sigma,t-1}$ is the total predicted reinforcement from all stimuli at trial $t-1$. $V_{\Sigma,t}$ is given by the same equation as in the RW model:

$$V_{\Sigma,t} = \sum_i V_{i,t} X_{i,t}. \tag{4.3}$$

where $X_{i,t}$ represents the presence ($X_{i,t} = 1$) or absence ($X_{i,t} = 0$) of the ith stimulus at time t. The associative strength update rule is proportional to the current value of the

associability, weighted by the values of the stimulus and reinforcer intensity (S_i and R_t respectively),

$$\Delta V_{i,t} = S_i \alpha_{i,t} R_t. \tag{4.4}$$

The Pearce-Hall model is able to account for a number of the standard conditioning phenomena including blocking and overshadowing. A more elaborate version of the model (Pearce & Hall, 1980) is needed to account for "inhibition effects" such as extinction and conditioned inhibition. Where the Pearce-Hall model distinguishes itself from the Rescorla-Wagner model is its prediction of attentional effects such as latent inhibition.

Latent inhibition (Lubow & Moore, 1959) (also known as the stimulus pre-exposure effect) is typically a two phase training procedure in which a subject experiences an extended phase of repeated exposure to an unpaired stimulus $A-$. Then, in a second phase, the same stimulus is pair with reinforcement, $A+$. Subjects who experience both phases of training are slower to acquire an association between the stimulus and reinforcement—requiring more trials to achieve an acquisition criterion—than subjects who did not experience unpaired presentations of A.

According to the Rescorla-Wagner model, nothing is learned without reinforcement prediction error. Hence, the large number of presentations of $A-$ that form the latent inhibition pre-training phase have no effect on the model. As a result, subsequent acquisition proceeds as it would if the unpaired stimulus presentations never occurred and the rate of acquisition is entirely unaffected.

Unlike Rescorla-Wagner theory, the Pearce-Hall model is able to offer a satisfactory explanation of latent inhibition. According to this perspective, initial presentations of $A-$ establish a low associability for A. With the associative strength $V_{A,t}$ initially low, unpaired presentations of A result in an associability near zero, $\alpha_{A,t} \approx 0$ (from Equation 4.2). The reduced associability resulting from $A-$ pre-training leads to a reduction in the associative strength learning rate (through Equation 4.4) which in turn results in the empirically observed retarded rate of acquisition characteristic of latent inhibition.

While the Pearce-Hall explanation of latent inhibition is superior to that of Rescorla-Wagner, it does not withstand a great deal of quantitative scrutiny. In stark contrast to the empirical observed phenomenon of latent inhibition, the theory of Pearce-Hall, as presented above, predicts that only a single trial of $A-$ pre-training is required to establish maximal latent inhibition. In their original exposition of the model, Pearce and Hall (1980) address this issue directly and suggest using an alternate form of Equation 4.2:

$$\alpha_{i,t} = \frac{1}{c} \sum_{\tau=t-c}^{t-1} |R_\tau - V_{\Sigma,\tau}|, \tag{4.5}$$

Phase 1	Intermediary	Phase 2	Rate of Acquisition (Phase 2)
–	–	$A+$	fast
$A+$	–	$A+$	slow
$A+$	$A-$	$A+$	fast

Table 4.1. Summary of two experiments by Hall and Pearce (1979, 1982), demonstrating a stimulus pre-exposure effect with pre-training pairing with a weak shock, $A+$. $A+$ represents stimulus A being paired with a strong shock. Surprisingly, the presentation of a few $A-$ (unpaired) trials between the pre-training and acquisition phases eliminates the pre-exposure effect. The first row is the control group.

for some constant c. Such a modification to the model would slow the onset of maximal latent inhibition as a function of the number of $A-$ pre-training trials, putting the model more in line with the empirical effect. Regardless of these details, the theory of Pearce-Hall offers an account of conditioning phenomena that is not easily dismissed. In fact there are a significant number of conditioning phenomena for which the Pearce-Hall explanation is the most compelling. Among the most theoretically interesting of these phenomena are the establishment of latent inhibition using reinforced pre-training trials and the downwards unblocking phenomenon.

In a series of experiments, Hall and Pearce (1979) elaborated on the basic empirical phenomenon of latent inhibition. They demonstrated that repeated presentations of a stimulus paired with a weak shock $A+$ reduced the speed of acquisition during subsequent pairings of the stimulus with a stronger shock $A+$ relative to control subjects that did not receive preliminary training with weak shock. Hall and Pearce (1982) went on to show that by providing a few presentations of the stimulus in isolation ($A-$) between $A+$ and $A+$ training restored the acquisition speed to that of the control group that never received $A+$ training. These results are summarized in Table 4.1.

The Pearce-Hall model offers a rather elegant explanation of this experiment. During Phase 1 training with $A+$, the associative strength $V_{A,t}$ is increased to match the reinforcement intensity of the weak shock and the associability is decreased in proportion to the prediction error. When Phase 2 training begins the associability is small and learning is slow when compared to the control condition without Phase 1 training. According to the Pearce-Hall account, adding a few trials of $A-$ between the $A+$ and $A+$ trials results in a nonzero prediction error and that translates (via Equation 4.2) into a recovery of significant associability.

Further evidence in favor of Pearce-Hall theory is provided by Dickinson et al. (1976) with their famous *downwards unblocking* experiment. They demonstrate that if rats are pre-trained with a stimulus paired with two separate reinforcements (electric shocks separated

by a time interval of 8 seconds), $A+ \rightarrow +$, and then presented with a compound of two stimuli similarly reinforced, $AB+ \rightarrow +$, they exhibit blocking to B (i.e. they respond significantly less to B than did control subjects who did not receive pre-training with $A+ \rightarrow +$). However when the second reinforcement is omitted during compound stimulus training, $AB+$, the rats respond equally to B whether or not they received pre-training with $A+ \rightarrow +$. From this Dickinson et al. conclude that the surprising omission of the second reinforcement causes "unblocking".

In the case of downwards unblocking, the Rescorla-Wagner model actually predicts that the opposite outcome should occur: the absence of the second reinforcement should cause B to become a conditioned inhibitor rather than the empirically observed conditioned excitor. The Pearce-Hall account of this phenomenon is simple and persuasive. The absences of the second reinforcement generates a negative prediction error that translates to significant associability for both A and B. Subsequent training with $AB+$ allows B to accrue significant associative strength.[1]

The previous experiments illustrate the considerable predictive power of the Pearce-Hall theory. The experiments highlight a feature of the model that is a large part of its predictive success: reinforcement prediction error drives the speed of learning. That is, the learning is controlled by the extent to which reinforcement events are surprising. There are numerous examples of variations and elaborations of these basic phenomena (for a review see Holland (1997)). All of these share the same basic feature: a surprising reinforcement event with a corresponding increase in the rate of change of responding.

It is worthwhile to take a moment and reflect on what these models and experiments are telling us about what animals are learning with regard to change in the environment and how it affects behavior. The Rescorla-Wagner model possesses the simplest of mechanisms to track changing contingencies: a constant learning rate. This turns out to be sufficient to explain (at least qualitatively) the simplest of contingency changing paradigms such as extinction. Yet it seems clear from the experimental evidence presented thus far in this chapter that the learning rates of organisms are not constant: they decrease with extended periods of predictable outcomes and they increase when outcomes change suddenly. The Pearce-Hall model formalizes this notion and has been remarkably successful in accounting for the behavior of animals. Latent inhibition, the reinforced pre-exposure

[1] Interestingly, if the two reinforcements are delivered in quick succession (with little or no delay between them) then the Rescorla-Wagner prediction turns out to be correct and the Pearce-Hall prediction is wrong. It is not clear how the extended Pearce-Hall model that can account for conditioned inhibition effects would negotiate this situation. For a nice account of this phenomenon I refer the interested reader to the work of Dayan and Long (1998).

effect and downwards unblocking are all parsimoniously explained with the Pearce-Hall mechanism for associability change. The mechanism specifies a detailed theory of *how* surprising reinforcements give rise to faster learning. Regarding the question of *why* unexpected events cause an increase in the learning rate, Pearce-Hall offers little insight. However, the statistical account of Dayan and Kakade and in particular the account of Dayan and Yu possess an illuminating perspective on the question of why.

4.2.3. A statistical account of change

Dayan and Kakade (2001) offer one of the very few statistically normative accounts of learning in an environment subject to change. Building on the inference model suggested by Sutton and colleagues (Sutton, 1992; Gluck, Glauthier, & Sutton, 1992), Dayan and Kakade introduce an alternative formulation of the Rescorla-Wagner model, relating reinforcement intensity R_t directly to the stimuli $X_{i,t}$ and their associative strengths or weights $W_{i,t}$:

$$R_t = \sum_i W_{i,t} X_{i,t} + \epsilon_{i,t} \tag{4.6}$$

where $\epsilon_{i,t}$ is an additive perturbation term, drawn independently from a zero mean Gaussian: $\epsilon_{i,t} \sim \mathcal{N}(0, \tau)$. Unlike the original Rescorla-Wagner formulation that specifies a point prediction $V_{\Sigma,t}$ for the reinforcement intensity; Equation 4.6 specifies a full intensity distribution, $R_t \sim \mathcal{N}(W_t \dot{X}_t, \tau)$, the vectors W_t and X_t have elements $W_{i,t}$ and $X_{i,t}$ respectively.

The weights $W_{i,t}$ are equivalent to the associative strengths and represent the contingencies present in the environment. In non-stationary environments, these contingencies are subject to change. To capture these changing contingencies, Dayan and Kakade propose a generative *model of change* over the weights:

$$W_{i,t+1} = W_{i,t} + \eta_{i,t}, \tag{4.7}$$

where $\eta_{i,t}$ is another additive perturbation term, drawn independently from a zero mean Gaussian for each stimulus i and trial t: $\eta_{i,t} \sim \mathcal{N}(0, \sigma)$. Equation 4.7 describes how the weights are assumed to change: by small increments of $\eta_{i,t}$ on every time-step t.

According to Dayan and Kakade, the task of the animal is to infer the weights $W_{i,t}$ given the observations of the stimuli $X_{i,t}$ and reinforcements R_t. With the prior uncertainty in $W_{i,t}$ represented as a Gaussian distribution, inference over $W_{i,t}$ reduces to the well known Kalman filter algorithm. Conditional on the observations, the vector of stimulus weights W_t is described by a Gaussian distribution with mean $\hat{W}_t$ and covariance Σ_t: $W_t \sim \mathcal{N}(\hat{W}_t, \Sigma_t)$. In accordance with the Kalman filter, the temporal updates of these

quantities are given by:

$$\hat{W}_{t+1} = \hat{W}_t + \frac{\Sigma_t X_t}{X_t \Sigma_t X_t + \tau^2}(R_t - \hat{W}_t X_t) \quad (4.8)$$

$$\Sigma_{t+1} = \Sigma_t + \sigma^2 \mathbb{I} - \frac{\Sigma_t X_t X_t \Sigma_t}{X_t \Sigma_t X_t + \tau^2} \quad (4.9)$$

where $\mathbb{I}$ is the identity matrix. If Σ_t is diagonal then Equation 4.8 reduces to the standard Rescorla-Wagner associative strength update rule.

While the model proposed by Dayan and Kakade is quite similar to the original Rescorla-Wagner model, the generative model of changing weights represents a significant departure. For instance, by exploiting the difference between the relatively large prior uncertainty in the value of the weights and the relatively small asymptotic uncertainty due to the diffusion term η_t, the Dayan-Kakade model is capable of demonstrating latent inhibition. A large initial uncertainty (or variance) results in fast learning on early presentation of a stimulus; however during pre-training, the unreinforced presentations of the stimulus cause the uncertainty of the weight to shrink to its asymptotic value which in turn decreases the Kalman gain, $(\Sigma_t X_t)/(X_t \Sigma_t X_t + \tau^2)$, slowing adaptation to future changes in reinforcement contingencies.

Comparing the Dayan-Kakade model to that of Pearce-Hall reveals a significant difference in how the effective learning rate (Dayan-Kakade: Kalman gain; Pearce-Hall: associability) is determined. For a particular stimulus i, the Kalman gain of the Dayan-Kakade model is determined entirely by the two perturbation parameters σ_i and τ and by the history of stimulus i presentations through the effect of $X_{i,t}$ in Equations 4.8 and 4.9. The presence or absence of reinforcement, surprising or not, has the same impact on the rate of weight adaptation as in the Rescorla-Wagner model. In contrast, the associabilities of the Pearce-Hall model are determined solely by the degree to which the reinforcement intensity is surprising.[2]

Naturally, this distinction leads to a divergence of predictions across a number of experiments, including the well known downwards unblocking effect. As discussed above, under the right conditions, the omission of an expected second stimulus in an otherwise standard blocking paradigm results in the *unblocking* of the normally blocked stimulus B. With the associability being tied to surprising reinforcement events, including omission, the Pearce-Hall model is able to explain relatively normal acquisition to B as a direct result of the prediction error resulting from the omission of the second reinforcement. The Dayan-Kakade model is unable to account for this type of unblocking both because the

[2]This distinction was not lost on Dayan and Kakade (Dayan et al., 2000; Dayan & Kakade, 2001) as they suggested a modification to the Kalman update equations originally introduced by Sutton (Sutton, 1992) incorporating prediction error directly into the determination of the gain.

Kalman filter offers no mechanism to increase learning rates as a result of reinforcement prediction error, and because the omission of an expected reinforcement could only result in a negative prediction error rendering B a conditioned inhibitor rather than an unblocked excitor.

In work exploring the role of the neuromodulators acetylcholine and norepinephrine on learning and inference, Dayan and Yu (2003) (also see Yu and Dayan (2003)) elaborated on the model of weight change introduced by Dayan and Kakade. In addition to the slow drifting caused by the η_t perturbations, Dayan and Yu considered larger occasional perturbations representing moments of dramatic change in the environment. Such dramatic changes in contingencies are common in conditioning paradigms, an example being the shift from acquisition trials to extinction trials. These infrequently occurring dramatic perturbations were incorporated into the model of weight change through an additive term:

$$W_{t+1} = W_t + \eta_t + c_t\phi_t \tag{4.10}$$

where η_t is as before and $\phi_t \sim \mathcal{N}(0, \rho)$, an additional additive independent Gaussian perturbation term, with variance $\rho^2 >> \sigma^2$. The variable $c_t \in \{0, 1\}$ signals the occurrence of dramatic change. With $c_t = 1$ a large change in the weight value is possible over a single time-step.

Crucially, Dayan and Yu take c_t to be an unknown latent variable and, like W_t, subject to inference from the observed stimuli and reinforcement deliveries.[3] In contrast to the straightforward Kalman filter of Dayan and Kakade, the model of change of Dayan and Yu induces uncertainty regarding the diffusion process on each time-step. Depending on the value of the unobserved variable c_t, the drift is either likely to be small or large. As a consequence of having to infer the value of c_t, the inference procedure comes to incorporate prediction error. Large reinforcement prediction error is evidence that the weights have changed dramatically and hence $P(c_t = 1) \approx 1$.

According to the generative model of Dayan and Yu (2003), reinforcement contingencies are subject to occasional large deviations. These spontaneous deviations are unobserved and their occurrence must be inferred from observations. Consequently, an unexpected reinforcement or the omission of an expected reinforcements constitutes evidence in favor of the occurrence of a large deviation. In the event of a potentially large perturbation in the weight values, uncertainty in these quantities increases, which has the effect—according to the rules of rational inference or simply Bayes' rule—of giving more credence

[3] As Dayan and Yu (2003) point out, the addition of the second occasional perturbation term renders the inference intractable as the number of modes of the mixture distribution grows exponentially in time. Dayan and Yu resort to a heuristic approximation of the inference.

to new observations. If these new observations provide evidence that the weights have changed, then this information will manifest itself in a more rapid change in the weight estimates than it would if the large deviation event did not occur. Thus, from the perspective of the Dayan-Yu model, unexpected reinforcement events gives rise to uncertainty in the weights and this, in turn, precipitates an increase in the learning rate.

Though the model of Dayan and Yu (2003) has not yet been applied to the battery of Pearce-Hall style experiments, where surprising reinforcement events signal fast learning, it should in principle be capable of capturing many of the same sorts of phenomena, with at least one notable exception: the downwards unblocking experiment of Dickinson et al. (1976). The mechanism for this prediction failure is the same as for its predecessors, the Dayan-Kakade model and the Rescorla-Wagner model. While the surprising omission of the second reinforcement causes the learning rate to increase, it also causes an over-expectation of reinforcement which in turn leads to negative incremental weight updates. Thus a would-be unblocked excitor becomes a conditioned inhibitor.

A further statistical perspective on change is offered by models of change detection (Gallistel, 1990; Gallistel, Mark, King, & Latham, 2001). Both of these models detect when the rate of reinforcement in the presence of a given stimulus has changed, though the model of Gallistel et al. (2001) has the advantage of being more formally grounded in a Bayesian statistical perspective, as opposed to the frequentist setting of the model of Gallistel (1990). Gallistel et al.'s (2001) model is very similar in spirit to the Dayan-Yu model in that change is conceptualized as a binary decision: change or no change, with the partial goal of the inference procedure to resolve when the discrete change events occur. It is noteworthy that, unlike the Dayan-Yu model, the model of Gallistel et al. (2001) is, in principle, capable of phenomena such as downwards unblocking. Its successful account centers on the rates of reinforcement, with the counts being restarted with the determination of change.

4.3. Change as a stochastic process

In contemplating Bayesian accounts of animal learning in a non-stationary environment, the interpretation of the prior must be handled with some care. Typically, in stationary applications of Bayesian inference, the prior distribution over the system parameters represents the uncertainty regarding the parameter values before the data are observed. This prior is then combined with the information from the data to recover the posterior distribution. In the stationary case, the order in which the data are processed has no impact on the final result. In the non-stationary case, the data is sampled from a system undergoing change. The observations at time t are a reflection of the system at that time; at time

$t+1$ the system has changed and the observations at $t+1$ are not exchangeable with the observations at time t. This has long been recognized and does not represent a significant conceptual problem. The Kalman filter naturally incorporates these considerations into its inference algorithm. The idea is simple: the posterior over the parameters at time-step t combines with the model of parameter change (Equation 4.7 in the Dayan-Kakade model) to determine the prior at time $t+1$. With recursive application of Bayesian inference, one recovers a distribution over the full trajectory of the parameters as they evolve.

A generative model of change describes how the parameters of the model evolve from time t to time $t+1$. Since the magnitude and direction of the parameter perturbations are assumed to be unobserved, the model of change defines a stochastic Markov process (or chain): a collection of random variables $\{\theta_t : t \in T\}$ where $T = \{0, 1, 2, \dots\}$ is the index set and

$$p(\theta_t \mid \theta_0, \dots, \theta_{t-1}) = p(\theta_t \mid \theta_{t-1}). \tag{4.11}$$

The transition function $p(\theta_t \mid \theta_{t-1})$ defines both the short term evolution as well as the long term behavior of the process.

The model of change of Dayan and Kakade and that of Dayan and Yu both define Markov chains over the space of weights with transition functions defined by Equations 4.7 and 4.10 respectively. Bringing the theory of stochastic processes to bear on the analysis of these two models reveals a number of important properties. They both define Markov chains that are irreducible, recurrent and aperiodic. Informally, this means that the weights may drift from any value, to any value and may return to any value at any time. These properties ensure that there is no pathological restriction on the paths that the weights may take as they evolve over time—an important property for a model of change.

With regard to their long term behavior, the model of change of Dayan and Kakade and that of Dayan and Yu both possess the Lebesgue (uniform) measure as their invariant measure and can be shown to be null recurrent (Robert & Casella, 1999). Informally this means that, as one might guess, these models of change cause probability mass to continue to spread out forever. To see how this property impacts their performance as models of conditioning, consider the following example (I will only discuss the Dayan-Kakade model, but the Dayan-Yu model behaves analogously). Imagine that, after a typical acquisition training regimen with a stimulus, $A+$, the subject is not presented with A for a very long time (measured with respect to the diffusion process). According to the Dayan-Kakade model of change, the variance of the corresponding weight grows and because A is absent, the Kalman filter equations have no way of stemming the growth. It will continue

to grow until A reappears. Then, with the variance on $W_{A,t}$ arbitrarily large, the model predicts that subsequent training should be nearly instantaneous, requiring only a single trial to reach arbitrarily close to asymptotic responding.[4]

Thus the models of change proposed by Dayan and Kakade and by Dayan and Yu lead to a situation where, in the absence of data, the uncertainty in the value of the weights can grow beyond any finite initial prior uncertainty (variance) in the weights—before any data is observed. This property is problematic for a model of change. How can one be more uncertain after having observed data a long time ago, than before one observes any at all?

A solution to this problem is found by recognizing that the animal's uncertainty regarding contingencies in the environment after a very long period without observing data should be equivalent to its uncertainty before observing any data. More precisely, after a very long time (measured with respect to the diffusion time constant of the change process), the animal's belief distribution over the generative model of the world should approach its prior distribution. In the language of stochastic processes, this implies that the diffusion process induced by the model of change should have an invariant distribution[5] equal to the prior distribution over parameters.

Specifying the model of change to admit the prior as the invariant distribution implies that the animal's beliefs regarding the environment are not dependent on some temporal origin that might signal the start of the training regimen. It also offers an interesting interpretation of the prior as the invariant distribution of the environment. The model of change I advance in this chapter will possess an invariant distribution equal to the model prior.

4.4. A latent cause model of change

In this section, I advance a version of the latent cause model that incorporates a generative model of change, similar to that of Dayan and Yu, that reflects the subject's beliefs regarding event contingencies. Building on the Bayesian model inference perspective explored in the previous chapter, I consider a latent cause framework where both model parameters and structure are simultaneously unknown and subject to change. Unlike the models previously considered in this chapter, the latent cause model attempts to recover the full generative process giving rise to all stimuli—not just reinforcements. As in the

[4] In fairness, both the Dayan-Kakade model and the Dayan-Yu model were never designed with an eye toward the rather extreme setting I propose in this thought experiment.

[5] The invariant distribution over the generative model is that which does not change through the action of the model of change. That is, if the distribution over the generative model at time $t = 0$ is the invariant distribution, then it remains the invariant distribution for all $t > 0$.

previous chapter, the basic latent cause model structure is specified by a sigmoid belief network.

4.4.1. Sigmoid belief networks

With stimuli represented as the random vector $y_t \in \{0,1\}^J$, $y_{j,t} = 1$ indicates the presence of the jth stimulus ($y_{j,t} = 0$ represents its absence). One element of y_t is distinguished as the reinforcer R. The correlations between stimuli (including the reinforcer) are encoded through common connections to a vector of L latent causes, $x \in \{0,1\}^L$.

$$P(y_{j,t} \mid x_t, w_t, m_t) = \sigma\left(\sum_{i=1}^{i=L} w_{ij,t} x_{i,t} + w_{0j}\right), \tag{4.12}$$

where $\sigma(z) = (1 + \exp(-z))^{-1}$ is the logistic sigmoid. For each model structure m_t, the influence of the latent causes x_t on stimulus $y_{j,t}$ is encoded in the weight vector $w_{j,t}$. The bias weight w_{0j} ensures that spontaneous events are rare. In the simulations reported here, unless otherwise specified, w_{0j} is fixed at -6.

Conditioned on the set of causes the stimuli are all independent, permitting the probability of a stimulus pattern $\mathcal{S}_t$ to be expressed as the product of the probabilities of the individual stimuli,

$$P(\mathcal{S}_t \mid x_t, w_t, m_t) = \prod_{j \in \mathcal{S}_t} P(y_{j,t} \mid x_t, w_t, m_t) \prod_{j \notin \mathcal{S}_t} (1 - P(y_{j,t} \mid x_t, w_t, m_t)), \tag{4.13}$$

where the notation $j \in \mathcal{S}_t$ indicates all stimuli that are present in the stimulus pattern $\mathcal{S}_t$.

In the previous chapter the prior probability of a cause being active was specified by a real valued parameter passed through a logistic sigmoid (to recover a probability), here the priors are specified directly with the probabilities $w_{i0,t}$. Since the causes are *a priori* independent, the probability of the vector x is given by a product of the probabilities of the activations of the individual causes: $P(x_t \mid w_t, m_t) = \prod_i P(x_{i,t} \mid w_t, m_t) = w_{i0,t}$.

4.4.2. A model of structural change

A key feature of the conditioning theory advanced in the previous chapter is the role of structural inference (model learning) in conditioning. The extension of the structural inference framework to the non-stationary case requires a generative model of structural change. Such structural change is modeled as the appearance and disappearance of latent causes. In an effort to simplify inference over the model, I do not explicitly consider the creation and destruction of individual links as I did in the previous chapter. Instead, each latent cause is assumed to possess a weight connecting it to every stimulus.

As discussed above, the model of change should admit an invariant distribution and that distribution should reflect the subject's prior beliefs as they pertain to the model structure and weights. In the previous chapter, I specified a geometric distribution as the prior over the number of latent causes. As an invariant distribution, the geometric distribution corresponds to a process where at any given time, there is a constant probability of the number of latent causes decrementing by one and a smaller constant probability of the number of latent causes incrementing by one. This is sometimes referred to as a *simple birth-death* process (Bhat & Miller, 2002). This process has the property that the probabilities of destruction of individual causes are not independent.[6] At any given time t, the probability of a particular cause being destroyed is p/n where p is the probability of a cause destruction event and n is the number of causes in existence at t.

As a model of conditioning, this property would seem to suggest that the more distinct causes the subject observes, the less likely it would be for any one cause to disappear. Such model behavior is inconsistent with the assumption that latent causes are independent entities. As an alternative, I model each latent cause as simply having a small constant probability ρ of disappearing on each time-step. A cause is created with constant probability ω at each time-step. This model corresponds to a death-immigration model within the population modeling literature (Bhat & Miller, 2002) with invariant distribution given by the Poisson distribution with parameter ω/ρ:

$$P(m = m^* \mid \omega, \rho) = \frac{e^{-\frac{\omega}{\rho}}}{m^*!} \left(\frac{\omega}{\rho}\right)^{m^*} \tag{4.14}$$

The substitution of the Poisson prior in the model of Chapter 3 would not have significantly changed any of the model predictions. The creation and destruction of latent causes is not directly observed and must be inferred from data. In my simulations: $\omega = \rho = 1 \times 10^{-6}$, corresponding to an assumption that the creation and destruction of causal events are rare occurrences.

In addition to random perturbations in the model structure, the weights representing the influence of the latent causes on the stimuli are also subject to drift through a diffusion process as described in the following section.

4.4.3. A model of weight change

In the case of weight diffusion, the choice of a particular transition function from time t to $t+1$ is constrained by the choice of the target invariant distribution which is in turn constrained to equal the prior distribution over weights. In the previous chapter, the weight

[6] Thanks go to Peter Dayan who originally pointed this out to me.

prior was given as a Laplace distribution:

$$p(w_{ij} \mid \alpha) = \frac{1}{2\alpha} e^{-\frac{|w_{ij}|}{\alpha}}, \tag{4.15}$$

where the hyperparameter $\alpha = 3$ for the simulations I present here (unless otherwise stated)[7].

It can be shown (Bibby, Skovgaard, & Sorensen, 2005) that one may define a diffusion process with the Laplace probability density as the invariant distribution. Such a process is described with the stochastic differential equation (SDE):

$$dw_{ij} = -\theta_i w_{ij} dt + \sqrt{2\theta_i \alpha^2 \left(1 + \frac{|w_{ij}|}{\alpha}\right)}\, dB_t, \tag{4.16}$$

where $t \in [0, \infty)$ is continuous, θ_i is a rate of change parameter and B_t is a standard Brownian motion or Wiener process. The diffusion model given in Equation 4.16 induces an autocorrelation function for $w_{ij}(t)$ given by

$$\mathrm{Corr}(w_{ij}(s+t), w_{ij}(s)) = e^{-\theta_i t}. \tag{4.17}$$

The autocorrelation function describes how a diffusing weight is correlated with its past values. With this diffusion model, past values of w_{ij} will have an exponentially decaying influence on the present value weighted by the rate of change parameter θ_i.

The stochastic differential equation provided in Equation 4.16 is a continuous process. Since I am modeling time as trial length intervals, a discrete version of 4.16 is required. By employing an Euler discretization scheme, the SDE in 4.16 may be approximated by the discrete process:

$$w_{ij,t} = (1 - \theta_{i,t}) w_{ij,t-1} + \sqrt{2\theta_{i,t} \alpha^2 \left(1 + \frac{|w_{ij,t-1}|}{\alpha}\right)}\, \epsilon_{ij,t} \tag{4.18}$$

where time is once again discrete: $t \in [0, 1, 2, \ldots]$.[8]. In Equation 4.18 I have also introduced a temporal index on $\theta_{i,t}$ to emphasize that this quantity is also subject to change over time (as discussed below). Equation 4.18 defines the generative model for weight change. Comparing it to the generative model of change of Dayan and Kakade, given in Equation 4.7, we see two distinct differences. First, the process described in 4.18 is mean reverting (with mean zero), that is, in the absence of data the distribution over $w_{ij,t}$ shifts toward zero at a rate proportional to $\theta_{i,t}$. Second, the perturbations $\epsilon_{ij,t} \sim \mathcal{N}(0, 1)$ are amplified as a function of $|w_{ij,t-1}|$, a consequence of the heavier tails of the Laplace distribution (relative to Gaussian).

[7] The value of α was changed from the previous chapter (where $\alpha = 2$) because the model of change acts to limit the influence of the data. Consequently, a weaker prior is appropriate.

[8] I have set the discretization interval parameter that normally appears in the Euler discretization to one, relying on the relatively small size of $\theta_{i,t}$ to ensure adequately small discretization error.

The probability $w_{i0,t}$ is assumed to be unknown and also subject to change. I define a beta distribution over $w_{i0,t}$:

$$p(w_{i0,t} \mid a_{i,t}, b_{i,t}) = \frac{\Gamma(a_{i,t} + b_{i,t})}{\Gamma(a_{i,t})\Gamma(b_{i,t})} w_{i0,t}^{a_{i,t}-1}(1 - w_{i0,t})^{b_{i,t}-1}, \tag{4.19}$$

with hyperparameters $a_{i,t}$ and $b_{i,t}$ that, in the absence of data, decay to their prior values:

$$a_{i,t} = (1 - \psi)a_{i,t-1} + \psi a_0 + x_{i,t} \tag{4.20}$$

$$b_{i,t} = (1 - \psi)b_{i,t-1} + \psi b_0 + (1 - x_{i,t}) \tag{4.21}$$

The (hyper)hyperparameters $a_0 = 1$ and $b_0 = 10$ define the prior values of $a_{i,t}$ and $b_{i,t}$ respectively, and $\psi = 0.01$.

4.4.4. Determination of $\theta_{i,t}$

The parameter $\theta_{i,t}$ plays a key role in the model of change. According to the generative model, a single $\theta_{i,t}$ is sampled per latent cause per time-step (according to Equation 4.18) and is shared across all weights associating that cause with the stimuli. As highlighted by the autocorrelation function in Equation 4.17, $\theta_{i,t}$ encodes the degree to which the weight values are correlated with their past. In other words, $\theta_{i,t}$ represents the degree to which the contingencies associated with the ith cause change from $t - 1$ to t with larger values of $\theta_{i,t}$ leading to greater degrees of change. If $\theta_{i,t}$ were known, equation 4.18 would fully specify the model of weight change. However, following the insights of Dayan and Yu (2003), $\theta_{i,t}$ is modeled as unknown and subject to inference.

For equation 4.18 to define a sensible change model, $\theta_{i,t}$ must lie in the interval $[0, 1]$. Once again, I employ the logistic sigmoid function, $\theta_{i,t} = \sigma(\eta_{i,t})$, to map $\eta_{i,t} \in \mathbb{R}$ to $\theta_{i,t} \in [0, 1]$. The parameter $\eta_{i,t}$ evolves according to a jump-diffusion process. Jump-diffusion dynamics capture the notion that while rates of change tend to be relatively constant, they are occasionally themselves subject to large changes. Thus when a new latent cause i emerges, $\eta_{i,t}$ is sampled from a Gaussian prior distribution: $\eta_0 \sim \mathcal{N}(\mu_\eta, \sigma_\eta)$. At time t and with probability p_{jump}, the process performs a *jump*, meaning $\eta_{i,t}$ is resampled from the prior. Otherwise with probability $(1 - p_{jump})$, the process evolves according to the diffusion model:

$$\eta_{i,t} = (1 - \kappa)\eta_{i,t-1} + \kappa\mu_\eta + \sqrt{2\kappa}\sigma_\eta\xi_{i,t}, \tag{4.22}$$

where $\xi_{i,t} \sim \mathcal{N}(0, 1)$ and κ plays a role analogous to $\theta_{i,t}$ in Equation 4.18. In the simulations $\kappa = 1 \times 10^{-5}$. Equation 4.22 defines a diffusion in $\eta_{i,t}$ with invariant distribution equal to the prior: $\mathcal{N}(\mu_\eta, \sigma_\eta)$.

Surprising events lead to inferences of larger values of $\theta_{i,t}$ which in turn lead to faster learning—the key insight of Pearce-Hall. One key distinction between this model and that

of Pearce-Hall is that this model attempts to predict all stimuli and not just reinforcement. As a result, any surprising stimulus event—be it an omission of an expected stimulus or the presence of an unexpected stimulus—leads to an increase in the inferred value of $\theta_{i,t}$. These process equations describe a kind of hybrid diffusion over the space of models. They specify an invariant distribution over the model space that represents the subject's *a priori* beliefs concerning the probability of encountering particular patterns of stimuli.

As described above, the non-stationary latent cause model is rather complicated. Figure 4.1 provides a summary of the dependencies among the models constituent random variables. The figure shows a graphical representation of the relationships described in Equations 4.124.22 and the surrounding text.

4.5. Inference over trajectories

As always, the process of conditioning is modeled as the subject's approximation to Bayesian inference over the generative process that gave rise to the training data. Similar to the Kalman filter model of Dayan and Kakade, I model animal learning as a process of recursive Bayesian inference through time, where the posterior at time $t-1$ combines with the models of structural and parameter change to determine the prior distribution at time t,

$$\begin{aligned} P(M_t \mid \mathcal{D}_{t-1}) &= P(m_t \mid m_{t-1}) P(M_{t-1} \mid \mathcal{D}_{t-1}) \\ p(\Theta_{i,t} \mid M_t, \mathcal{D}_{t-1}) &= p(\theta_{i,t} \mid \theta_{i,t-1}, m_t) p(\Theta_{i,t-1} \mid M_{t-1}, \mathcal{D}_{t-1}) \\ p(W_{ij,t} \mid \Theta_{i,t}, M_t, \mathcal{D}_{t-1}) &= p(w_{ij,t} \mid \theta_{i,t}, m_t, w_{ij,t-1}) p(W_{ij,t-1} \mid \Theta_{i,t-1}, \mathbf{m}_{t-1}, \mathcal{D}_{t-1}), \end{aligned}$$

where M_t is the vector $[m_t, m_{t-1}, m_{t-2}, \ldots]$ ($W_{ij,t}$ and $\Theta_{i,t}$ are defined analogously) and $\mathcal{D}_{t-1}$ is the training data presented up to time $t-1$.

The prior at time t is combined—through Bayes rule—with the likelihood of the training stimulus pattern, d_t, given by $P(d_t \mid w_t, x_t, m_t)$ to form the posterior at time t:

$$p(W_t, x_t, \Theta_t, M_t \mid \mathcal{D}_t) \propto P(d_t \mid w_t, x_t, m_t) p(W_t, x_t, \Theta_t, M_t \mid \mathcal{D}_{t-1}), \tag{4.23}$$

where $\mathcal{D}_t = \{d_t, \mathcal{D}_{t-1}\}$ and w_t refers to all $w_{ij,t}$ including the biases $w_{i0,t}$ and w_{0j}. The recursion continues until all the training trials are processed.

With the posterior distribution over the model parameters and structures, one can compute any number of quantities. Of particular interest is the determination of the probability of reinforcement on the test trials—the quantity I take as a correlate to the subject's response magnitude. The probability of reinforcement R, conditioned on the test stimulus pattern $\mathcal{T}$ and the complete training data $\mathcal{D}_T$, is determined by marginalizing over all the

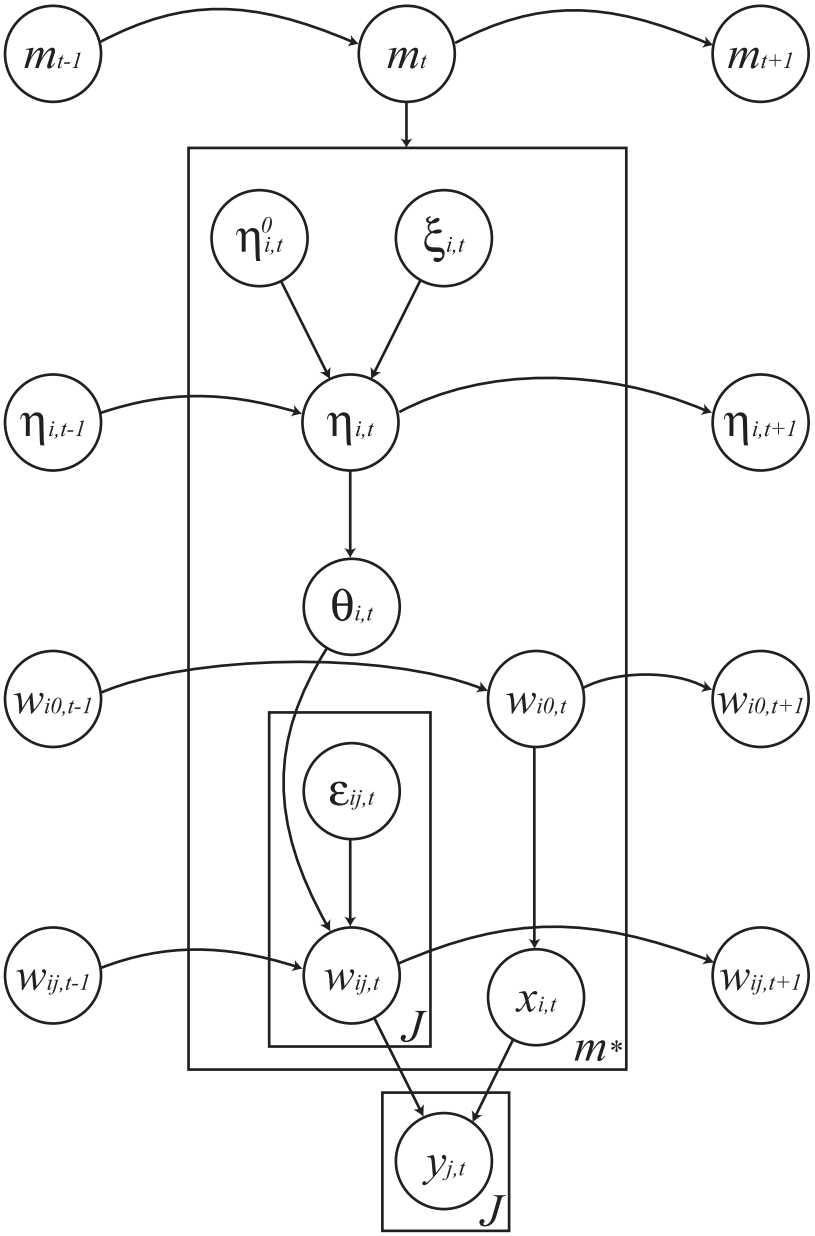


Figure 4.1. A graphical representation of the non-stationary latent cause model using templates. The large box (or template) represents m copies of the variables included inside it (indexed by i), one for each latent cause. Each of these, in turn, contains an additional template representing J copies of the weights $w_{ij,t}$ and $\epsilon_{ij,t}$. Shown in the figure are the dependencies between the variables at time index t and their dependence on the variables at times $t-1$ and $t+1$. The dependencies of the variables on the model parameters are suppressed in the interest of clarity.

unknown parameters, W_T, Θ_T, M_T and x_T:

$$P(R \mid \mathcal{T}, \mathcal{D}_T) = \sum_m \iint \left(\sum_x P(R \mid x, w_T, m_T) p(x \mid \mathcal{T}, w_T, m_T) p(w_T, \theta_T, m_T \mid \mathcal{D}_T) \right) dw_T \, d\theta_T. \tag{4.24}$$

During testing, inference over the latent causes x proceeds as it does during training. The effect of latent cause inference plays a similar role in this incarnation of latent cause theory as it did in the previous two chapters: to effect generalization across observed stimulus patterns.

4.6. Monte Carlo integration

Unsurprisingly, the generative model under consideration is not amenable to analytical inference. In the previous chapter, I described a Markov chain Monte Carlo (MCMC) method to recover samples from the posterior distribution over model structures and parameter values. In the present circumstance with the data no longer IID, using MCMC would require sampling from the whole trajectories over model structures and parameters as they evolve in time. The resulting sample space is prohibitively large and beyond the practical limits of most MCMC methods.

Fortunately, the particle filter is an alternative Monte Carlo based approximate inference scheme that is well suited to the non-IID nature of the problem at hand. The algorithm is essentially an iterative application of importance sampling, where samples are drawn from some proposal distribution and then weighted by the relative "importance" of that sample under the posterior distribution. The particle filter leverages sample populations from the posterior at one time-step to recover samples from the posterior at the next time-step by relying on the substantial correlation between these models.

I adopt a common and simple particle filter approach to sample the unknown parameters from a proposal distribution and weight the resulting samples proportionally to their posterior probability and inversely proportional to their probability under the proposal distribution. The strategy is then to sample from the model structure process, the change process, the weights and the activations of the latent causes and then evaluate the data likelihood to determine the particle weights. The particle filter is presented in Algorithm 1. The proposal probabilities are: destruction of a cause, $q_d = 0.01$; creation of a cause, $q_b = 0.05$; resample from $\eta_{i,t}$ prior (jump), $q_{jump} = 0.05$. Note that the proposal probabilities of cause destruction and creation, given by q_d and q_b respectively, are distinct from the prior probabilities of cause destruction (ω) and creation (ρ).

The only deviation from a straightforward application of particle filtering involves the sampling of the latent cause activations $x_{i,t}$. The training dataset contained the training

trials interspersed with empty trials representing the intertrial period. Randomly sampling $x_{i,t}$ from its prior would lead to poor performance of the particle filter approximation as both the particles that sample no active causes during non-empty trials and those that sample an active cause during empty trials would receive very small weights—reducing the effective number of particles. Therefore cause activations $x_{i,t}$ are sampled one at a time, in a random order, in proportion to their posterior probability:

$$P(x_{i,t} = 1 \mid d_t, w_t, x_{\bar{i},t}, m_t) \propto P(d_t \mid w_t, x_{i,t} = 1, x_{\bar{i},t}, m_t) P(x_{i,t} = 1 \mid w_t, m_t), \tag{4.25}$$

where $\bar{i}$ represents all latent causes other than i. This is like a single pass of Gibbs sampling through the cause activations. The process is initialized by sampling from the prior.

4.7. Simulations

In this section I present simulations of the non-stationary latent cause model on a number of experiments. The experiments were selected both because of their relevance to the question of how animals reason about change and because they highlight various aspects of the non-stationary latent cause model. All simulations were run with 50,000 particles.

4.7.1. Acquisition and extinction

The exploration of the behavior of the non-stationary latent cause model begins with the basic conditioning phenomena of acquisition and extinction. As shown in Figure 4.2, the model successfully tracks changing reinforcement contingencies through both acquisition training ($A+$) for the first 25 trials and extinction training (A) for the remaining 25 trials.

Figure 4.2 reveals an interesting aspect of the non-stationary latent cause model: the rapid rate of initial acquisition. This property is a result of model structure inference. When acquisition training begins in the "naive" model, there are no causes to claim responsibility for the $A+$ trials and the resulting data likelihood is very low. Thus when one or a few particles are sampled that includes a latent cause close to the high likelihood region (in this case with large positive weights on A and reinforcement and small or negative weights on all other stimuli), they are assigned a large portion of the probability mass through particle weighting. At that point the probability of reinforcement given A makes a significant abrupt increase. While this is not the typical acquisition curve there is some recent evidence (Gallistel, Fairhurst, & Balsam, 2004) that this is qualitatively closer to actual animal behavior than the gradually increasing curved predicted by Rescorla and Wagner (1972) among others. This point is discussed further in the concluding chapter. Figure 4.2 shows that in

Algorithm 2 Description of the particle filter approximate inference algorithm for the non-stationary latent cause model. The quantity $q^{(k)}$ is the proposal probability for particle k and $p^{(k)}$ is the prior probability of the proposed update for particle k under the established model assumptions.

for all time t **do**
  for all particles k **do**
    for all causes i **do**
      Sample f from Uniform(0,1)
      if $f < q_d$ **then**
        Destroy the ith cause
        Set $q^{(k)} = q_d$ and $p^{(k)} = \rho$
      else
        Set $q^{(k)} = (1 - q_d)$ and $p^{(k)} = (1 - \rho)$
      end if
    end for
    Sample g from Uniform(0,1)
    if $g < q_b$ **then**
      Create uth cause
      Set $q^{(k)} = q^{(k)} \times q_b$ and $p^{(k)} = p^{(k)} \times \omega$
    else
      Set $q^{(k)} = q^{(k)} \times (1 - q_b)$ and $p^{(k)} = p^{(k)} \times (1 - \omega)$
      Sample $\eta_{u,t}^{(k)}$ from $\mathcal{N}(\mu_\eta, \sigma_\eta)$
    end if
    for all causes i **do**
      Sample h from Uniform(0,1)
      if $h < q_{jump}$ **then**
        Set $q^{(k)} = q^{(k)} \times q_{jump}$ and $p^{(k)} = p^{(k)} \times p_{jump}$
        Sample $\eta_{i,t}^{(k)}$ from $\mathcal{N}(\mu_\eta, \sigma_\eta)$
      else
        Set $q^{(k)} = q^{(k)} \times (1 - q_{jump})$ and $p^{(k)} = p^{(k)} \times (1 - p_{jump})$
        Sample $\epsilon_{ij,t}^{(k)}$ from $\mathcal{N}(0, 1)$
        Compute $\eta_{i,t}^{(k)}$ and $\theta_{i,t}^{(k)}$ from Equation 4.22
      end if
      for all stimuli j **do**
        Sample $w_{ij,t}^{(k)}$ from Equation 4.18
      end for
    end for
    for all causes i **do**
      Sample $x_{i,t}^{(k)}$ from Equation 4.25 and set $q^{(k)}$ accordingly
      $p^{(k)} = p^{(k)} \times P(x_{i,t}^{(k)} \mid w_t^{(k)}, m_t^{(k)})$
    end for
    Compute weight $s_t^{(k)} \propto s_t^{(k)} P(d_t \mid w_t^{(k)}, x_t^{(k)}, m_t^{(k)}) p^{(k)} / q^{(k)}$
  end for
  Normalize $s_t^{(k)} \leftarrow s_{t-1}^{(k)} / \sum_k s_t^{(k)}$
  (Optional) Resample particles from weights $s_t^{(k)}$
end for

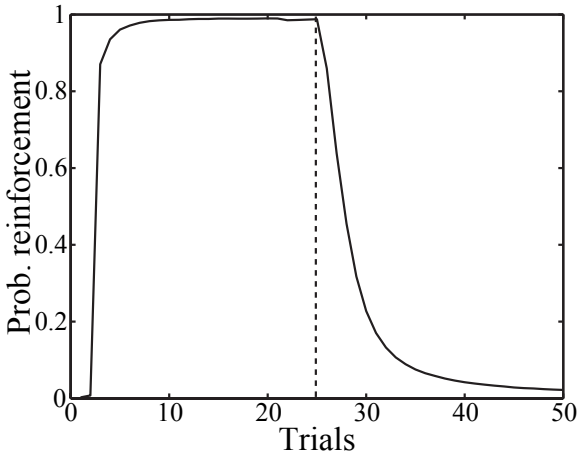


Figure 4.2. A simulated learning curve (responding to A) through an acquisition training phase ($A+$) for the first 25 trials, followed by an extinction training phase ($A-$) at the dashed line.

the early stages of extinction, the model predicts a precipitous decline in the expectation of reinforcement. This property of the model appears to be a poor quantitative match to the behavior of animals. Animals undergoing extinction tend to taper off conditioned responding gradually with presentations of unreinforced stimuli (Gallistel & Gibbon, 2000).

4.7.2. Latent inhibition

Latent inhibition describes the retardation of acquisition resulting from (usually) unreinforced pre-exposure of the target stimulus (i.e. repeated trials of $A-$). Figure 4.3 illustrates the model simulations in both a pre-exposed condition where 300 trials of $A-$ preceded $A+$ acquisition training and a control condition that lacked $A-$ pre-exposure. The inferred model structure (possessing the vast majority of the probability mass) was a single latent cause spanning both training phases. As evident from the figure, the latent cause model successfully accounts for the empirically observed slower rate of acquisition with the pre-exposure treatment. The mechanism responsible of this behavior—the reduction in weight uncertainty with pre-exposure—is the same mechanism for latent inhibition as offered by the Dayan-Kakade Kalman filter model. As the uncertainty in the distribution over weights decreases, more evidence is required to push the weights to the large positive values required to represent acquisition.

4.7.3. Partial reinforcement extinction effect

The partial reinforcement extinction effect is a mainstay of empirical research within the field of animal learning (Fitzgerald, 1963; Gibbon, Farrell, Locurto, Duncan, & Terrace,

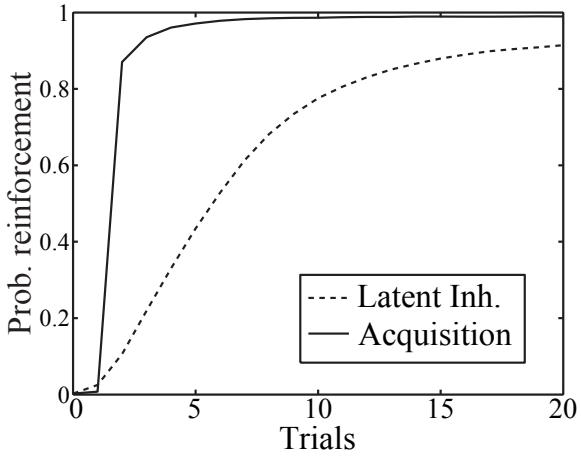


Figure 4.3. Comparison of simulated acquisition curves for a pre-exposed stimulus and a non pre-exposed stimulus. The plot shows only the second phase of reinforced trials ($A+$) and not the 300 pre-exposure trials of $A-$.

1980; Rescorla, 1999; Haselgrove, Aydin, & Pearce, 2004). When animals are presented a sitmulus with a partial schedule of reinforcement (for example, reinforced on only half of the presentations of a stimulus) $A+$ / $A-$, their response to A extinguishes more slowly compared to animals who received continuous reinforcement ($A+$). Despite its long empirical history, the partial reinforcement extinction effect presents a challenge to many models of conditioning. None of the models of Rescorla and Wagner, Pearce and Hall, Dayan and Kakade or Dayan and Yu offer an adequate explanation for the effect.[9] Interestingly, the Pearce-Hall model predicts the opposite of the empirical result. With a partial reinforcement schedule the prediction error remains high and as a result the corresponding associability also remains high which in turn leads to a greater rate of extinction from partial reinforcement relative to the continuous reinforcement.

As shown in Figure 4.4, the non-stationary latent cause is capable of accounting for the partial reinforcement extinction effect. Figure 4.4(a) illustrates the probability of reinforcement as a function of extinction trials ($A-$), after either 400 reinforced stimulus presentations ($A+$) or 200 reinforced stimulus presentations interspersed with 200 non-reinforced stimulus presentations ($A+$ / $A-$). As Figure 4.4(a) shows, the partial reinforcement schedule results in slower extinction. Most probability mass is centered on there being only one latent cause, with significant links to A and reinforcement established during acquisition training (continuous or partial) and then through extinction the weight associated with reinforcement decreases.

[9]The models of Mackintosh (1975) and Gallistel and Gibbon (2000) both offer an explanation of the partial reinforcement effect.

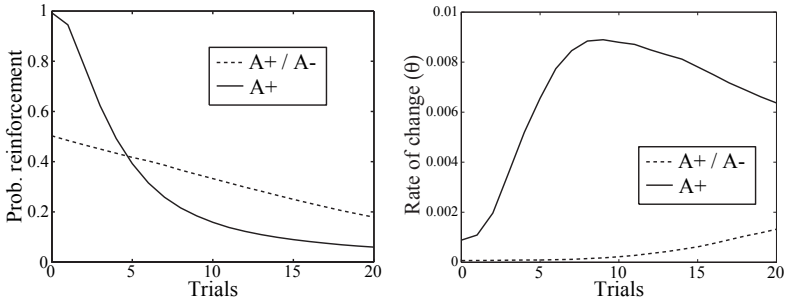
Insight into the model's account of the effect may be gleaned from considering the mean of the inferred distribution over the rate of change parameter $\theta_{1,t}$ as shown in Figure 4.4(b). As the plot reveals, early into the extinction training the change parameter associated with the continuous reinforcement schedule begins to increase significantly in response to the significant prediction error. Speaking at the level of the particle filter inference mechanism, the vast majority of the particles represent models that are expecting reinforcement with high probability and the failure to observe reinforcement results in these particles receiving little weight (low data likelihood). However particles that happen to sample larger values of $\theta_{1,t}$ are likely to move away from this low likelihood model and, by the action of the invariant distribution on the model of change, they more likely to move to closer to a more likely model with a smaller reinforcement weight. As a result these particles are more heavily weighted in the particle filter algorithm and this in turn leads to the distribution over $\theta_{1,t}$ increasing in value (on average). The effect of the increase in $\theta_{1,t}$ is, of course, to beget faster extinction as, according to Equation 4.18, the weights are permitted to move more each time-step. Thus there is a kind of positive feedback loop resulting in faster extinction than would be found in, for instance, the Kalman filter model of Dayan and Kakade. In the case of partial reinforcement, extinction trials are not nearly so unlikely (in fact they occur on every second trial), and as a result particles that sample larger values of $\theta_{1,t}$ are not weighted as highly as in the continuous reinforcement case. The net effect is that the whole positive feedback process is slowed down and extinction proceeds more slowly.

The explanation offered by the non-stationary latent variable model for the partial reinforcement extinction effect is, in essence, that the shift between partial reinforcement and extinction is less dramatic than the shift between continuous reinforcement and extinction. Less of a change is inferred and consequently there is less uncertainty about the model weights which in turn results in a slower adaptation of their values.

4.7.4. Unblocking

The downwards unblocking experiment of Dickinson et al. (1976) is ill-suited to the stimulus representation scheme of the latent cause model. The model represents stimuli as atomic entities that are either present or absent. Changes in reinforcement magnitude, or the occurrence of two reinforcements as in the case of Dickinson et al. (1976), are not easily expressed within this framework.[10] Thus I will turn to a closely related experiment by

[10]It's important to point out that this has nothing to do with the either the change model, nor the claim that animals are attempting to recover the causal structure of the environment. It is a consequence of a modeling choice made to simplify the description of stimuli.



(a) Simulated learning curve (b) Rate of change: $\theta_{1,t}$

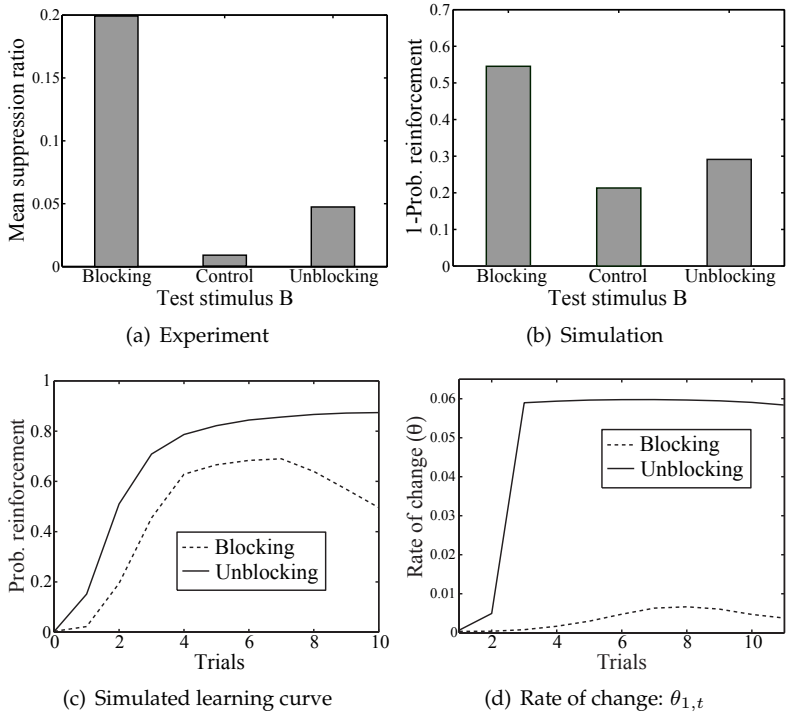
Figure 4.4. The partial reinforcement extinction effect. (a) Simulated response to A during extinction training ($A-$) for both a continuously reinforced stimulus $A+$ and a partially reinforced stimulus (50% reinforcement schedule) $A+$ / $A-$. (b) The corresponding average value of $\theta_{1,t}$.

Group	Phase 1	Phase 2	Test $\rightsquigarrow$ Result
Control	A–*footshock*	CB–*footshock*	$B \rightsquigarrow Resp.$
Blocking	A–*footshock*	AB–*footshock*	$B \rightsquigarrow -$
Unblocking	A–*footshock*	AB–*ice-water-dunking*	$B \rightsquigarrow Resp.$

Table 4.2. Experiment of Blaisdell et al. (1999).

Blaisdell et al. (1997): unblocking with a qualitative change of reinforcer (also see Holland (1988)). Here Blaisdell et al. trained subjects on presentations of A–*footshock*. Following this treatment, some of the subjects (the unblocking group) received pairings of AB–*ice-water-dunking*, others (the blocking group) received pairings of AB–*footshock* and still others (the control group) received pairings of CB–*footshock* where both C and B were novel stimuli. With respect to the blocking group, the unblocking group experienced significant unblocking, meaning experimental subjects responded significantly to presentations of the would-be blocked stimulus B. In designing the experiment, Blaisdell et al. took care to balance the two reinforcers to ensure that they had equal potential to elicit responding. The experiment is summarized in Table 4.2.

In accounting for this finding, the latent cause model was simulated with 150 trials of A–*footshock* before further training with either AB–*footshock* (blocking group) or AB–*ice-water-dunking* (unblocking group). The results of the simulation are presented in Figure 4.5. As is apparent from a comparison of Figures 4.5(a) and 4.5(b), the latent cause model captures the basic effect of unblocking. That is, a qualitative change in the reinforcer—represented in the model as a change from one atomic stimulus to another—results in a greater probability of reinforcement at test time.



(a) Experiment (b) Simulation

(c) Simulated learning curve (d) Rate of change: $\theta_{1,t}$

Figure 4.5. Unblocking with a qualitative change in reinforcement. (a) The experimental results of Blaisdell et al. (1997). The mean suppression ratio measures the level of bar pressing for water in the presence of the test stimulus relative to bar pressing in the absence of the stimulus. Smaller values of the ratio indicate a greater degree of conditioning. (b) The simulated mean probability of reinforcement after 3 $AB+$ trials (where $+$ is *footshock* for the blocking and control groups and *ice-water-dunking* for the unblocking group). (c) The simulated mean probability of reinforcement and (d) the mean value of $\theta_{1,t}$ as functions of the number of trials for both the blocking and unblocking training schedules.

Figures 4.5(c) and 4.5(d) illustrate the learning curves (mean probability of reinforcement) and the corresponding mean values of the change parameter $\theta_{1,t}$ for the blocking and unblocking simulations. These plots reveal that inference over change is a key mechanism of the latent cause theory's account of unblocking. During the first phase of A–*footshock*, the relative stability in the training sequence results in a relatively small value being inferred for $\theta_{1,t}$ (the majority of the probability mass rests on there being only one latent cause). Then, for the unblocking group, the qualitative change in reinforcement on top of the addition of novel stimulus B constitutes significant evidence that the value of $\theta_{1,t}$ has increased (as reflected in Figure 4.5(d)). In comparison, the blocking group only observed the addition of the novel stimulus B and consequently the inferred change is relatively minor. As a consequence of the difference in inferred values of $\theta_{1,t}$, the unblocking group learns faster, as is demonstrated in Figure 4.5(c).

4.7.5. Latent inhibition and overshadowing counteract each other

Much of the literature discussing the Pearce-Hall model rather casually describes a mechanism where associabilities are boosted when a surprising event occurs. Throughout this chapter, I have endeavored to be a bit more careful with this description, emphasizing that associabilities are boosted when surprising *reinforcement* events occur. Since the Pearce-Hall model—like the vast majority of conditioning models—is discriminative, reinforcements are the only events that are predicted and hence the only ones that can be surprising.[11]

The non-stationary latent cause model makes no distinction between reinforcement and other stimuli, with the weights evolving identically for each stimulus (including reinforcement). As a consequence, the model predicts that unexpected neutral stimuli should also be capable of inducing faster learning. The experiments that bear on this question seem to support the latent cause perspective. The clearest examples of these are the experiments showing that a change in context (Dexter & Merrill, 1969; Lantz, 1973; Lubow, Rifkin, & Alek, 1976), the addition of unexpected stimuli (Blaisdell et al., 1998), and the omission of expected stimuli (Wilson, Boumphrey, & Pearce, 1992; Holland, 1997) all serve to restore rates of acquisition to pre-exposed stimuli in a manner similar to the reinforcement manipulations of Hall and Pearce (1982) discussed above. This and the following sections present a number of simulations of the latent cause model highlighting this point.

One of the clearest examples of the effect of surprising neutral stimuli on the rate of learning is offered by an investigation of Blaisdell et al. (1998). They trained three groups of rats in a two-phase procedure[12] described in Table 4.3. One group was trained with a typical overshadowing paradigm, the second group was trained with a typical latent inhibition paradigm and the third group was trained with a paradigm combining latent inhibition with overshadowing. Blaisdell et al. showed that by combining the two procedures, the two effects—both known to reduce the response relative to simple acquisition training—seem to counteract each other with the two together leading to greater response than either separately. From the point of view of latent inhibition, this implies that the introduction of a novel stimulus, at the transition from pre-exposure training to acquisition training, causes a reduction of the latent inhibition effect.

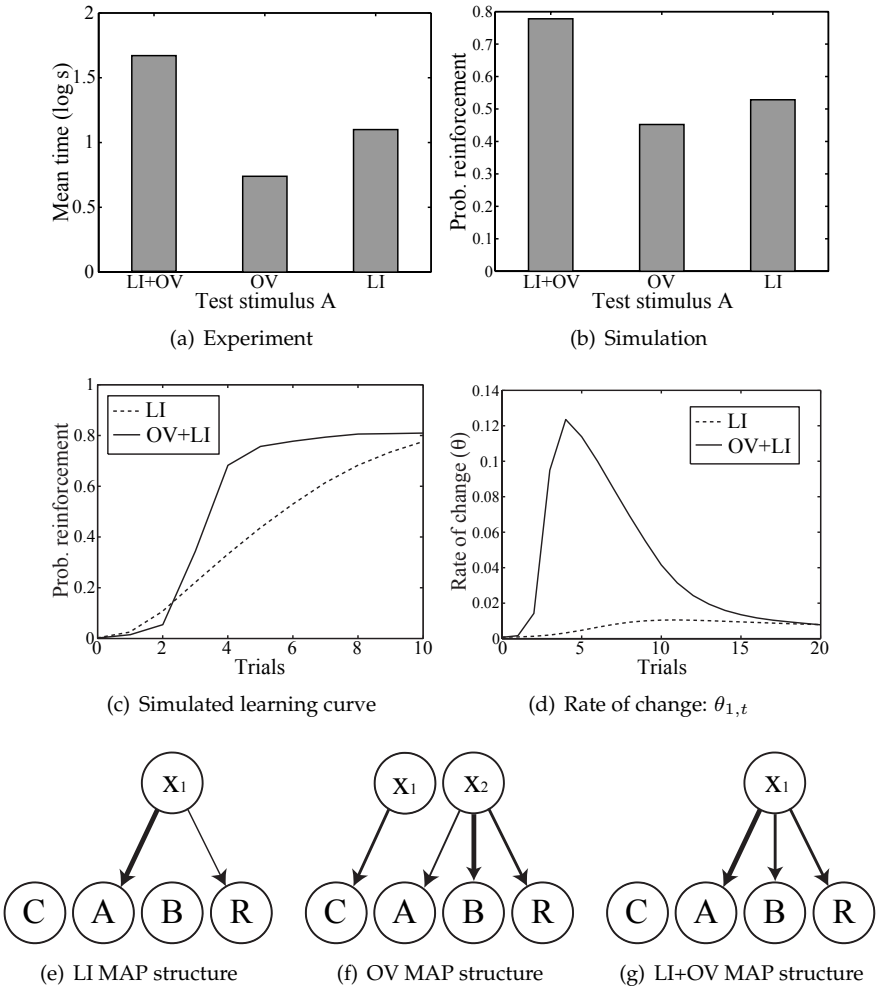
To produce a significant overshadowing effect Blaisdell et al. used a relatively low salience stimulus A relative to the overshadowing stimulus B. In order to capture this

[11] This is also the case for the normative Dayan-Kakade and Dayan-Yu models.

[12] Blaisdell et al. also trained two additional control groups that I will not consider.

Group	Phase 1	Phase 2	Test $\rightsquigarrow$ Result
LI + OV	$A-$	$AB+$	$A \rightsquigarrow Resp.$
OV	$C-$	$AB+$	$A \rightsquigarrow -$
LI	$A-$	$A+$	$A \rightsquigarrow -$

Table 4.3. Summary of Experiment 1 of Blaisdell et al. (1998). LI is latent inhibition and OV is overshadowing.



(a) Experiment (b) Simulation

(c) Simulated learning curve (d) Rate of change: $\theta_{1,t}$

(e) LI MAP structure (f) OV MAP structure (g) LI+OV MAP structure

Figure 4.6. Latent inhibition and overshadowing counteract each other. (a) The experimental results of Blaisdell et al. (1998). The response is measured in the time to takes 5 cumulative licks of water. (b) The mean simulated probability of reinforcement after 6 $AB+$ trials (phase 2). (c) The simulated probability of reinforcement as a function of the number of phase 2 trials. (d) the mean value of $\theta_{1,t}$ as a function of the number of phase 2 trials for the latent inhibition (LI) and latent inhibition plus overshadowing (LI+OV) training schedules. (e)–(g) The MAP model structures (with small weights excluded) for each of the training groups at the point of testing: after 6 trials of Phase 2 training.

effect, I require some notion of saliency to render the stimuli unequal. There are two mechanisms within the latent cause model that could reflect saliency: the priors over $w_{ij,t}$ (with hyperparameter α) and the stimulus biases $w_{0j,t}$. Together they dictate how much behavioral control one stimulus could exert over another through their mutual influence on the inference process over the latent cause (see section 3.4.2 for details). Therefore the less salient stimulus is represented with $\alpha = 2.5$, $w_{0j,t} = -5$ and the more salient stimulus is represented with $\alpha = 3.5$, $w_{0j,t} = -7$. For all other stimuli $\alpha = 3$, $w_{0j,t} = -6$.

Figure 4.6 illustrates the latent cause account of the Blaisdell et al. experiment. According to the model, stimulus pre-exposure counteracts overshadowing by establishing a strong association between the overshadowed stimulus A and the latent cause. When overshadowing training ($AB+$) begins, the latent cause encoding the presence of A also begins to encode B and reinforcement, but the pre-exposure training allows A to overcome its low saliency and play a greater role in the inference over the latent cause (recall the explanation of overshadowing from the previous chapter). Thus, during testing with A alone, its increased weight, relative to the overshadowing condition with no A pre-exposure, leads to an increased probability that the latent cause is active. The increased probability in turn leads to the empirically observed increase in probability of reinforcement.

The experiment affords an opportunity to examine the interaction between structural and parametric change during the inference process. As described above, in the case of stimulus pre-exposure training preceding overshadowing training (Group LI+OV in Table 4.3), most probability mass is assigned to the interpretation that the Phase 2 trials of $AB+$ are derived from a modified version of the Phase 1 trials of $A-$, due to the overlapping stimulus A. As shown in Figure 4.6(g), both phases of training are attributed to a single, albeit changing latent cause. In contrast, the training schedule of Group OV (in Table 4.3) of overshadowing ($AB+$) preceded by pre-exposure to a third stimulus C results in the majority of probability mass being assigned to models with two latent causes. As shown in Figure 4.6(f), one latent cause accounts for $C-$ trials; while the other, encoding the $AB+$ trials, is inferred to have come into existence with the onset of overshadowing training. The lack of any overlapping stimuli caused the particle filter algorithm to infer that these two phases of training resulted from two distinct latent causes, with neither experiencing significant parametric change.

The account of how the addition of the neutral stimulus B alleviates the effects of latent inhibition is more relevant to the present discussion of change and its effect on behavior than is the model's account of how pre-exposure mitigates overshadowing. The transition between the pre-exposure training phase and the overshadowing phase is marked

Group	Phase 1	Phase 2	Test $\rightsquigarrow$ Result
SOC	$A+$	$AB-$	$B \rightsquigarrow Resp.$
Blocking	$A+$	$AB+$	$B \rightsquigarrow -$

Table 4.4. Summary of an experiment of Cheatle and Rudy (1978). The SOC Group received second-order conditioning training.

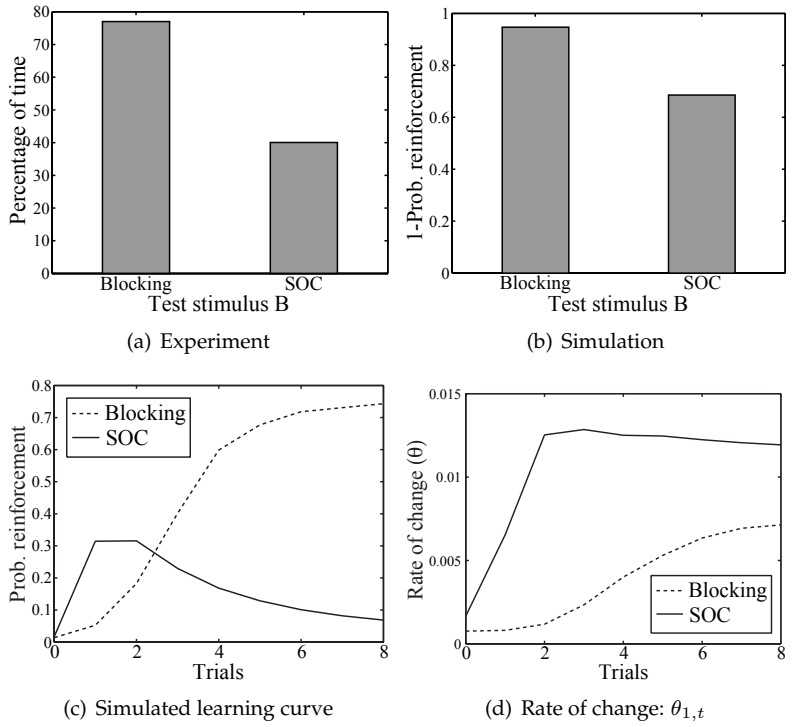
by the introduction of both a reinforcer and stimulus B. As depicted in Figures 4.6(c) and 4.6(d), the introduction of these two stimuli constitute significant evidence of an increased value of $\theta_{1,t}$. Larger values of $\theta_{1,t}$ result in greater uncertainty in the value of the weights on all links emanating from that latent cause via Equation 4.18. This, in turn, leads to an increased rate of acquisition for A– mitigating the retarding effect of latent inhibition.

4.7.6. Second-order conditioning and blocking

The addition of the model of change to the latent cause model offers an opportunity to explain the puzzling finding of an experiment of Cheatle and Rudy (1978) exploring the relationship between second-order conditioning and blocking. The typical training protocols of these two conditioning phenomena are strikingly similar. Both begin with a phase of acquisition training $A+$, followed by a phase of training with $AB-$ trials for second-order conditioning or $AB+$ trials for blocking.[13] Using sequential versions of both second-order conditioning and blocking, Cheatle and Rudy demonstrate that the second-order conditioning protocol establishes greater responding to B than does the blocking protocol (their experiment is summarized in Table 4.4). This result is striking as the only difference between these two procedures is the presence of reinforcement in the second phase of the blocking procedure. Somehow the presence of reinforcement in the second training phase results in a decrease in responding to B. Cheatle and Rudy suggest that this result may be due to an associative memory interference effect with the presentation of reinforcement disrupting the formation of an association between B and A. The non-stationary latent cause model proposes a different account.

Figure 4.7 shows the experimental phenomenon explored in the Cheatle and Rudy investigation and the latent cause simulation results. The simulations involved 100 trials of $A+$ followed by 8 trials of either $AB-$ training for the second-order conditioning (SOC) group or $AB+$ training for the blocking group. Comparing Figures 4.7(a) and 4.7(b), the latent cause model demonstrates the empirically observed effect of greater responding to B

[13] Usually second-order conditioning is established in a sequential paradigm, with second phase training $B \rightarrow A \rightarrow +$, and blocking is typically established with the simultaneous presentation of A and B, though both simultaneous second-order conditioning (Yin et al., 1994) and sequential blocking (Cheatle & Rudy, 1978) have been reported.



(a) Experiment (b) Simulation

(c) Simulated learning curve (d) Rate of change: $\theta_{1,t}$

Figure 4.7. Second-order conditioning and blocking. (a) The experimental results of Cheatle and Rudy (1978). Responses are measured in the percentage of time spent on in the presence of the odor stimulus. Subjects tend to spend less time in the presence of the stimuli that are predictive of reinforcement. (b) The quantity 1-probability of reinforcement (to correspond to the behavior measure of Cheatle and Rudy (1978)) when testing with B after one trial one trial of either AB (SOC group) or $AB+$ (blocking group). (c) The mean probability of reinforcement as a function of training. (d) The mean value of $\theta_{1,t}$ as a function of training.

in the SOC condition than in the blocking condition. According to the latent cause model, the addition of B together with the absence of the expected reinforcement during SOC training is strong evidence in favor of a significant change in the generative model and this is reflected in the inferred value of $\theta_{1,t}$ and consequently in the rate of acquisition of the weight associated with B (as shown in Figures 4.7(c) and 4.7(d)). When testing with B, the greater value of $w_{1B,t}$ relative to the blocking condition causes inference over latent cause activation to place significantly more probability mass on the corresponding latent cause being active.

As is apparent from Figure 4.7(c), the latent cause model predicts that both second-order conditioning and blocking are transient phenomena. In the case of second-order conditioning, the rapid decrease (extinction) of the weight associating the latent cause and reinforcement results—according to the model—in a decrease in responding to B. Meanwhile, in the case of blocking, the weight $w_{1B,t}$ eventually approaches the magnitude of

that of $w_{1,A}$ and blocking gives way to an overshadowing effect (i.e. blocking disappears with respect to a control procedure that does not receive $A+$ pre-training). The transient nature of both second-order conditioning (Yin et al., 1994) and blocking (Miller, 2002) has been empirically established.

4.8. Discussion

In this chapter, I have set out to correct a central weakness of the model presented in the previous chapter: the assumption of stationarity and the resulting disregard of trial order effects. This issue was addressed by incorporating a generative model of change into the latent cause framework. The resulting non-stationary latent cause model offers a unique perspective on how animals reason about change in their environment. Through the mechanism of inference over change, the model is capable of capturing classic Pearce-Hall effects such as latent inhibition and unblocking. But more importantly, by the underlying latent cause structure, the model generalizes the previous notion of how *associabilities* are manipulated to include a role for surprising neutral stimuli. There appears to be broad empirical support for the non-stationary latent cause perspective.

4.8.1. Perspectives on change

The model of change developed in this chapter shares some important characteristics with that of Dayan and Yu (2003) . Both models propose probabilistic generative models that describe how the model parameters and structure evolve over time. They both offer the same mechanism of inference over a rate of change parameter ultimately permitting observations to influence the learning rate. However they differ in how they treat the dynamics governing change itself.

The Dayan-Yu model assumes that the amount of change is usually fixed to a known quantity and that this is occasionally interrupted by discrete rapid change events. Inference over change amounts to inferring whether or not a rapid change event has occurred. Once a change event occurs, the change parameter returns to its usual value.

The present model of change takes a somewhat different perspective. While it also allows for sudden changes in the rate of change (due to the jump process), the value of the change parameter, $\theta_{i,t}$, is also assumed to change slowly over time. This property gives the model the capability to infer that the environment is in a state of relatively greater change over extended durations and should, in principle be able to account for phenomena such as the empirical effect of increasing learning rates with successive retraining (Pavlov, 1927;

Rescorla, 2003b). However, interestingly this experiment presents a particular challenge to the non-stationary model of change that is worth discussing in some detail.

For naive subjects, the acquisition of a conditioned response to a novel stimulus is relatively slow compared to reacquisition to the same stimulus after it had been extinguished through presentations of the stimulus in isolation. It has been observed (Pavlov, 1927) that multiple consecutive phases of acquisition and extinction training lead to increasingly fast acquisition and extinction (as measured by achieving some behavioral criterion for each).

Through inference over the unknown rate of change parameter $\theta_{i,t}$ the non-stationary latent cause model is capable of tracking changes in the rate of change. The experience of successive acquisition and extinction training phases cause progressively increasing values of the corresponding $\theta_{i,t}$, which in turn leads to increasingly rapid learning. The model accurately describes the empirical phenomenon for every acquisition cycle save the first. According to the model, the first acquisition phase causes an inferred model structure change: the addition of a new latent cause encoding the relationship between the stimulus and reinforcement. There is a mismatch between the parametric model of change and the structural model of change with the onset of structure change being typically too rapid (as shown in the acquisition curve of Figure 4.2) relative to the dynamics of parametric change. It seems one could simply adjust the prior on structural change to make it more improbable and slow acquisition due to structural change. However the wider issue of how to balance these relatively independent models of change remains. It is not clear that the distinction between parametric change and structural change as described in this chapter is ultimately justified by the empirical behavior of animals. A more parsimonious account would involve either a single mechanism for change in both structure and parameters or some notion of dependence or interaction between these two processes.

The impact of the invariant distribution of the model of change is not particularly apparent in the exposition of the simulation results; however, there is one critical consequence of its existence that is worth a brief mention. Through the mean reverting property of the model of change, the invariant distribution prevents the values of the weights from growing too large. With excessively large weights, the generalization property of the model—as mediated through inference over the latent variables—would be compromised.

Consider a latent cause i encoding associations between stimuli A and B and reinforcement and imagine that the weight associated with each of these elements is extremely large with respect to the fixed bias term ($w_{0j} = -6$). If such a model were presented A alone, the absence of an expected B would be so improbable with large values of $w_{iB,t}$ that nearly all probability mass would be assigned to the latent cause being inactive and

complete overshadowing would result. This situation would also compromise most of the phenomena described in this chapter that rely on some amount of generalization in order to incorporate two distinct training phases on a single latent cause. The existence of the invariant distribution permits weights to diffuse and yet remain within a reasonable range of magnitude.

One crucial component of the non-stationary latent cause model is the sharing of the change parameter $\theta_{i,t}$ across all weights associated with a given cause. This is how the surprising appearance of one stimulus ultimately leads to an increased learning rate for a second stimulus. While sharing $\theta_{i,t}$ across weights attached to a single latent cause is clearly important, it is unclear whether it is appropriate to envision independent $\theta_{i,t}$ for each latent cause as I have done, or rather if these $\theta_{i,t}$ might be correlated, with evidence of change within one cause affecting reasoning about change within other causes. At present, there seems to be little empirical data to inform these sorts of speculations.

4.8.2. Inference over structure, parameters and change

From one perspective, the model of change enters the negotiation of the tradeoff between model complexity and data fitting as a simplifying influence: it is much easier for a single cause to explain a variety of data if it is permitted to change its parameters between observations. On the other hand, a rapidly changing model is much more complicated in the marginal likelihood sense[14] than a slowly changing model. All of these considerations are handled seamlessly within the Bayesian framework. Animals too must negotiate between competing interpretations of their environment and decide, for example, whether a surprising observation represents a change in existing contingencies or the introduction of something entirely new. The non-stationary latent cause model offers a perspective on the computational principles underlying these issues.

While the particle filter algorithm does permit inference over model trajectories, unfortunately it appears to be significantly less powerful in terms of inference over model structures than the MCMC method, employed in the previous chapter. As a result, the size of the models that can be reliably inferred from data is much more restricted in this case. Fortunately, most phenomena that appear to exercise inference over change also appear to require relatively small models with only one or two latent causes.

[14] See section 3.3 for a discussion of marginal likelihoods.

4.8.3. Unblocking

Finally, I wish to raise a point regarding one of the experiments that I considered in this chapter that is not directly concerned with the issue of change and learning rates. The unblocking experiment of Blaisdell et al. (1997) seems to be in conflict with the experimental phenomenon of transreinforcer blocking (Williams, 1994; Ganesan & Pearce, 1988). As the name suggests, transreinforcer blocking experiments demonstrate continued cue competition effects despite a qualitative change in reinforcement. The latent cause model clearly predicts some sort of unblocking resulting from a change in reinforcer and as such also finds itself in conflict with transreinforcer blocking.

One potential solution to this conflict is offered by a recent theory of Daw, Niv, and Dayan (2005) who, with regard to instrumental conditioning phenomena, suggest that there are two parallel models of behavior competing for behavioral control. One model is generative, similar to the kind considered in this thesis, the other is discriminative. While Daw et al. (2005) were primarily interested in how these models interact in decision making, their perspective offers an interesting solution to the present unblocking/transreinforcer blocking conflict. It seems reasonable to suggest that while the generative model represents reinforcer identity, the discriminative model is restricted to encode some quantitative value of reinforcement. When presented with the unblocking experiment of Blaisdell et al. (1997), the generative system would predict unblocking (as shown above), while the discriminative system, with no means to represent the qualitative difference in reinforcer, would predict blocking. The experimental parameters controlling which of these alternative model predictions dominate at any given point remains to be elucidated and is beyond the scope of this thesis.

CHAPTER 5

Learning temporal structure

The models presented in the previous two chapters were trial level models, with a single trial being the basic unit of time. While this kind of abstraction is typical of models of conditioning, extensive experimentation has revealed that the temporal structure of stimuli within the trial has a significant impact on learning and behavior. In this chapter, I consider how the latent cause theory can be used to gain some insight into the nature of time and its role in learning and behavior.

As described in Chapter 2, the time between stimulus and reinforcement has long been understood to play an important role in conditioning experiments. For instance, a stimulus presented immediately preceding a reinforcer (forward conditioning) supports robust responding. However, if the stimulus onset occurs simultaneously with or after the reinforcer onset (simultaneous and backward conditioning respectively), then responding to the stimulus is significantly weakened. The temporal difference (TD) model of conditioning (Sutton & Barto, 1990) and others have been developed to account for these sorts of temporal phenomena. They suggest that the stimulus is only useful as a predictor if it immediately precedes reinforcement. If the stimulus follows reinforcement, or precedes it by an extended interval, then the predictive ability of the stimulus is compromised and learning is subsequently impaired. While the explanation offered by the TD style models is broadly consistent with canonical temporal conditioning phenomena, a more recent line of experimentation by Miller and colleagues raises serious questions regarding this interpretation of the role of time in learning (Cole, Barnet, & Miller, 1995; Cole & Miller, 1999; Barnet, Grahame, & Miller, 1993; Barnet & Miller, 1996).

In an attempt to explain these later experiments, Miller and Barnet (1993) proposed the *temporal coding hypothesis* (TCH). As will be described in more detail in Section 5.2, the TCH states that animals learn *timeline* representations of trials, encoding the temporal relationships among all stimuli. The notion of a timeline of stimuli may be naturally expressed

	Name	Phase 1	Phase 2	Test ⇒ Result	Test ⇒ Result
Expt. 1	Simultaneous Cond.	(4)$T+$	(4)$C \rightarrow T$	$T \Rightarrow$ –	$C \Rightarrow$ Resp.
Expt. 2A	Sensory Precond.	(12)$T \rightarrow C$	(8)$T \rightarrow +$	$C \Rightarrow$ –	
Expt. 2B		(12)$T \rightarrow C$	(8)$T \longrightarrow +$	$C \Rightarrow$ Resp.	
Expt. 3	Backward Cond.	(96)$L \rightarrow + \rightarrow X$	(8)$B \rightarrow X$	$X \Rightarrow$ –	$B \Rightarrow$ Resp.

Table 5.1. Experimental Paradigms. Phases 1 and 2 represent two stages of training trials, each presented (n) times. The short arrow ($\rightarrow$) indicates one stimulus immediately following another; and the long arrow ($\longrightarrow$) indicates a 5 sec gap between stimulus offset and the following stimulus onset (simultaneous stimuli are written together). For Expt. 1, the tone T, click train C, and footshock reinforcer $+$ were all of 5 sec duration. For Expt. 2, the tone and click train durations were 5 sec and the footshock $+$ lasted 0.5 sec. For Expt. 3, the light L, buzzer B, and auditory stimulus X (either a tone or white noise) were all of 30 sec duration, while the footshock $+$ lasted 1 sec. Resp. indicates a conditioned response to the test stimulus.

in the latent cause framework by extending the binary causes of the previous two chapters to a multinomial variable encoding a latent chain of states. Under this scheme, the model may be interpreted as a hidden Markov model (HMM) and constitutes a computational version of the TCH.

In this chapter I present three of the Miller et al. experiments and I show how the existing computational models of conditioning are inadequate to explain these data. I next develop an intratrial version of the latent cause framework and show that the model accounts for the Miller et al. experiments via inference over the latent cause. This chapter concludes with a discussion of the implications of this model to wider computational issues in conditioning.

5.1. Experiments

Over the past two decades a large body of empirical evidence has given rise to the idea that, in the context of classical conditioning, animals explicitly learn the temporal relationship between stimuli. In a series of experiments beginning in the late 1980s Ralph Miller and his colleagues at SUNY Binghamton have systematically demonstrated that in many classical conditioning paradigms (e.g. delay conditioning (Barnet, Arnold, & Miller, 1991a), trace conditioning (Cole et al., 1995), backward conditioning (Cole & Miller, 1999), blocking (Barnet et al., 1993) and conditioned inhibition (Barnet & Miller, 1996)) animals learn the temporal structure of the stimuli presented to them. In this chapter, we focus on three experiments: simultaneous conditioning, sensory preconditioning, and backward conditioning. Table 5.1 summarizes the three procedures and the behavior results.

5.1.1. Experiment. 1: Simultaneous conditioning

Response to a stimulus is impaired when it is presented simultaneously with reinforcement rather than preceding the reinforcement. As previously mentioned, the failure of the simultaneous conditioning procedure to elicit a response is a well established result in the classical conditioning literature (Pavlov, 1927). Barnet et al. (1991a) reported an interesting second-order extension of the classic paradigm. In the first phase of training, they presented subjects with a tone stimulus (T) simultaneously with a footshock ($+$). In the second phase, the subjects experienced a click train (C) preceding T. Barnet et al found that while testing with T itself results in minimal response, testing with C evoked a significant freezing response from the subjects.

5.1.2. Experiment. 2: Sensory preconditioning

In a second experiment, Cole et al. (1995) exposed rats to a tone (T) immediately followed by a click train (C). In a second phase, the tone was followed by a footshock ($+$) that occurred either coincident with click train onset (variant A), or 5 sec later, coincident with click train offset (variant B). They found that when the click train and shock both immediately followed the tone, little conditioned response was elicited by the click train. However, when the shock occurred 5 sec after tone offset and click train onset, then the click train did come to elicit a response.

5.1.3. Experiment. 3: Backward conditioning

In another experiment by Cole and Miller (1999), subjects were presented with a flashing light L, followed by a footshock $+$, followed by an auditory stimulus X (either a tone or white noise). In phase 2, a buzzer B was followed by X. Testing revealed that while X did not elicit a response (in fact, it became a conditioned inhibitor), X seems to have imparted an excitatory association to B which did evoke a response from the subjects.

5.2. Previous models

The existing literature on models of conditioning offers few mechanisms to describe the temporal structure learning described by Miller and colleagues. Many models such as Rescorla-Wagner are trial level models and as such cannot represent within-trial associations. Since the results of the three experimental protocols seem to hinge on the existence of explicit within-trial associations, Rescorla-Wagner fails to predict the various outcomes. It is perhaps more fair to say that these experiments are simply beyond the domain of applicability of Rescorla-Wagner.

There are a number of models that have been developed with the specific goal of capturing the temporal representation of stimuli within a trial. Some of these models are extensions of the temporal difference model of conditioning (Sutton & Barto, 1990), described in Chapter 2, that has successfully accounted for a number of conditioning phenomena that deal specifically with the timing of stimuli.

While the TD model possesses the ability to represent within-trial associations between stimuli, it fails to account for the experiments in Table 5.1. The reason is the overly simplistic mechanism to integrate the timelines of different trials. TD can only impart predictive information about reinforcement held by a stimulus X to stimuli preceding X. In none of the experiments is this kind of temporal integration sufficient to explain the results. In Experiment 1, some versions of TD can account for the reduced associative strength of a CS when its onset occurs simultaneously with $+$, but no version of TD can explain why the second-order stimulus C should acquire greater associative strength than T. In Experiment 2, no learning occurs in Phase 1 with TD because no prediction error is generated by pairing T with C. As a result, no response is elicited by C after T has been paired with $+$ in Phase 2. In Experiment 3, TD fails to predict the results because X is not predictive of $+$; thus X acquires no associative strength to pass on to B in the second phase.

One of the problems with the TD model of conditioning is its focus on predicting the reinforcement signal to the exclusion of all other stimuli. The successor representation of Dayan (1993), the world model of Sutton and Pinette (1985), and the basal ganglia model of Suri and Schultz (2001) all attempt to predict future stimuli in some sense. Though the models differ in how they represent the future stimuli, they all assume the world is Markov with respect to the current vector of stimulus representations,

$$P(x_f \mid x(t), x(t-1), x(t-2), \dots) = P(x_f \mid x(t)), \tag{5.1}$$

where $x(t)$ is a vector representing the stimuli at time t and x_f is a vector representation of future stimuli.

While each of these models improves upon the explanation of conditioning phenomena offered by the original TD model—for example, the model of Suri and Schultz can reproduce the sensory preconditioning effect (Suri & Schultz, 2001)—they all fail to predict the results of the Miller et al experiments for essentially the same reason. They fail because they do not make the necessary inferences about relations among stimuli. By assuming that the world is Markov *in the stimuli* one cannot recover the findings in Table 5.1.

Consider Experiment 1: since T is simultaneously conditioned and not predictive of $+$, it does not amass significant associative strength. Subsequent presentations of C followed

by T may lead to the associative strength of C matching that of T but never exceeding it as the empirical results would suggest. This is the identical failure mechanism as found in the traditional TD model.

Some of these models, such as the Suri and Schultz model, are able to learn that T leads to C in the first phase of Experiment 2. They then learn that T also leads to reinforcement $+$ with some delay. However there is no mechanism that would fuse these two instances into a prediction of reinforcement with the presentation of C as required. In Experiment 3, these models fail for the same reason as they did in Experiment 2. The individual chains of stimuli may be learned, but they are not integrated as required. These models are capable of one kind of chain integration that backs up reinforcement predictions from later times to earlier times, however none of the three experiments described above have this particular structure.

Gallistel and Gibbon (2000) suggest that the combination of scalar expectancy theory (SET) and rate estimation theory (RET), as described in Section 2.6.1, can account for the results in Table 5.1. In their account, they describe how, with a given test stimulus, SET is able to predict when the reinforcer is expected and only initiates a response if the reinforcer is expected in the immediate future. While this is certainly a partial explanation of the Miller et al. experiments, it does not provide any kind of mechanism to explain the vexing issue of how the second-order associative transfer comes about and how time is involved in that learning process. Presumably, this could be explained with the RET and SET models fully integrated; however, the integrated model does not appear to be explicitly described and it is not obvious how such an integration would explain the second-order conditioning experiments under consideration.

To explain and summarize the findings in Table 5.1, Miller et al. have developped the *temporal coding hypothesis* (TCH). It posits that temporal contiguity, rather than reinforcement contingency, is sufficient to produce an association between stimuli. The association is not a simple scalar quantity; instead, information about the temporal relationships among stimuli is encoded implicitly in memory. The temporal coding hypothesis also suggests that memory representations of trials with similar stimuli become integrated, preserving the relative temporal structure of the stimuli.

Though not a formal computational model, the temporal coding hypothesis has considerable predictive power and provides a simple intuitive explanation of the three experiments of section 5.1 and the host of other similar experiments by Miller and colleagues. Applying the temporal coding hypothesis to Experiment 1, we get the integrated memory

representation, $C \rightarrow T+$. If we interpret the response as a prediction of imminent reinforcement, then we arrive at the correct prediction of a strong response to C and a weak response to T. Integrating the hypothesized memory representations of the two phases of Experiment 2 results in: (A) $T \rightarrow C+$ and (B) $T \rightarrow C \rightarrow +$. The stimulus C is only predictive of reinforcement in variant B, consistent with the experimental findings. For Experiment 3, an integrated memory representation of the two phases produces $LB \rightarrow + \rightarrow X$. Stimulus B is predictive of reinforcement while X is not. By first creating and then integrating event timelines, the temporal coding hypothesis is able to account for the results of each of the three experiments.

5.3. A latent cause model of temporal structure

In the remainder of this chapter, I explore a version of the latent cause model that embodies the principles of the temporal coding hypothesis and successfully predicts the results of the three experiments highlighted in Table 5.1. A key idea we take from the TCH is the need to represent a timeline of stimuli. A straightforward formalization of a timeline is a Markov chain of states. Each state represents one instant of time, and at each time-step a transition is made to the next state in the chain. This restricted representation is key to capturing the phenomena underlying the empirical results.

Due to the added complexity of incorporating time into the latent cause model, I do not consider inference over an unknown model structure as was done in Chapters 3 and 4. Instead I will assume a fixed model structure, with two timelines emanating from a single holding state and each timeline consisting of six temporal states. The transitions out of this holding state are the only probabilistic and adaptive transitions in the model. These transition probabilities represent the frequency with which the timelines are experienced. Figure 5.1 illustrates the model structure used in all simulations.

As suggested in Figure 5.1, associated with each state is a stimulus observation. "Stimulus space" is an n-dimensional continuous space, where n is the number of distinct stimuli that can be observed (tone, light, shock, etc.). Each state has an expectation concerning the stimuli that should be observed when that state is occupied. This expectation is modeled by a probability density function, over this space, defined by a mixture of two multivariate Gaussians. The probability density at stimulus observation $\mathbf{y}(t)$ in state $x(t) = i$ at time t is,

$$p(\mathbf{y}(t) \mid x(t) = i) = (1 - \omega_i) \cdot \mathcal{N}(\mathbf{x}(t) \mid \mu_{i0}, \sigma_{i0}^2) + \omega_i \cdot \mathcal{N}(\mathbf{x}(t) \mid \mu_{i1}, \sigma_{i1}^2), \tag{5.2}$$

where ω_i is a mixture coefficient for the two Gaussians associated with state $x(t) = i$. The Gaussian means μ_{i0} and μ_{i1} and variances σ_{i0}^2 and σ_{i1}^2 are vectors of the same dimension

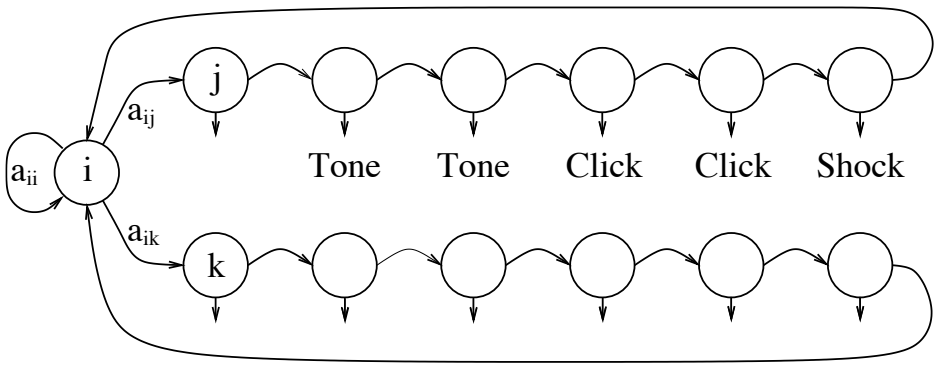


Figure 5.1. A depiction of the state and observation structure of the model (caution, this is not a Bayes net). Shown are two timelines, one headed by state **j** and the other headed by state **k**. State **i**, the holding state, transitions to states **j** and **k** with probabilities $\mathbf{a}_{ij}$ and $\mathbf{a}_{ik}$ respectively. Below the timeline representations are a sequence of observations: Tone, Click (click train) and Shock (footshock). The Tone and Click (click train) stimuli appear for two time steps each to simulate their presentation for an extended duration in the experiment.

as the stimulus vector $\mathbf{y}(t)$. Given knowledge of the state, the stimulus components are assumed to be mutually independent (covariance terms are zero). For each state, the first Gaussian pdf is non-adaptive, meaning μ_{i0} is fixed about a point in stimulus space representing the absence of stimuli. σ_{i0}^2 is fixed as well. For the second Gaussian, μ_{i1} and σ_{i1}^2 are adaptive. The observation model encodes a representation of the stimulus in a continuous perceptual stimulus space (adaptive Gaussian) while also explicitly encoding a representation of the absence of the stimulus.

5.4. Parameter inference

Fixing the model structure has important consequences for a normative account of these phenomena. Each of the experiments under consideration consists of two distinct sequences of observed stimuli (two timelines). In modeling these experiments, the goal is to demonstrate that the trained model integrates the timelines. That is, to demonstrate that both sequences of stimuli arise from a single Markov chain. It is a *conceptually* straightforward extension of the model of Chapter 3 to imagine how inference over the model structure might cause timelines to be integrated. As was seen with multiple simulations of experiments in Chapter 3 (e.g. second-order conditioning and early stages of configural conditioning), similar trial types become integrated through the latent causes because the quantity of data is insufficient to warrant more complex model that would separate trial types (with each emanating from a separate latent cause).

By fixing the model to the two-timeline structure (as shown in Figure 5.1), normative or optimal inference, in terms of data likelihood, would place the two distinct stimulus sequences in two separate Markov chains and would not demonstrate the desired integration of the timelines. As we shall see, the parameter learning procedure employed in this model succeeds in integrating the timelines; however it does so as a result of a suboptimal local minimum of the inference procedure. Thus the model presented in this chapter cannot provide a truly normative account of the merging of timelines.

A normative perspective on the model presented here is still possible. If one accepts that a true normative approach to this problem would involve inference over the model structure, then—as suggested by the results in Chapter 3—the timelines are still integrated just as the HMM model presented in this chapter implies (albeit through an entirely different inferential mechanism than that offered here). Consequently, in terms of an analysis of rational behavior, the contribution of the model presented here is to specify *how* the timelines integrate and how the relative timing of events affects animal behavior.

While I now provide these details of the implementation for completeness, they are not crucial to the exposition presented here. In principle, I could have used any of a number of approximate inference techniques (e.g. variational methods, particle filter methods, etc.) to estimate the parameters of the HMM model.

As already mentioned, in the case of a latent Markov chain of states, learning reduces to the problem of inference over the hidden state and parameters of the model. This implies the determination of the joint distribution over the parameters ω_i, μ_{i1} and σ_{i1}^2 for each state i, and a_{ij} for each state transition (out of the holding state), conditional on the sequence of observations.

Recursive Bayesian inference is one possible online learning scheme. It offers the appealing property of combining prior beliefs about the world with current observations through the recursive application of Bayes' theorem, $p(\lambda \mid \mathbf{Y}(t)) \propto p(\mathbf{y}(t) \mid \mathbf{Y}(t-1), \lambda) p(\lambda \mid \mathbf{Y}(t-1))$. The prior distribution, $p(\lambda \mid \mathbf{Y}(t-1))$ reflects the belief over the parameter λ before the observation at time t, $\mathbf{y}(t)$. $\mathbf{Y}(t-1)$ is the observation history up to time $t-1$, i.e. $\mathbf{Y}(t-1) = \{\mathbf{y}(t-1), \mathbf{y}(t-2), \dots\}$. The likelihood, $p(\mathbf{y}(t) \mid \mathbf{Y}(t-1), \lambda)$ is the probability density over $\mathbf{y}(t)$ as a function of the parameter λ.

Unfortunately, the implementation of exact recursive Bayesian inference for a hidden Markov model (HMM) is computationally intractable. This is a consequence of there being missing data in the form of hidden state. With hidden state, the posterior distribution over

the model parameters, after the observation, is given by

$$p(\lambda \mid \mathbf{Y}(t)) \propto \sum_{i=1}^{N} p(\mathbf{y}(t) \mid x(t) = i, \mathbf{Y}(t-1), \lambda) p(x(t) = i \mid \mathbf{Y}(t-1), \lambda) p(\lambda \mid \mathbf{Y}(t-1)), \quad (5.3)$$

where we have summed over the N hidden states. Computing the recursion for multiple time steps results in an exponentially growing number of terms contributing to the exact posterior.

Instead I use a recursive approximate inference scheme developed by Huo and Lee (1997), who employ a quasi-Bayes approach (Smith & Makov, 1978). The quasi-Bayes approach exploits the existence of a repeating distribution (natural conjugate) over the parameters for the *complete-data* HMM (where missing data such as the state sequence is taken to be known). Briefly, a distribution over the latent state is inferred via a forward-backward algorithm. Next, the inferred state values are used together with the observations to update the hyperparameters governing the prior distribution over the parameters. This results in an approximation to the posterior distribution over the HMM parameters within the conjugate family of the complete-data HMM. See Huo and Lee (1997) for a more detailed description of the algorithm.

Inferring the hidden state involves estimating transition probabilities between states, $\xi_{ij}^{(\tau)} = P(x(\tau) = i, x(\tau+1) = j \mid \mathbf{Y}(t), \lambda)$, and joint state and mixture component label probabilities $\zeta_{ik}^{(\tau)} = P(x(\tau) = i, l(\tau) = k \mid \mathbf{Y}(t), \lambda)$. Here $l(\tau) = k$ is the mixture component label indicating which Gaussian, $k \in \{0, 1\}$, is the source of the stimulus observation at time τ. λ is a vector of the current estimate of all model parameters.

I use an online version of the forward-backward algorithm (Krishnamurthy & Moore, 1993) to infer $\xi_{ij}^{(\tau)}$ and $\zeta_{i1}^{(\tau)}$. The forward pass computes the joint probability over state occupancy (taken to be both the state value and the mixture component label) at time τ and the sequence of observations up to time τ. The backward pass computes the probability of the observations in a memory buffer from time τ to the present time t given the state occupancy at time τ. The forward and backward passes over state/observation sequences are combined to give an estimate of the state occupancy at time τ given the observations up to the present time t. In the simulations reported here the memory buffer was 7 time steps long ($t - \tau = 6$).

The values of $\xi_{ij}^{(\tau)}$ and $\zeta_{i1}^{(\tau)}$, computed using the forward-backward algorithm, are used to update the hyperparameters. For the HMM, the prior is taken to be a product of Dirichlet probability density functions (pdfs) for the transition probabilities (a_{ij}), beta pdfs for the observation model mixture coefficients (ω_i) and normal-gamma pdfs for the Gaussian parameters (μ_{i1} and σ_{i1}^2). The basic hyperparameters are exponentially weighted

counts of events, with recency weighting determined by a forgetting parameter ρ. For example, κ_{ij} is the number of expected transitions observed from state i to state j, and is used to update the estimate of parameter a_{ij}. The hyperparameter ν_{ik} estimates the number of stimulus observations in state i credited to Gaussian k, and is used to update the mixture parameter ω_i. The remaining hyperparameters ψ, ϕ, and θ serve to define the pdfs over μ_{i1} and σ^2_{i1}. The variable d in the equations below indexes over stimulus dimensions. S_{i1d} is an estimate of the sample variance, and is a constant in the present model.

$$\kappa_{ij}^{(\tau)} = \rho(\kappa_{ij}^{(\tau-1)} - 1) + 1 + \xi_{ij}^{(\tau)} \tag{5.4}$$

$$\nu_{ik}^{(\tau)} = \rho(\nu_{ik}^{(\tau-1)} - 1) + 1 + \zeta_{ik}^{(\tau)} \tag{5.5}$$

$$\psi_{i1d}^{(\tau)} = \rho\psi_{i1d}^{(\tau-1)} + \zeta_{i1}^{(\tau)} \tag{5.6}$$

$$\phi_{i1d}^{(\tau)} = \rho(\phi_{i1d}^{(\tau-1)} - \frac{1}{2}) + \frac{1 + \zeta_{i1}^{(\tau)}}{2} \tag{5.7}$$

$$\theta_{i1d}^{(\tau)} = \rho\theta_{i1d}^{(\tau-1)} + \frac{\zeta_{i1}^{(\tau)} S_{i1d}}{2} + \frac{\rho\psi_{i1d}^{(\tau-1)}\zeta_{i1}^{(\tau)}}{2(\rho\psi_{i1d}^{(\tau-1)} + \zeta_{i1}^{(\tau)})}(x_d^{(\tau)} - \mu_{i1d}^{(\tau-1)})^2 \tag{5.8}$$

In the last step of the inference procedure, I update the estimate of the model parameters as the mode of their approximate posterior distribution. While this is an approximation to proper Bayesian inference on the parameter values, the mode of the approximate posterior is guaranteed to converge to a mode of the exact posterior. In the equations below, N is the number of states in the model.

$$a_{ij}^{(\tau)} = \frac{\kappa_{ij}^{(\tau)} - 1}{\sum_{n=1}^{N}(\kappa_{in}^{(\tau)} - 1)} \tag{5.9}$$

$$\omega_i^{(\tau)} = \frac{\nu_{i1}^{(\tau)} - 1}{\nu_{i0}^{(\tau)} + \nu_{i1}^{(\tau)} - 2} \tag{5.10}$$

$$\mu_{i1d}^{(\tau)} = \frac{\rho\psi_{i1d}^{(\tau-1)}\mu_{i1d}^{(\tau-1)} + \zeta_{i1}^{(\tau)} x_d^{(\tau)}}{\rho\psi_{i1d}^{(\tau-1)} + \zeta_{i1}^{(\tau)}} \tag{5.11}$$

$$(\sigma^2_{i1d})^{(\tau)} = \frac{2\rho\theta_{i1d}^{(\tau-1)} + \rho\psi_{i1d}^{(\tau-1)} \cdot (\mu_{i1d}^{(\tau)} - \mu_{i1d}^{(\tau-1)})^2 + \zeta_{i1}^{(\tau)} S_{i1d} + \zeta_{i1}^{(\tau)}(x_d^{(\tau)} - \mu_{i1d}^{(\tau)})^2}{\rho(2\phi_{i1d}^{(\tau-1)} - 1) + \zeta_{i1}^{(\tau)}} \tag{5.12}$$

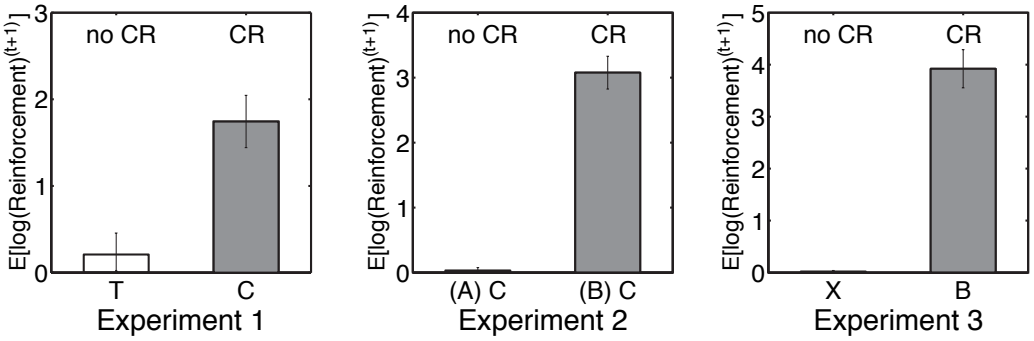


Figure 5.2. Results from 20 runs of the model simulation with each experimental paradigm. On the ordinate is the total reinforcement, on a log scale, above the baseline (an arbitrary perception threshold) expected to occur on the next time step. The error bars represent two standard deviations away from the mean.

The model contained two timelines (Markov chains). Let i denote the holding state and j, k the initial states of the two chains. The transition probabilities were initialized as $a_{ij} = a_{ik} = 0.025$ and $a_{ii} = 0.95$. Adaptive Gaussian means μ_{i1d} were initialized to small random values around a baseline of 10^{-4} for all states. The exponential forgetting factor was $\rho = 0.9975$, and both the sample variances S_{i1d} and the fixed variances σ^2_{i0d} were set to 0.2.

The goal is to show how the model integrates the timelines of the two training phases of each experiment. As previously mentioned, in the context of the HMM model, integrating timelines amounts to both phases of training becoming associated with a single chain of states. Figure 5.1 shows the integration of the two phases of Expt. 2B. During the second phase of the experiments, the second Markov chain (shown in Figure 5.1 starting with state **k**) offers an alternative to the chain associated with the first phase of learning. If we successfully integrate the timelines, the second chain is not used.[1]

5.5. Results

We trained the model on each of the experimental protocols of Table 5.1, using the same numbers of trials reported in the original papers. The model was run continuously through both phases of the experiments with a random intertrial interval. Figure 5.2 shows the simulation results from each of the three experiments. If we assume that the response varies monotonically with imminent reinforcement prediction, then in each case, the model's predicted responses agreed with the observations of Miller et al.

[1]In separate simulated experiments with two non-overlapping stimulus sequences, the two stimulus sequences were each associated with a separate chain of states.

The mechanism responsible for the account of the three experiments is state inference. When the second phase of each experiment is presented, the forward-backward algorithm determines the probability of each state. The states in the chain encoding the first experimental phase are always found to be the most likely. This is a direct result of there being stimuli in common between the two phases of training. Thus as expressed in the temporal coding hypothesis, the model integrates timelines with similar stimuli. During the test phase, the presentation of the test stimulus causes the inference procedure to place most of the probability mass on the state directly associated with that observation. If imminent reinforcement is predicted, by looking one time-step down the chain, then according to the model a response is evoked.

5.6. Discussion

The application of the latent cause framework to modeling the temporal structure of stimuli has drawn out a number of interesting features of the theory. First and foremost, latent cause theory is unique among the computational theories of conditioning because it is able to offer an unambiguous account of the fascinating set of experiments of Miller et al. exploring the nature of time in conditioning. In accounting for this data, the model provides a unique view of how animals represent and reason about the temporal order of events.

As demonstrated above, inference over the latent cause (state) is the mechanism responsible for stimulus pattern integration and hence the source of the model's account of the experiments of Miller et al. As was explored in the previous two chapters, inference over the latent state may also be seen as a mechanism for cue competition effects such as blocking and overshadowing—both phenomena accounted for by the TD model.

Where TD distinguishes itself from the present instantiation of the latent cause model is in accounting for conditioned inhibition phenomena, where a stimulus comes to predict the *absence* of reinforcement. Like the Rescorla-Wagner model, the TD model is capable of encoding negative associative strengths which represent a negative correlation with reinforcement and permit an explanation of conditioned inhibition. The latent cause model described in this chapter possesses only positive correlations between the stimuli and the latent cause. As a result, it cannot easily account for conditioned inhibition effects. However, as we saw in Chapter 3, by considering an alternative formulation of the theory—one that admits both negative correlations with the latent cause and inference over the model structure—the latent cause framework is able to offer an explanation of conditioned inhibition.

5.6.1. Contiguity versus Contingency Revisited

As discussed in Chapter 2, there is a question within the conditioning community regarding the essential ingredients for conditioning to occur. The currently dominant position, supported by such associative learning theories as the Rescorla-Wagner model, holds that mere temporal contiguity of stimulus and reinforcement is insufficient for the establishment of an association. Rather, for learning to occur the reinforcement must be interpreted as being contingent on the stimulus. The less fashionable opposing position claims that contiguity is entirely sufficient for the formation of an association.

Previously, the asymmetry between forward and backward conditioning as well as the perceived ineffectiveness of simultaneous conditioning (both are reviewed in Chapter 2 and briefly in the introduction to this chapter) was interpreted as significant evidence in favor of the contingent perspective. They suggest that the stimulus is only useful as a predictor if it precedes reinforcement. Otherwise the reinforcement event cannot be expressed as contingent on the stimulus event and no significant associative strength is acquired. This perspective is consistent with—and closely related to—the causal interpretations of the Rescorla-Wagner model. Our experience of the physical world dictates that causes must precede effects. If the stimulus is to be interpreted as the cause of the reinforcement then it must precede it. However the failure of the contingency based models to account for the experiments of Miller et al. demonstrates that this interpretation may be overly simplistic.

With their temporal coding hypothesis, Miller and Barnet (1993) offer a contiguity based perspective that not only accounts for the set of experiments considered above, but also offers an explanation of the asymmetry between forward and backward conditioning. They argue that simultaneous and backward conditioning results in weak responding, not because the corresponding associations are weak, but rather because they are not properly assessed. Typical classical conditioning paradigms measure anticipatory responses such as the blink of an eye when a puff of air is expected, or freezing in anticipation of a foot-shock. These types of behavioral indicators are appropriate measures of forward associations, but not of simultaneous or backward associations. Miller and Barnet suggest that temporal relationships among the stimuli and reinforcer are encoded within the associations acquired during learning. When testing with the stimulus, the temporal relationship is recalled and if the stimulus previously preceded the reinforcer, then an appropriate response would be made in anticipation of the imminent reinforcement. On the other hand, if the stimulus

previously followed the reinforcer, then the appearance of the stimulus would not signal an impending reinforcement and no response would be elicited.

Despite being declared a causal model and ultimately concerned with event contingencies, the latent cause model's commitment to learning the generative model of all stimuli, regardless of the nature of their reinforcement contingency, places it firmly in the sufficiency of contiguity camp. In fact, the temporally extended version of the latent cause model developed in this chapter essentially constitutes a computational implementation of the temporal coding hypothesis. Critically, the latent cause model offers a rigorous mechanism for timeline integration: inference over the latent state.

5.6.2. Learning and Performance

As described in the section 5.5, inference over the latent state determines whether a reinforcement is imminent during testing. State inference localizes the subject within the event sequence and, if the test stimulus typically occurred before reinforcement during training, then the presence of the test stimulus results in a high probability of impending reinforcement and evokes a response from the subject. Consider the case of Experiment 3. While testing with the backwards conditioned stimulus X no significant response was measured. Yet it is clear that there is an association involving X—that is, that something is learned about X.

This touches on another issue of some contention within the conditioning literature centering around the distinction between a true association and our measurement of that quantity as revealed by the performance of the subject at the time of testing. The operational definition of learning, offered in the introductory chapter, confounds the concepts of learning and performance. While learning often results in a change in behavior, it need not always be so. There are many examples within the conditioning literature where learning is latent, that is unobservable and unmeasurable, only to be revealed by subsequent experimental manipulations.

While the distinction between learning and performance was recognized by the early behaviorists,[2] more modern associative learning theories have largely neglected this distinction. Recently the controversy has centered around whether cue competition effects such as overshadowing and blocking are the result of a learning deficit—where little or no

[2]The distinction between learning and performance was recognized as early as Skinner (Skinner, 1938) where he separately defined a *reflex reserve* reflecting learning and a *reflex strength* reflecting performance potential. However of the early theorists, it was Tolman (1955) who was chiefly responsible for promoting the learning–performance distinction.

association is formed between the overshadowed (or blocked) stimulus (Holland, 1999)—or a performance deficit—where an association is formed but responding during testing is otherwise blocked (Arcediano et al., 2004). The latent cause model contributes to this debate by offering a rare computational account that favors the performance deficit explanation. The role of test performance on the predictions of the model is obviously crucial to the latent cause account of the present set of experiments.

5.6.3. Over-generalization

The question of how animals judge similarity between patterns of stimuli is currently an active area of research within the conditioning community and serves as an illustration of the difference between the model presented in this chapter, with a fixed model structure, and those proposed in Chapters 3 and 4, where the model structure itself was subject to inference. By considering formal Bayesian inference over the space of model structures, I showed that not only can the model structure be inferred from data, but the act of inference over models can bring about the kind of trial type integration witnessed in this chapter as the result of truly rational (or globally optimal) inference.

According to the model presented in this chapter, measurements of similarity are mediated by inference over the state of a fixed model structure. Consider a probe stimulus pattern possessing some number of stimuli in common with a stimulus pattern that was present repeatedly during training and presumably is encoded as an event timeline. During testing, the latent cause associated with the trained stimulus pattern is inferred to have given rise to the probe stimulus pattern with probability determined by the degree of stimulus overlap. If the stimulus patterns are sufficiently similar, then predictions of subsequent stimuli (in particular reinforcement) encountered during training are *generalized* to the probe stimulus pattern.

While this ability to generalize from previous experience to novel situations is an absolutely essential feature of any successful learning system, equally essential is the ability to discriminate between similar experiences when necessary. With training, animals are eventually capable of distinguishing between events with a significant amount of overlap in the pattern of stimuli. Unfortunately the present formulation of the latent cause model has a seriously impaired ability to discriminate between similar patterns of stimuli. For example, all the experiments considered in this chapter were variants of second-order conditioning effects. It is reasonably well documented that second order conditioning is a transient effect, disappearing as the number of second-order pairings increases (Yin et al.,

1994). The current model formulation completely misses this transition and simply continues to merge the trial types *ad nauseum*.

The failure to discriminate can, to a certain degree, be dismissed as a local maximum effect resulting from my choice of inference schemes. For in every experiment we consider here, the data likelihood would actually be maximized by splitting the two trial types between the two timelines that form the model structure. While the HMM model developed in this chapter successfully reproduces animal behavior on a set of surprising experiments, it does not provide a satisfactory answer to the question: why do animals integrate timelines? As discussed in Section 5.4, the model presented in Chapter 3 suggests a promising answer to this question.

CHAPTER 6

Concluding remarks

This thesis presented an exploration of the latent cause theory of classical conditioning. According to the theory, animals assume an unseen cause for observed events and attempt to recover the corresponding generative model that gives rise to these events. I have demonstrated how the latent cause theory may be applied to three distinct areas of classical conditioning. In each case it offers a novel account of empirical phenomena, many of which have otherwise resisted a computational explanation.

In Chapter 3 I explored the notion of generalization and discrimination in considerable detail. I advance the claim that animals negotiate between the competing requirement of discrimination and generalization based on the principle of the Bayesian (or automatic) Occam's razor. This proposal was made explicit with the introduction of a latent cause model where the generative structure is assumed unknown and subject to inference. The model was applied to demonstrate how model structure inference accounts for classic configural conditioning effects, second-order conditioning and acquired equivalence effects.

In Chapter 4 I introduce a model of change to the modeling framework advanced in Chapter 3. The model of change describes how the parameters and structure of the generative world model change over time. Its introduction allows the resulting non-stationary latent cause model to model trial order effects such as extinction—a paradigm the model of Chapter 3 is unable to accommodate. The model of change includes an unknown rate-of-change parameter. Through inference surprising events provide evidence that the rate of change has increased, and consequently that the model weights have similarly experienced significant change. The model negotiates between structural and parametric change with the same Bayesian inference mechanism that gave rise to the model complexity / data fidelity tradeoff in Chapter 3.

The non-stationary latent cause model offers a novel perspective on the factors that influence animal judgments about changes in the environment. In particular, the model

correctly predicts that the introduction of a novel neutral stimulus with the onset of acquisition training in a latent inhibition training procedure should alleviate the retarded acquisition characteristic of latent inhibition. The non-stationary latent cause model may be regarded as a generalization of the Pearce-Hall model—well known for its successful account of animal behavior in the presence of changing reinforcement contingencies. The non-stationary latent cause model is also used to account for a number of other change-related phenomena including unblocking and the partial reinforcement extinction effect.

In Chapter 5, I introduced a version of a latent cause model that explicitly encodes intratrial time as a sequence of causes (or states). This sequence of causes forms a *timeline* to which observed events such as stimuli and reinforcements are associated—in their temporal order. In this context the latent cause model is equivalent to a hidden Markov model with a single cause as the hidden state and the stimuli (plus reinforcement) as the observations. This model was used to demonstrate how the latent cause framework offers a computational implementation of the Temporal Coding Hypothesis (Miller & Barnet, 1993) and accounts for a number of interesting experiments (Barnet et al., 1991a; Cole et al., 1995; Cole & Miller, 1999). Collectively these experiments provide compelling evidence that animals encode the temporal relationship between stimuli and reinforcements and use this representation to predict impending reinforcement. The latent cause account of these phenomena is mediated by inference over the latent cause structures that merges similar trial-type timelines together.

While the model presented in Chapter 5 offers an account of the temporal coding experiments of Miller et al., the inference procedure over the hidden state is non-convex and thus subject to suboptimal learning. One consequence of this is that the model is unable to learn to discriminate similar trial-types. For example, the model would fail to discriminate between the three trial-types of negative patterning ($A+$, $B+$, and $AB-$), a discrimination routinely performed by animals. Instead of discriminating the three trial-types to three separate timelines, the overlapping stimuli would cause the model in Chapter 5 to merge the three together on a single timeline representing the occurrence of A, B and reinforcement. Because this solution is a local minimum of the approximate posterior distribution, the pseudo-Bayes inference scheme never reaches a representation where the three trial-types are encoded separately. As previously discussed, the model presented in Chapter 3 offers a suggestion of how these timelines might come to be discriminated.

6.1. Model Predictions

As an explicit computational modeling framework, the latent cause theory is capable of making explicit and verifiable predictions regarding the outcomes of experiments. In this section, I briefly describe two experiments and the corresponding latent cause theory predictions.

Consider an experiment with two interspersed trial types: $AB+$ (where A and B have been matched for saliency[1]) and a variable number of interspersed $A-$ trials. Following the example of the second-order conditioning account presented in Chapter 3, with a few interspersed presentations of $A-$, the latent cause model would give significant probability mass to models merging the two trial types ($AB+$ and $A-$) into a single latent cause. The latent cause would possess a strong association (weight) with A since it is presented in every trial, and weaker associations with B and reinforcement because these do not appear in the $A-$ trials. Thus on testing, this should translate to a moderate increase in responding to A relative to B.

This is a somewhat counterintuitive prediction: a few extinction trials with A actually augments responding to it. According to the model this effect would only be temporary and eventually—with increasing numbers of $A-$ trials, it will be inferred that there are two latent causes, an $AB+$ cause and an $A-$ cause (alternatively: a $B+$ cause and an $A-$ cause). According to the model, testing at this point would reveal that responding to A and B has reversed with B eliciting a larger response than A. To the best of my knowledge, this experiment has not been performed and so this prediction would constitute a fair test of the latent cause interpretation of classical conditioning.

The second prediction is derived from the non-stationary version of the latent variable model developed in Chapter 4. According to the model of change, the addition of a novel stimulus constitutes strong evidence in favor of change. In Chapter 4, we saw how this successfully explained a decrease in the latent inhibition effect when a novel stimulus was presented together with the target stimulus during second stage acquisition training. According to the latent cause model, the same effect should be apparent in extinction. That is, training with $A+$ followed by $AB-$ should generate faster extinction to A relative to controls that experience $A+$ followed by $A-$. In contrast, the Rescorla-Wagner model would predict that the addition of a novel stimulus B with the onset of extinction training would result in reduction in the rate of extinction, as B would claim some responsibility for the

[1] With similar saliency there is no strong asymmetric overshadowing effect arising from $AB+$ training and a moderate amount of overshadowing is observed when testing with both A and B (Mackintosh, 1975)

absence of expected reinforcement. According to this perspective B would become a conditioned inhibitor and A would be relatively protected from extinction.

While there seems to be no empirical evidence that bears directly on this effect, there are some closely related experiments that offer some insight. The experiment appears similar to the paradigm responsible for the protection from extinction effect (Rescorla, 2003b). Here a trained excitor is extinguished in the presence of an established conditioned inhibitor. The presence of the conditioned inhibitor acts to preserve the excitor through extinction rather than promote faster extinction. While this result would seem to contradict the latent cause model prediction, it is known that the paradigm suggested above to provide accelerated extinction ($A+$ followed by $AB-$) does not render B a conditioned inhibitor (Yin et al., 1994) (countering the prediction of the Rescorla-Wagner model).[2] Thus it seems that the protection from extinction effect may not apply in this case.

Indirect evidence in favor of the latent cause prediction is provided by an experiment demonstrating that extinction may be facilitated by an increase or a decease in the trial duration during delay conditioning (Haselgrove & Pearce, 2003). In particular, facilitated extinction was observed when a 10 second stimulus was lengthened during extinction training and similarly when a 60 second stimulus was shortened during extinction training[3]. According to the non-stationary latent cause perspective[4], change in trial duration would be interpreted as evidence in favor of change within the environment which in turn results in a facilitation of the extinction effect as the parameters adapt at a greater rate (see Chapter 4 for details). I suggest that the addition of a novel stimulus B would have an equivalent effect.

6.2. Implications for associative learning theories

Throughout this thesis, numerous experiments were described where the latent cause account involves the integration of various trial-types into association with a single latent variable, sometimes merging stimuli that were never presented together onto a single latent cause. These stimuli, once merged, are subject to the same effects of facilitation (such as second order conditioning) and competition (such as blocking and overshadowing)—mediated through the latent cause inference mechanism—as stimuli that were originally paired together. This constitutes a rather distinct property of the latent variable model that

[2] Rescorla's (2003b) experiment used a previously established conditioned inhibitor.

[3] Rate estimation theory (Gallistel & Gibbon, 2000), with its famous quantitative prediction regarding the number of trials required for extinction, has considerable trouble explaining why both shortening and lengthening the stimulus duration would facilitate extinction (Haselgrove & Pearce, 2003)

[4] It is important to note that the model presented in Chapter 4 has no capacity to represent trial length and so could not be directly applied to this experiment

it shares with no other computational model of conditioning with the possible exception of the Gluck and Myers (1993) model of hippocampus and similar network based models (O'Reilly & Rudy, 2001). This feature of the model has important implications for associative learning theory, which holds as one of its most central tenets that stimulus competition effects arise as a result of stimuli being presented together.

A recent line of empirical inquiry seriously questions the traditional associative learning perspective with regard to stimulus competition effects (Miller & Escobar, 2002). Miller and Escobar draw on a series of experiments to illustrate that stimulus competition effects are arrived at through far more general means than those conceived in traditional associative learning theory. Most associative learning theories account for competition between cues (preceding stimuli) presented together. These are the basic phenomena of blocking (Kamin, 1967), overshadowing (Kamin, 1969) and the relative validity effect (Wagner et al., 1968) (as described in Chapter 2). However there also appears to be significant evidence of competition between outcomes (subsequent stimuli or reinforcers) trained together (Rescorla, 1980; Esmoriz-Arranz, Miller, & Matute, 1997; Miller & Matute, 1998), where for example, the training sequence: $A \rightarrow O_1$ followed by $A \rightarrow O_1O_2$ results in blocking to the association between cue A and outcome O_2 (Esmoriz-Arranz et al., 1997). Effects such as these are difficult to reconcile with the traditional associative learning perspective, yet the latent cause framework could, in principle, readily accommodate these kinds of effects since there is no distinct role played by the cue or the outcome. However, it is important to note that these effects are typically demonstrated within a sensory preconditioning paradigm as it has been observed that the biological significance of the outcome has an important effect on the establishment of outcome competition effects (Miller & Matute, 1998). The latent cause model, as described in the thesis, is unable to account for the role of biological significance.

From the point of view of latent cause theory, the most interesting claim by Miller and Escobar is that competition effects are evident between cues and outcomes that are trained *apart* (Miller & Escobar, 2002). For example, following the training sequence $A \rightarrow O_1$ and $B \rightarrow O_1$, testing with A reveals an interference effect when compared to control subjects that do not experience $B \rightarrow O_1$ training (Escobar et al., 2001). More traditional associative learning theories lack a mechanism to account for interference effects between stimuli that are never paired together. However, according to the latent cause model such an effect is parsimoniously explained as a result of merging trial-types. In the above example the two trial-types are merged based on their both possessing the outcome O_1. The latent cause model offers one of the few parsimonious accounts for all four of these interference

phenomena: 1) cues trained together, 2) outcomes trained together, 3) cues trained apart, and 4) outcomes trained apart.

6.3. Implications of the causal interpretation of the model

As discussed in Chapters 1 and 2, the latent cause model is based on the assumption that animals interpret events in the context of a classical conditioning experiment as all arising independently from an unseen cause. While this frequently appears to be a valid interpretation of events, it is not always so. As discussed in Chapter 2 and again in Chapter 3, there are some instances where the more traditional Rescorla-Wagner causal assumption of the cue *causing* the reinforcement is the most reasonable interpretation of events. Perhaps the most compelling example of a violation of the latent cause assumption arises when food and nausea are paired. It is entirely appropriate for an animal to interpret a direct causal association between food and subsequent illness—even at a considerable temporal distance. As discussed in Chapter 3, when the latent causal assumption is invalid, the predictions of the model diverge from animal behavior. Extending the model to incorporate multiple possible causal structures is a nontrivial, but potentially important, avenue for future work. The extension could be interpreted as an instance of theory-based causal inference (Tenenbaum & Griffiths, 2003) where each causal structure—latent cause or direct causal link between stimuli—constitutes a hypothesis of the causal relationship among stimuli. As is clear from the empirical evidence (Garcia & Koelling, 1966), certain of these causal hypotheses will be *a priori* more likely in particular stimulus contexts. This would not present a fundamental problem to the theory-based causal inference framework since it admits priors over structure—though it remains to fully specify the relationship between stimulus-reward pairs and the corresponding prior distribution over causal structures.

Even without completely specifying the theory-based causal inference extension, the extension offers an explanation of some interesting experimental results. For example, it has recently been reported (Batson & Batsell, 2000) that facilitation rather than blocking arises from a sequential training schedule of $odor \rightarrow nausea$ followed by $odor \cdot food \rightarrow nausea$. According to both traditional associative theories and the latent cause model, the specific identity of the stimuli would not alter the prediction of blocking of the $food$ stimulus. However, the extended model could be imagined to first encode a latent cause association between $odor$ and $nausea$ and then with $odor \cdot food \rightarrow nausea$ training, the prior preference for a direct connection between $food$ and $nausea$ could lead to a model structure shown in Figure 6.1 with two paths from $food$ to $nausea$, one direct and one through

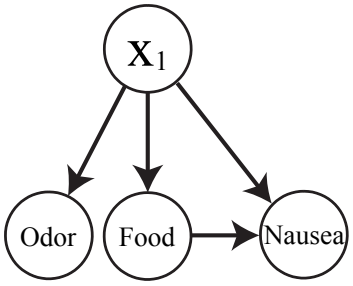


Figure 6.1. A possible model structure that would account for the observed facilitation of responding to *food* due to pre-training with $odor \rightarrow nausea$ before training with $odor \cdot food \rightarrow nausea$. This procedure would typically be expected to produce blocking to *food*, but because of the double path between *food* and $nausea$ the proposed theory-based causal inference model would predict facilitation.

the latent cause, mediated by their common association with $odor$. The two connections would presumably sum (in a manner consistent with sigmoid belief network properties) leading to a increased predicted probability of $nausea$ compared to control subjects that do not experience $odor \rightarrow nausea$ training.

6.4. Exploiting the rational analysis approach

In this thesis, I have held firm to the rational analysis perspective that animals behave in accord with the normative standard of Bayesian inference over the latent cause model structure. In each of Chapters 3, 4 and 5, I proposed a distinct version of the latent cause theory. In each case, the modeling framework was adapted to accommodate a set of data or to explore an aspect of the theory. Each chapter contains a different scheme to approximate inference over the model parameters and, in the case of Chapters 3 and 4, the model structure itself. In Chapter 3, I used reversible-jump Markov chain Monte Carlo to generate samples from the posterior. In Chapter 4, a particle filter was employed to sample over trajectories of model parameters and structures. Finally in Chapter 5, a quasi-Bayes deterministic approximate inference scheme was used. The combination of choosing a generative modeling framework based on specific research goals and then choosing an appropriate inference scheme based on convenience has been exploited to provide a wide ranging and diverse exploration of the latent cause theory and its implications with regard to animal behavior in classical conditioning.

If I had instead focused on a specific learning rule—as is the standard approach to modeling classical conditioning—it would have been much more difficult to adapt that rule to new modeling paradigms where, for instance, we may be interested in encoding the temporal order of stimuli. Adhering to the rational approach of Bayesian inference

defined the *correct* learning rule for a given modeling paradigm, be it encoding the temporal order of stimuli or assuming an unknown model structure. Even if the exact *correct* learning rule is not analytically tractable—as was the case in every model I considered—the rational approach admits the exploration of the considerable wealth of approximate inference schemes that exist within the machine learning and statistical literatures.

The rational approach also provides significant potential to explore more sophisticated models than would otherwise be practical. For instance, without the guidance of Bayesian inference it would have been a formidable task to attempt to specify mutually consistent learning rules over structure and parameters. However, within the Bayesian framework the equivalent learning rules are fully specified by the combination of the joint prior distribution and the basic calculus of probability theory. I suggest that one of the main contributions of this thesis is to demonstrate the power of the rational approach in developing models of animal behavior.

6.5. Extensions to the latent cause model

In each of Chapters 3, 4 and 5, I have shown how the latent cause framework can offer insight into distinct and rarely intersecting avenues within the classical conditioning literature. However, it remains to truly provide an integrated model of all the features presented in these chapters. Such a model would simultaneously encode the temporal order of stimuli within a trial, permit inference over the model structure and incorporate a model of change over model structure and parameters.

There are a number of serious issues that must be confronted in pursuing such a model. Perhaps most importantly, an appropriate approximate inference scheme must be determined. The particle filter employed in Chapter 4 struggled with the more complex experiments considered there, and would very likely be inadequate for models that also encode a representation of the temporal sequence of stimuli within a trial. It is possible that an analytic approximation over some of the real valued parameters would ameliorate the situation[5], though even if this were effective, the model structure space alone is sufficiently large to present serious challenges to existing particle filter algorithms.

A further issue of some concern in pursuing the integrated latent cause model is the relatively few experiments that exercise a significant portion of the model. Conditioning experiments tend to cluster around theoretically interesting questions raised by competing theories. Presently there is little guidance as to what are theoretically interesting experiments that incorporate all aspects of the model and consequently very few experiments

[5] Analytic approximations proved ineffective in improving inference in the model presented in Chapter 4.

exist that would even warrant such an integrated model. However there are a few exceptions, with one of the more interesting experiments being an exploration of facilitation and interference effects in serial compound training (Schreurs, Kehoe, & Gormezano, 1993).

In this experiment, one group of subjects (Group Compound) is exposed to serial compound training $A \rightarrow B \rightarrow C \rightarrow D+$ where each arrow indicates a 0.4 second time delay between each stimulus (also of 0.4 seconds in duration). A second group of subjects (Group Element) receives interleaved presentations of each of A, B, C and D separately paired with reinforcement in such a way that the individual temporal relationships are preserved between the two groups. That is, Group Element receives A followed 2.4 seconds later by reinforcement and separately B followed 1.6 seconds later by reinforcement and so on. Early in training, relative to Group Element, Group Compound reveals significant facilitation of response when tested with A. However with further training, the relative amounts of responding to A between the two groups reverses with Group Compound showing evidence of overshadowing of A as Group Element continues to slowly increase responding.

The occurrence of facilitation followed by overshadowing brought on by training with intervening stimuli would be predicted by an integrated model. Early in training, the intervening stimuli provide evidence in support of a temporally extended model structure (a more complex model structure that is not preferred *a priori*). This would have the effect of facilitating an encoding of an association between A and reinforcement. However with additional training, the weights between the intervening stimuli and the latent timeline of states (causes) would increase to a point where, in testing A in isolation, the absence of the intervening stimuli constitutes considerable evidence against the latent cause being in a state predictive of reinforcement. This is the temporally extended version of the latent cause model's mechanism for overshadowing provided in Chapter 3.

6.6. Implications for computational theories of brain function

The theory advanced in this thesis is very abstract and as such does not offer much in the way of guidance for neurophysiological mechanisms underlying learning in classical conditioning. However there are a few basic computational principles that appear to have correlates in more physiologically motivated models of learning. As previously mentioned in Chapter 3, there are many similarities between the latent cause theory and a recent cortical-hippocampal model (Gluck & Myers, 1993). The Gluck and Myers connectionist model represents the hippocampus as an autoencoder with a relatively small number of units in the middle layer encoding configurations of stimuli in a manor analogous to the causes within latent cause theory. Comparing these two models leads one to

the possibility that perhaps the latent cause structure may be mapped to the hippocampus; however, such a hypothesis does not appear to survive a more penetrating analysis. Animals with lesioned hippocampi have been shown to be readily capable of many learning tasks including traditional configural conditioning tasks (Gallagher & Holland, 1992; Whishaw & Tomie, 1991). Yet, without the latent cause structure, the latent cause model has no mechanism for associating stimuli and hence no mechanism for learning at all. Thus, it seems it is overly simplistic to attribute the latent cause model function exclusively to the hippocampus.

The cortical-hippocampal model of O'Reilly and Rudy suggests a somewhat different division of labor between the cortex and the hippocampus (O'Reilly & Rudy, 2001) that offers an interesting point of contact with the latent cause model. Following (McClelland et al., 1995), O'Reilly and Rudy suggest that there is a fundamental tradeoff in learning between the competing requirements of "extracting and representing the *general* properties of the environment" and learning and remembering important *specific* details of the environment. They argue that these two opposing requirements justify two separate learning systems: the cortex and the hippocampus, with the cortex specialized to learn generalities of the environment and the hippocampus specialized to learn quickly and represent specific details of the environment. The tradeoff that O'Reilly and Rudy describe is remarkably similar to the fundamental tradeoff between discrimination and generalization advanced in Chapter 3. One possible interpretation of the relationship between these models is to consider the Bayesian Occam's razor effect of balancing the requirements between generalization and discrimination being embodied in the competition between the hippocampal and cortical influence on the output units. Such an interpretation leads to the hypothesis that lesioning the hippocampus would bias the tradeoff toward simpler models and retard acquisition challenging discriminations such as negative patterning. Unfortunately once again, such a prediction does not appear to be supported by the available empirical evidence (Gallagher & Holland, 1992; Whishaw & Tomie, 1991).

6.7. Implications for inference mechanisms in the brain

Ultimately, organisms of limited computational means cannot be capable of the kinds of computations I impute to them. They must resort to solving these difficult inference problems approximately. This line of reasoning naturally leads one to wonder: what sort of approximate inference schemes might exist in the brains of animals?

An interesting answer to this question is provided by a recent line of research interpreting the computations performed by the visual cortex as an instance of hierarchical

Bayesian inference (Lee & Mumford, 2003). The authors argue that the tightly coupled feedback loops within the visual cortex may be computing something akin to loopy belief propagation or particle filtering or some combination of the two. The readily distributed nature of these two algorithms make them attractive candidates for implementation in the brain.

A recent investigation by Gallistel et al. (2004) is more relevant to the question of what kind of approximate inference scheme supports learning in the classical conditioning context. They demonstrate that the stereotypical smooth and gradual learning curve that has become a classic symbol of conditioning is a fiction—a result of averaging together the learning curves of individual subjects. Gallistel et al. (2004) shows that the learning curves of individual subjects often show abrupt step-like increases from untrained levels of responding to near asymptotic responding. The onset of the abrupt transition is widely distributed across subjects. Interestingly, this kind of abrupt acquisition behavior is reminiscent of the fast acquisition apparent in the particle filter inference algorithm employed in Chapter 4. In fact, as the number of particles used in the simulation decreases, the abrupt acquisition largely remains, but across independent simulations its onset becomes more widely distributed. This might suggest that animal brains implement a kind of particle filter algorithm with a limited number of hypotheses (or particles). Such a hypothesis is consistent with the proposal put forth by Lee and Mumford (2003).

Bibliography

A. G. Baker, Robin A. Murphy, F. V.-T., & Mehta, R. (2001). Contingency learning and causal reasoning. In Mowrer, R. R., & Klein, S. B. (Eds.), *Handbook of Contemporary Learning Theories*, chap. 7, pp. 255–306. Lawrence Erlbaum Associates, Mahwah, New Jersey.

Anderson, J. R. (1990). *The adaptive character of thought*. Erlbaum, Hillsdale, NJ.

Arcediano, F., Escobar, M., & Miller, R. R. (2004). Is stimulus competition an acquisition deficit or a performance deficit?. *Psychonomic Bulletin & Review*, *11*(6), 1105–1110.

Aydin, A., & Pearce, J. M. (1997). Some determinants of response summation. *Animal Learning and Behavior*, *25*(1), 108–121.

Balleine, B. W., & Dickinson, A. (2000). The effect of lesions of the insular cortex on instrumental conditioning: Evidence for a role in incentive memory. *Journal of Neuroscience*, *20*, 8954–8964.

Bao, S., Chan, V. T., & Merzenich, M. M. (2001). Cortical remodelling induced by activity of ventral tegmental dopamine neurons. *Nature*, *412*, 79–83.

Barnet, R. C., Arnold, H. M., & Miller, R. R. (1991a). Simultaneous conditioning demonstrated in second-order conditioning: Evidence for similar associative structure in forward and simultaneous conditioning. *Learning and Motivation*, *22*, 253–268.

Barnet, R. C., Grahame, N. J., & Miller, R. R. (1991b). Comparing the magnitudes of second-order conditioning and sensory preconditioning effects. *Bulletin of the Psychonomic Society*, *29*(2), 133–135.

Barnet, R. C., Grahame, N. J., & Miller, R. R. (1993). Temporal encoding as a determinant of blocking. *Journal of Experimental Psychology: Animal Behavior Processes*, *19*(4), 327–341.

Barnet, R. C., & Miller, R. R. (1996). Temporal encoding as a determinant of inhibitory control. *Learning and Motivation*, *27*, 73–91.

Batson, J. D., & Batsell, W. R. (2000). Augmentation, not blocking, in an A+/AX+ flavor-conditioning procedure. *Psychonomic Bulletin & Review*, *7*(3), 466–471.

Bhat, U. N., & Miller, G. K. (2002). *Elements of Applied Stochastic Processes* (3rd edition).

Wiley, Hoboken, NJ.

Bibby, B. M., Skovgaard, I. M., & Sorensen, M. (2005). Diffusion-type models with given marginal distribution and autocorrelation function. *Bernoulli*, *11*(2), 191–220.

Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Oxford University Press, Oxford, England.

Blaisdell, A. P., Bristol, A. S., Gunther, L. M., & Miller, R. R. (1998). Overshadowing and latent inhibition counteract each other: Support for the comparator hypothesis. *Journal of Experimental Psychology: Animal Behavior Processes*, *24*(3), 335–351.

Blaisdell, A. P., Denniston, J. C., & Miller, R. R. (1997). Unblocking with qualitative change of unconditioned stimulus. *Learning and Motivation*, *28*, 268–279.

Blaisdell, A. P., Gunther, L. M., & Miller, R. R. (1999). Recovery from blocking achieved by extinguishing the blocking CS. *Animal Learning & Behavior*, *27*, 63–76.

Bouton, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of pavlovian conditioning. *Psychological Bulletin*, *114*, 80–99.

Brandon, S. E., Vogel, E. H., & Wagner, A. R. (2000). A componential view of configural cues in generalization and discrimination in Pavlovian conditioning. *Behavioural Brain Research*, *110*, 67–72.

Brogden, W. J. (1939). Sensory preconditioning. *Journal of Experimental Psychology*, *25*, 323–332.

Brown, P. L., & Jenkins, H. M. (1968). Auto-shaping of the pigeon's key peck. *Journal of Experimental Analysis of Behavior*, *11*, 1–8.

Cheatle, M. D., & Rudy, J. W. (1978). Analysis of second-order odor-aversion conditioning in neonatal rats: Implications for Kamin's blocking effect. *Journal of Experimental Psychology: Animal Behavior Processes*, *4*(3), 237–249.

Cole, R. P., Barnet, R. C., & Miller, R. R. (1995). Temporal encoding in trace conditioning. *Animal Learning and Behavior*, *23*(2), 144–153.

Cole, R. P., & Miller, R. R. (1999). Conditioned excitation and conditioned inhibition acquired through backward conditioning. *Learning and Motivation*, *30*, 129–156.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and striatal systems for behavioural control. *Nature Neuroscience*, *8*(12), 1704–1711.

Dayan, P., & Long, T. (1998). Statistical models of conditioning. In *Advances in Neural Information Processing Systems 10*, pp. 117–123.

Dayan, P. (1993). Improving generalization for temporal difference learning: the successor representation. *Neural Computation*, *5*, 613–624.

Dayan, P., & Kakade, S. (2001). Explaining away in weight space. In Todd K. Leen, T. G. D.,

& Tresp, V. (Eds.), *Advances in Neural Information Processing Systems 13*, Vol. 13, pp. 451–457.

Dayan, P., Kakade, S., & Montague, P. R. (2000). Learning and selective attention. *Nature Neuroscience*, *3*, 1218–1223.

Dayan, P., & Yu, A. (2003). Uncertainty and learning. *IETE Journal of Research*, *49*, 171–182.

Delamater, A. R., Sosa, W., & Katz, M. (1999). Elemental and configural processes in patterning discrimination learning. *Quarterly Journal of Experimental Psychology*, *52B*(2), 97–124.

Dexter, W. R., & Merrill, H. K. (1969). Role of contextual discrimination in fear conditioning. *Journal of Comparative and Physiological Psychology*, *69*, 677–681.

Dickinson, A., Hall, G., & Mackintosh, N. J. (1976). Surprise and the attenuation of blocking. *Journal of Experimental Psychology: Animal Behavior Processes*, *2*, 213–222.

Escobar, M., Matute, H., & Miller, R. R. (2001). Cues trained apart compete for behavioral control in rats: Convergence with the associative interference literature. *Journal of Experimental Psychology: General*, *130*(1), 97–115.

Esmoriz-Arranz, F. J., Miller, R. R., & Matute, H. (1997). Blocking of subsequent and antecedent events. *Journal of Experimental Psychology: Animal Behavior Processes*, *23*, 145–156.

Fanselow, M. (1990). Factors governing one-trial contextual conditioning. *Animal Learning & Behavior*, *18*, 264–270.

Fitzgerald, R. D. (1963). Effects of partial reinforcement with acid on the classically conditioned salivary response in dogs. *Journal of Comparative and Physiological Psychology*, *56*, 1056–1060.

Friston, K. J., Tononi, G., Reeke, G. N. J., Sporns, O., & Edelman, G. M. (1994). Value-dependent selection in the brain: Simulation in a synthetic neural model. *Neuroscience*, *59*, 229–243.

Gallagher, M., & Holland, P. C. (1992). Preserved configural conditioning and spatial learning impairments in rats with hippocampal damage. *Hippocampus*, *2*, 81–88.

Gallistel, C. R. (1990). *The organization of learning*. MIT Press, Cambridge, MA.

Gallistel, C. R., Fairhurst, S., & Balsam, P. (2004). The learning curve: Implications of a quantitative analysis. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(36), 13124–13131.

Gallistel, C. R., & Gibbon, J. (2000). Time, rate and conditioning. *Psychological Review*, *107*, 289–344.

Gallistel, C. R., Mark, T. A., King, A. P., & Latham, P. E. (2001). The rat approximates an

ideal detector of changes in rates of reward: Implications of the law of effect. *Journal of Experimental Psychology: Animal Behavior Processes*, *27*(4), 354–372.

Ganesan, R., & Pearce, J. M. (1988). Effects of changing the unconditioned stimulus on appetitive blocking. *Journal of Experimental Psychology: Animal Behavior Processes*, *14*(280–291).

Garcia, J., & Koelling, R. (1966). Relation of cue to consequence in avoidance learning. *Psychonomic Science*, *4*, 123–124.

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, *84*, 279–325.

Gibbon, J., Farrell, L., Locurto, C. M., Duncan, H. J., & Terrace, H. S. (1980). Partial reinforcement in autoshaping with pigeons. *Animal Learning & Behavior*, *8*, 45–59.

Gibson, E. J. (1959). A re-examination of stimulus generalization.. *Psychological Review*, *66*, 340–342.

Gluck, M., Glauthier, P., & Sutton, R. S. (1992). Adaptation of cue-specific learning rates in network models of human category learning. In *Proceedings of the fourteenth annual conference of the cognitive science society*, pp. 540–545. Erlbaum.

Gluck, M. A., & Myers, C. (1993). Hippocampal mediation of stimulus representation: A computational theory. *Hippocampus*, *3*(4), 491–516.

Glymour, C. (2001). *The Mind's Arrows: Bayes Nets and Graphical Causal Models in Psychology*. MIT Press, Cambridge, MA.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, *82*, 711–732.

Grice, G. R. (1965). Investigations of response-mediated generalization.. In Mostofsky, D. I. (Ed.), *Stimulus generalization*, pp. 373–382. Stanford University Press, Stanford, CA.

Guthrie, E. R. (1935). *The psychology of learning*. Harper, New York.

Guthrie, E. R. (1959). Association by contiguity. In Koch, S. (Ed.), *Psychology: A study of a science*, Vol. 2. McGraw-Hill, New York.

Hall, G., & Pearce, J. M. (1979). Latent inhibition of a CS during CS-US pairings. *Journal of Experimental Psychology: Animal Behavior Processes*, *5*, 162–177.

Hall, G., & Pearce, J. M. (1982). Restoring the associability of a preexposed CS by a surprising event. *Quarterly Journal of Experimental Psychology*, *34B*, 127–140.

Haselgrove, M., Aydin, A., & Pearce, J. M. (2004). A partial reinforcement extinction effect despite equal rates of reinforcement during Pavlovian conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, *30*(3), 240–250.

Haselgrove, M., & Pearce, J. M. (2003). Facilitation of extinction by an increase or a decrease

in trial duration. *Journal of Experimental Psychology: Animal Behavior Processes*, *29*(2), 153–166.

Hilgard, E. R., & Marquis, D. G. (1961). *Conditioning and Learning* (2nd edition). Appleton-Century-Crofts, New York.

Holland, P. C. (1988). Extinction and inhibition in unblocking. *Journal of Experimental Psychology: Animal Behavior Processes*, *14*, 261–279.

Holland, P. C. (1997). Brain mechanism for changes in processing of conditioned stimuli in Pavlovian conditioning: Implication for behavior theory. *Animal Learning & Behavior*, *25*, 373–399.

Holland, P. C. (1998). Amount of training affects associatively-activated event representation. *Neuropharmacology*, *37*, 461–469.

Holland, P. C. (1999). Overshadowing and blocking as acquisition deficits: No recovery after extinction of overshadowing or blocking cues.. *Quarterly Journal of Experimental Psychology: Comparative and Physiological Psychology*, *52B*(4), 307–333.

Honey, R. C., & Hall, G. (1989). Acquired equivalence and distinctiveness of cues. *Journal of Experimental Psychology: Animal Behavior Processes*, *15*(4), 338–346.

Honey, R. C., & Watt, A. (1998). Acquired relational equivalence: Implications for the nature of associative structures. *Journal of Experimental Psychology: Animal Behavior Processes*, *24*(3), 325–334.

Honey, R. C., & Watt, A. (1999). Acquired relational equivalence between contexts and features. *Journal of Experimental Psychology: Animal Behavior Processes*, *25*(3), 324–333.

Houk, J. C., Adams, J. L., & Barto, A. G. (1995). A mode of how the basal ganglia generate and use neural signals that predict reinforcement. In Houk, J. C., David, J. L., & Beiser, D. G. (Eds.), *Models of Information Processing in the Basal Ganglia*, pp. 249–270. MIT Press.

Hull, C. L. (1943). *Principles of behavior*. Appleton-Century-Crofts, New York.

Huo, Q., & Lee, C.-H. (1997). On-line adaptive learning of the continuous density hidden Markov model based on approximate recursive Bayes estimate. *IEEE Transactions on Speech and Audio Processing*, *5*(2), 161–172.

Iba, Y. (2001). Extended ensemble Monte Carlo. *International Journal of Modern Physics C*, *12*(5), 623–656.

James, W. (1890). *Principles of psychology*. Holt, New York.

Justin A. Harris, Megan L. Jones, G. K. B., & Westbrook, R. F. (2000). Contextual control over conditioned responding in an extinction paradigm. *Journal of Experimental Psychology: Animal Behavior Processes*, *26*(2), 174–185.

Kakade, S., & Dayan, P. (2000). Acquisition in autoshaping. In Solla, S. A., Leen, T. K., & Müller, K.-R. (Eds.), *Advances in Neural Information Processing Systems*, pp. 24–30. MIT Press.

Kakade, S., & Dayan, P. (2002). Acquisition and extinction in autoshaping. *Psychological Review*, *109*, 533–544.

Kamin, L. J. (1967). Attention-like processes in classical conditioning. In *Miami Symposium on the Prediction of Behavior: Aversive Stimulation*, pp. 9–31. University of Miami Press, Miami.

Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In Campbell, B. A., & Church, R. M. (Eds.), *Punishment and aversive behavior*, pp. 276–296. Appleton-Century-Crofts, New York.

Kehoe, E. J. (1986). Summation and configuration in conditioning of the rabbit's nictitating membrane response. *Journal of Experimental Psychology: Animal Behavior Processes*, *12*, 186–195.

Kehoe, E. J., Graham-Clarke, P., & Schreurs, B. G. (1989). Temporal patterns of the rabbit's nictitating membrane response to compound and component stimuli under mixed CS-US intervals. *Behavioral Neuroscience*, *103*, 283–295.

Kehoe, E. J., Schreurs, B. G., & Graham, P. (1987). Temporal primacy overrides prior training in serial compound conditioning of the rabbit's nictitating membrane response.. *Animal Learning and Behavior*, *15*, 455–464.

Krishnamurthy, V., & Moore, J. B. (1993). On-line estimation of hidden Markov model parameters based on the Kullback-Leibler information measure. *IEEE Transactions on Signal Processing*, *41*(8), 2557–2573.

Lantz, A. E. (1973). Effects of number of trials, interstimulus interval and dishabituation during CS habituation on subsequent conditioning in a CER paradigm. *Animal Learning & Behavior*, *1*, 273–277.

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of Optical Society of America, A*, *20*(7), 1434–1448.

Lubow, R. E., & Moore, A. U. (1959). Latent inhibition: the effect of nonreinforced preexposure to the conditional stimulus. *Journal of Comparative and Physiological Psychology*, *52*, 415–419.

Lubow, R. E., Rifkin, B., & Alek, M. (1976). The context effect: The relationship between stimulus preexposure and environmental preexposure determines subsequent learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *2*, 38–47.

MacKay, D. J. C. (1991). Bayesian model comparison and backprop nets. In *Advances in*

Neural Information Processing Systems 4, Cambridge, MA. MIT Press.

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

Mackintosh, N. J. (1983). *Conditioning and Associative Learning*. Oxford University Press, New York.

Marr, D. (1971). Simple memory: A theory for archicortex. *Philosophical Transactions of the Royal Society B*, *262*, 23–81.

Marr, D. (1982). *Vision*. Freeman, San Francisco.

Matute, H., & Pineño, O. (1998). Stimulus competition in the absence of compound conditioning. *Animal Learning and Behavior*, *26*(1), 3–14.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*, 419–457.

Medina, J. F., Nores, W. L., Ohyama, T., & Mauk, M. D. (2000). Mechanisms of cerebellar learning suggested by eyelid conditioning. *Current Opinions in Neurobiology*, *10*(6), 717–724.

Miller, N. E., & Dollard, J. C. (1941). *Social learning and imitation*. Yale University Press, New Haven, CT.

Miller, R. R. (2002) Personal communication.

Miller, R. R., & Escobar, M. (2002). Associative interfence between cues and between outcomes presented together and presented apart: an integration. *Behavioral Processes*, *57*, 163–185.

Miller, R. R., & Matute, H. (1998). Biological significance in forward and backward blocking: resolution of a discrepancy betwee animals and human causal judgement. *Journal of Experimental Psychology: General*, *125*, 370–386.

Miller, R. R., & Matzel, L. D. (1988). The comparator hypothesis: a response rule for the expression of associations. In Bower, G. H. (Ed.), *The psychology of learning and motivation*, pp. 51–92. Academic, San Diego, CA.

Miller, R. R., & Barnet, R. C. (1993). The role of time in elementary associations. *Current Directions in Psychological Science*, *2*(4), 106–111.

Miller, R. R., Barnet, R. C., & Grahame, N. J. (1995). Assessment of the Rescorla-Wagner model. *Psychological Bulletin*, *117*(3), 363–386.

Mirenowicz, J., & Schultz, W. (1994). Importance of unpredictability for reward responses in primate dopamine neurons. *Journal of Neurophysiology*, *72*, 1024–1027.

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive hebbian learning. *Journal of Neuroscience*, *16*, 1936–1947.

Mowrer, R. R., & Klein, S. B. (1989). Traditional learning theory and the transition to contemporary learning theory. In Mowrer, R. R., & Klein, S. B. (Eds.), *Contemporary Learning Theories*, chap. 1, pp. 3–17. Lawrence Erlbaum Associates.

Mowrer, R. R., & Klein, S. B. (2001). The transitive nature of contemporary learning theory. In Mowrer, R. R., & Klein, S. B. (Eds.), *Handbook of Contemporary Learning Theories*, chap. 1, pp. 1–21. Lawrence Erlbaum Associates, Mahwah, New Jersey.

Myers, K. M., Vogel, E. H., Shin, J., & Wagner, A. R. (2001). A comparison of the Rescorla-Wagner and Pearce models in a negative patterning and a summation problem. *Animal Learning & Behavior*, *29*(1), 36–45.

Nakajima, S. (1997). Failure of inhibition by B over C after A+, AB-, ABC+ training. *Journal of Experimental Psychology: Animal Behavior Processes*, *23*(4), 482–490.

Nakajima, S., & Urushihara, K. (1999). Inhibition and facilitation by B over C after A+, AB-, ABC+ training with multimodality stimulus combinations. *Journal of Experimental Psychology: Animal Behavior Processes*, *25*(1), 68–81.

O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, *34*, 171–175.

O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map*. Oxford University Press, Oxford, England.

O'Reilly, R. C., & Rudy, J. W. (2001). Conjunctive representations in learning and memory: Principles of hippocampal and cortical function.. *Psychological Review*, *108*, 311–345.

Pavlov, I. P. (1927). *Conditioned Reflexes*. Oxford University Press.

Pearce, J. M. (1994). Similarity and discrimination: A selective review and a connectionist model. *Psychological Review*, *101*, 587–607.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: variations in the effectiveness of conditioned but not unconditioned stimuli.. *Psychological Review*, *87*, 532–552.

Pearce, J. M., Aydin, A., & Redhead, E. S. (1997). Configural analysis of summation in autoshaping. *Journal of Experimental Psychology: Animal Behavior Processes*, *23*(1), 84–94.

Pearce, J. M., & George, D. N. (2002). Summation: Further assessment of a configural theory. *Quarterly Journal of Experimental Psychology*, *55B*(1), 61–73.

Pearce, J. M., & Redhead, E. S. (1993). The influence of an irrelavent stimulus on two

discriminations.. *Journal of Experimental Psychology: Animal Behavior Processes*, *19*, 180–190.

Pearl, J. (2000). *Causality*. Cambridge University Press, Cambridge, UK.

Philips, R. G., & LeDoux, J. E. (1992). Differential contribution of amydgala and hippocampus to cued and contextual fear conditioning. *Behavioral Neuroscience*, *106*, 274–285.

Prewit, P. L. (1967). Number of preconditioning trials in sensory preconditioning using cer training. *Journal of Comparative and Physiological Psychology*, *64*, 360–362.

Redhead, E. S., & Pearce, J. M. (1995a). Similarity and discrimination learning. *Quarterly Journal of Experimental Psychology*, *48B*, 46–66.

Redhead, E. S., & Pearce, J. M. (1995b). Stimulus salience and negative patterning. *Quarterly Journal of Experimental Psychology*, *48B*, 67–83.

Rescorla, R. A. (1972). "Configural" conditioning in discrete-trial bar pressing. *Journal of Comparative and Physiological Psychology*, *79*(2), 307–317.

Rescorla, R. A. (1980). *Pavlovian Second-Order conditioning*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Rescorla, R. A. (1982). Simultaneous second-order conditioning produces S-S learning in conditioned suppression. *Journal of Experimental Psychology: Animal Behavior Processes*, *82*, 23–32.

Rescorla, R. A. (1997). Summation: Assessment of a configural theory. *Animal Learning and Behavior*, *25*, 200–209.

Rescorla, R. A. (1999). Within-subject partial reinforcement extinction effect in autoshaping. *Quarterly Journal of Experimental Psychology*, *52B*, 75–87.

Rescorla, R. A. (2003a). Elemental and configural encoding of the conditioned stimulus. *Quarterly Journal of Experimental Psychology*, *56B*(2), 161–176.

Rescorla, R. A. (2003b). Protection from extinction. *Learning & Behavior*, *31*(2), 124–132.

Rescorla, R. A., & Coldwell, S. E. (1995). Summation in autoshaping. *Animal Learning & Behavior*, *23*, 314–326.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement.. In Black, A. H., & Prokasy, W. F. (Eds.), *Classical Conditioning II*. Appleton-Century-Crofts.

Rescorla, R. A. (1999). Associative changes in elements and compounds when the other is reinforced. *Journal of Experimental Psychology: Animal Behavior Processes*, *25*(2), 247–255.

Rizley, R. C., & Rescorla, R. A. (1972). Associations in second-order conditioning and sensory preconditioning. *Journal of Comparative and Physiological Psychology*, *81*, 1–11.

Robert, C. P., & Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer.

Rudy, J. W., & Sutherland, R. J. (1989). The hippocampal formation is necessary for rats to learn and remember configural discriminations. *Behavioural Brain Research*, *34*, 97–109.

Rudy, J. W., & Sutherland, R. J. (1995). Configural association theory and the hippocampal formation: An appraisal and reconfiguration. *Hippocampus*, *5*, 375–389.

Saul, L. K., Jaakkola, T., & Jordan, M. I. (1996). Mean field theory for sigmoid belief networks. *Journal of Artificial Intelligence Research*, *4*, 61–76.

Savastano, H. I., Cole, R. P., Barnet, R. C., & Miller, R. R. (1999). Reconsidering conditioned inhibition. *Learning and Motivation*, *30*, 101–127.

Schmajuk, N. A., & DiCarlo, J. J. (1992). Stimulus configuration, classical conditioning, and hippocampal function. *Psychological Review*, *99*(2), 268–305.

Schooler, L. (2001). Rational theory of cognition in psychology. In Smelser, N. J., & Baltes, P. B. (Eds.), *International Encyclopedia of the Social & Behavioral Sciences*, pp. 12771–12775. Elsevier Science Ltd.

Schreurs, B. G., Kehoe, E. J., & Gormezano, I. (1993). Concurrent associative transfer and competition in serial conditioning of the rabbit's nictitating membrane response. *Learning and Motivation*, *24*, 395–412.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, *80*, 1–27.

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

Seligman, M. E. P. (1970). On the generality of the laws of learning. *Psychological Reviewychological*, *77*, 406–418.

Seward, J. P. (1942). An experimental study of guthrie's theory of reinforcement. *Journal of Experimental Psychology*, *30*, 247–256.

Skinner, B. F. (1938). *The behavior of organisms. An experimental analysis*. Appleton-Century-Crofts, New York.

Smith, A. F. M., & Makov, U. E. (1978). A quasi-Bayes sequential procedure for mixtures. *Journal of the Royal Statistical Society*, *40*(1), 106–112.

Suri, R. E., & Schultz, W. (2001). Temporal difference model reproduces anticipatory neural activity. *Neural Computation*, *13*(4), 841–862.

Sutherland, R. J., & Rudy, J. W. (1989). Configural association theory: The role of the hippocampus formation in learning, memory, and amnesia. *Psychobiology*, *17*(2), 129–144.

Sutton, R. S. (1988). Learning to predict by the methods of temporal difference. *Machine*

Learning, *3*, 9–44.

Sutton, R. S. (1992). Gain adaptation beats least squares?. In *Proceedings of the 7th Yale Workshop on Adaptive and Learning Systems*, pp. 161–166.

Sutton, R. S., & Barto, A. G. (1990). Time-derivative models of Pavlovian reinforcement. In Gabriel, M., & Moore, J. (Eds.), *Learning and Computational Neuroscience: Foundations of Adaptive Networks*, chap. 12, pp. 497–537. MIT Press.

Sutton, R. S., & Pinette, B. (1985). The learning of world models by connectionist networks. In Erlbaum, L. (Ed.), *Proceedings of the seventh annual conference of the cognitive science society*, pp. 54–64, Irvine, California.

Tenenbaum, J. B., & Griffiths, T. L. (2001a). Generalization, similarity and Bayesian inference. *Behavioral and Brain Sciences*, *24*, 629–640.

Tenenbaum, J., & Griffiths, T. (2001b). Structure learning in human causal induction. In Leen, T. K., Dietterich, T. G., & Tresp, V. (Eds.), *Advances in Neural Information Processing Systems 13*, pp. 59–65, Cambridge, MA. MIT Press.

Tenenbaum, J. B., & Griffiths, T. L. (2003). Theory-based causal inference. In S. Becker, S. T., & Obermayer, K. (Eds.), *Advances in Neural Information Processing Systems 15*, pp. 35–42, Cambridge, MA. MIT Press.

Thorndike, E. L. (1911). *Animal intelligence: experimental studies*. Macmillan, New York.

Thorndike, E. L. (1913). *The psychology of learning*. Teachers College Press, New York.

Thorpe, W. H. (1956). *Learning and instinct in animals*. Methuen, London.

Tolman, E. C. (1932). *Purposive Behavior in Animals and Men*. Century, New York.

Tolman, E. C. (1955). Principles of performance. *Psychological Review*, *62*, 315–325.

Tolman, E. C., & Brunswick, E. (1935). The organism and the causal texture of the environment. *Psychological Review*, *42*, 43–77.

Tolman, E. C., & Honzik, C. H. (1930). Introduction and removal of reward, and maze performance in rats. *University of California Publications in Psychology*, *4*, 257–275.

Underwood, B. J. (1966). *Experimental psychology* (2nd edition). Appleton-Century-Crofts, New York.

Wagner, A. R., Haberlandt, F. A., & Price, T. (1968). Stimulus selection in animal discrimination learning. *Journal of Experimental Psychology*, *76*(2), 171–180.

Wagner, A. R., & Rescorla, R. A. (1972). Inhibition in Pavlovian conditioning: Application of a theory. In Boakes, R. A., & Halliday, M. S. (Eds.), *Inhibition and Learning*, pp. 301–366. Academic Press, New York.

Wagner, A. R., & Brandon, S. E. (2001). A componential theory of pavlovian conditioning. In *Handbook of Contemporary Learning Theories*, chap. 2, pp. 23–64. Lawrence Erlbaum

Associates, Mahwah, New Jersey.

Wasserman, E. A., & Miller, R. R. (1997). What's elementary about associative learning. *Annual Review of Psychology*, *48*, 573–607.

Whishaw, I. Q., & Tomie, J. A. (1991). Acquisition and retention by hippocampal rats of simple, conditional, and configural tasks using tactile and olfactory cues: Implications for hippocampal function. *Behavioral Neuroscience*, *105*, 787–797.

Wickelgren, W. A. (1979). Chunking and consolidation: A theoretical synthesis of semantic networks, configuring in conditioning S-R versus cognitive learning, normal forgetting, the amnesic syndrome, and the hippocampus arousal system. *Psychological Review*, *86*, 44–60.

Williams, B. A. (1994). Blocking despite changes in reinforcer identity. *Animal Learning & Behavior*, *22*, 442–457.

Wilson, P. N., Boumphrey, P., & Pearce, J. M. (1992). Restoration of the orienting response to a light by a change in its predictive accuracy. *Quarterly Journal of Experimental Psychology*, *44B*, 17–36.

Yin, H., Barnet, R. C., & Miller, R. R. (1994). Second-order conditioning and Pavlovian conditioned inhibition: Operational similarities and differences. *Journal of Experimental Psychology: Animal Behavior Processes*, *20*(4), 419–428.

Yu, A. J., & Dayan, P. (2003). Expected and unexpected uncertainty: ACH and NE in the neocortex. In *Advances in Neural Information Processing Systems 15*, pp. 157–164.

Yuille, A. (2005). The Rescorla–Wagner algorithm and maximum likelihood estimation of causal parameters. In Saul, L. K., Weiss, Y., & Bottou, L. . (Eds.), *Advances in Neural Information Processing Systems 17*, pp. 1585–1592, Cambridge, MA. MIT Press.

Document Log:

Manuscript Version 1 — January 19, 2003

Typeset by $\mathcal{AMS}$-LaTeX — 14 June 2006

AARON C. COURVILLE

ROBOTICS INSTITUTE, CARNEGIE MELLON UNIVERSITY, 5000 FORBES AVE., PITTSBURGH, PA 15213, USA,

E-mail address: aaronc@ri.cmu.edu

Typeset by $\mathcal{AMS}$-LaTeX