CITY, UNIVERSITY OF LONDON

DOCTORAL THESIS

The Rescorla-Wagner Drift-Diffusion Model

Author:
André LUZARDO

Supervisor:
Dr. Eduardo ALONSO
First supervisor
Dr. Artur GARCEZ Second supervisor
Dr. Esther MONDRAGÓN
External advisor

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

in the

Machine Learning Group
Department of Computer Science

6th February 2018


CITY
UNIVERSITY OF LONDON
EST 1894

Declaration of Authorship

I, André LUZARDO, declare that this thesis titled, 'The Rescorla-Wagner Drift-Diffusion Model' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- The University Librarian may exercise his powers of discretion to allow this thesis to be copied in whole or in part without further reference to the author.

Signed:

Date:

Abstract

André LUZARDO

The Rescorla-Wagner Drift-Diffusion Model

Computational models of classical conditioning have made significant contributions to the theoretic understanding of associative learning, yet they still struggle when the temporal aspects of conditioning are taken into account. Interval timing models have contributed a rich variety of time representations and provided accurate predictions for the timing of responses, but they usually have little to say about associative learning. In this thesis we present a unified model of conditioning and timing that is based on the influential Rescorla-Wagner conditioning model and the more recently developed Timing Drift-Diffusion model. We test the model by simulating 11 experimental phenomena and show that it can provide an adequate account for 9, and a partial account for the other 2. We argue that the model can account for more phenomena in the chosen set than these other similar in scope models: CSC-TD, MS-TD, Learning to Time and Modular Theory. A comparison and analysis of the mechanisms in these models is provided, with a focus on the types of time representation and associative learning rule used.

Acknowledgements

I would like to thank my sponsors (CAPES, Brazil) for the financial support, and my supervisors for their guidance.

Contents

List of Figures

List of Tables

List of Abbreviations

BeT	Behavioural Theory of Time
CSC	Complete Serial Compound
CS	Conditioned Stimulus
CR	Conditioned Response
CV	Coefficient of Variation
E	Excitatory
I	Inhibitory
ISI	Inter Stimulus Interval
LeT	Learning to Time
LSM	Least Means Squares
LSTM	Long Short-Term Memory network
MS-TD	Microstimulus Temporal Difference
MLP	Multi-Layer Perceptron
MoT	Modular Theory
MTS	Multiple Time Scales
O	Organism
RW	Rescorla-Wagner
RWDDM	Rescorla-Wagner Drift-Diffusion Model
S-D	Schmajuk-DiCarlo Model
SET	Scalar Expectancy Theory
STM	Spectral Timing Model
SOP	Sometimes Opponent Processes or Standard Operating Procedures
SSCC	Simultaneous and Serial Configural-cue Compound
TD	Temporal Difference
TDDM	Timing Drift-Diffusion Model
TILT	Timing from Inverse Laplace Transform
US	Unconditioned Stimulus
UR	Unconditioned Response

List of RWDDM Mathematical Symbols

Ψ	accumulator or timer
A	accumulator rate
m	accumulator noise factor
θ	accumulator threshold
t	time
t^*	time of US occurrence
α_t	accumulator slope learning rate
x	stimulus representation
σ	standard deviation of the stimulus representation
V	associative strength
H	motivational parameter for the US
α_V	associative strength learning rate
n	trial number

Chapter 1

Introduction

Classical conditioning theories aim to understand how associations between stimuli are learned. Ever since Pavlov, 1927 the process of association formation has been understood to depend crucially on the temporal relations between stimuli (Savastano and Miller, 1998; Balsam, Fairhurst and Gallistel, 2006; Kirkpatrick, 2013). Yet, classical conditioning theories have so far struggled to work when time is taken into account as an attribute of the stimulus representation. The study of time as a mental representation is the object of a separate area of study known as interval timing. Interval timing theories have produced a rich variety of time representations (Gibbon, Church and Meck, 1984; Killeen and Fetterman, 1988; Machado, 1997; Staddon and Higa, 1999; Matell and Meck, 2004), and therefore are a natural place to look for ways to integrate time into classical conditioning. In this thesis I first analyse previous efforts in this direction before introducing a new hybrid classical conditioning and timing model.

The process of association formation is understood to be of fundamental survival value for both human and non-human animals. Prediction, which forms the core of classical conditioning, allows the organism to adapt to significant events in its surroundings. A prototypical experiment in classical conditioning, a type of associative learning, involves a neutral stimulus and an unconditioned stimulus (US) which is capable of eliciting an unconditioned response (UR). After repeated pairings of both stimuli in a specified order and temporal distance, the neutral stimulus comes to elicit a response similar to the UR. This response is called the conditioned response (CR) and the neutral stimulus is said to have become a conditioned stimulus (CS). Classical conditioning theories typically conceptualize this process as the formation of a link (association) between the internal representations of CS and US. Their basic building blocks are (Pearce and Bouton, 2001; Brandon, Vogel and Wagner, 2002): (a) the representations of stimuli, and (b) a learning rule to update the association weights between these representations. Although most theories do not attempt to find neurophysiological correlates, these constructs are nonetheless commonly assumed to be instantiated by (a) neural activity in the form of spike rates, and (b) synaptic plasticity (Moore, 2002; Klopf, 1988; Gallistel and Matzel, 2013). These have found some support in the neuroscientific literature, particularly studies of the role of dopamine in reward prediction (Schultz, Dayan and Montague, 1997; Dayan and

Niv, 2008; Niv, 2009; Eshel, 2016). However it is important to note that there is still no widely accepted complete neural mechanism for classical conditioning and that most theories stay at the computational level of explanation.

Stimulus representations are generally thought of as neural activation that is elicited by the stimulus, which may linger for a short time as a 'trace' after stimulus offset. Representations are commonly one of two types: molar or componential. Molar (or elemental) trace theories treat the stimulus as a single conceptualized unit whose activity is usually assumed to peak quite early following stimulus onset, and then gradually decrease (Hull, 1943; Wagner, 1981; Sutton and Barto, 1981; Schmajuk and Moore, 1988; McLaren and Mackintosh, 2000; Harris and Livesey, 2010). In contrast, componential trace theories break down the CS representation into smaller units, each capable of being associated with the US, with some units more active early during the CS and others late, but all leaving a trace after activation (Desmond and Moore, 1988; Grossberg and Schmajuk, 1989; Vogel, Brandon and Wagner, 2003; Ludvig, Sutton and Kehoe, 2008).

Learning rules may be classified according to different criteria. An important period in the recent history of the field gave rise to one of these criteria. Prior to 1970's conditioning used to be rooted in the stimulus-response tradition, which attributed crucial importance to the temporal pairing, or contiguity, of stimuli for the development of associations. The linear operator learning rule (Hull, 1943) is one of the products of that period. In the late 1960's and early 1970's important experimental discoveries using compound stimuli, that is, a stimulus formed by combining other individual stimuli, showed the contiguity view to be incomplete (Kamin, 1968; Rescorla, 1988; Gallistel and Gibbon, 2001). These compound experiments indicated that the formation of associations also depended on the reinforcement history of the individual elements forming the compound stimulus. This led to the development of new learning rules (Rescorla and Wagner, 1972; Mackintosh, 1975a; Pearce and Hall, 1980) capable of combining individual reinforcement histories in compounds, which the linear operator rule cannot. The first, and arguably still the most influential, of these learning rules is the Rescorla-Wagner (RW, Rescorla and Wagner, 1972). It has become famous for being the first model able to provide an account for the blocking effect (Kamin, 1968), where a novel CS does not become associated with the US if it is reinforced only in compound with a previously conditioned CS.

The CR is usually not a single event. Organisms time their responses so that they emerge gradually during the duration of the CS and reach maximum frequency or intensity around the time of reinforcement. Interval timing theories have attempted to provide an account for this *timing* of the CR. One of the fundamental properties of timing behaviour is that it is approximately timescale invariant, i.e. the whole response distribution scales with the interval being timed (Gibbon, 1977; Allman et al., 2014).One of the consequences of timescale invariance is that the coefficient of variation, that is the standard deviation divided by the mean, of the dependent measure of timing is approximately constant. A number of timing models have put

forth explanations for timescale invariance and other timing properties (how time is encoded, how it is stored in memory and how it gets translated into behaviour) by recourse to an internal pacemaker. The most influential pacemaker-based timing theory to date is Scalar Expectancy Theory (SET, Gibbon, Church and Meck, 1984; Gibbon and Church, 1984). The pacemaker is supposed to mark the passage of time by emitting pulses. These pulses can be gated to an accumulator via a switch which closes at the start of a relevant interval and opens when the interval is finished. The accumulator count is kept in working memory. At the end of the interval the current count is transferred to a long-term reference memory. Behaviour is guided by the action of a comparator which actively compares the count in working memory to the one retrieved from reference memory.

In spite of the considerable overlap, interval timing and classical conditioning are not easily integrated. Most conditioning theories are trial-based, that is they consider the trial as the unit of time. A trial is generally taken to be the state where a CS is present (or CSs in compound) and which may or may not contain a US (or USs). The most influential model in this category is the Rescorla-Wagner (RW, Rescorla and Wagner, 1972). In order to account for different stimulus durations, trial-based theories like RW must resort to some sort of time discretization, usually by subdividing the trial into 'mini-trials'. Each mini-trial is treated as a trial in its own right, which are then used to update associative links. This gives rise to the problem of deciding on a particular discretization. Also, given that humans experience time passing as a continuous flow, it is unlikely that animals discretize their conditioning experience in such a way. A more realistic approach to timing is taken by real-time theories. These theories attempt to formalize the concept of a continuous flow of time.

The Temporal Difference model (TD, Sutton and Barto, 1990; Sutton and Barto, 1998) was one of the earliest and still most influential real-time classical conditioning model. It may be thought of as a real-time version of RW. When used with stimulus representations such as the Complete Serial Compound (CSC, Moore, Choi and Brunzell, 1998), Microstimuli (MS, Ludvig, Sutton and Kehoe, 2008; Ludvig, Sutton and Kehoe, 2012) and the Simultaneous and Serial Configural-cue Compound (SSCC, Mondragón et al., 2014) it is capable of reproducing some timing phenomena like the gradual increase in anticipatory responding that occurs before a signalled reinforcer, and the lower response rates observed during longer CSs. However, only MS-TD has a time representation capable of approximating the most fundamental property of timing, timescale invariance. Another issue with the stimulus representations for TD is that their approach to timing resembles the strategy used by trial-based models, i.e. they all split the stimulus into a number of smaller units or states, the number of which being directly proportional to the duration of the stimulus. Given that conditioning is observed in a timescale that ranges from milliseconds to hours (Kehoe and Macrae, 2002, p. 189) this can lead to a very high number of units being required. The stimulus as a whole no doubt is a complex entity, and the

brain may be employing a large number of neurons to represent it, but to dedicate so many resources only for timing might not be the most energy-efficient strategy. Also, TD and its stimulus representations do not usually account for a change in timing that is not tied to reinforcement. Animals time the occurrence of different events, such as onset and offset of stimuli (see for example Meck and Church, 1984), but TD usually only allows for the timing of rewards.

On the other hand, timing models have made even fewer attempts at integrating aspects of classical conditioning. A notable exception is the Learning to Time (LeT, Machado, 1997; Machado, Malheiro and Erlhagen, 2009) model. It represents the passage of time by transitioning between internal states according to a stochastic pacemaker, an idea borrowed from an earlier timing model called the Behavioural Theory of Time (Killeen and Fetterman, 1988). Learning takes place by associating reinforcement presentation with the current internal state according to the linear operator, a standard classical conditioning rule. LeT offers an account of the basic dynamics of association formation, but it cannot explain cue-competition phenomena like blocking. In a blocking procedure, a CS is first paired with a US until a CR is acquired. The same CS is then presented together with a novel CS and both are paired with the US for a few trials. If the novel CS is now presented alone it elicits little or no responding, and so it is said to be blocked by the first CS. LeT's learning rule, the linear operator, has largely been supplanted by RW in classical conditioning modelling because it cannot explain cue-competition phenomena. Like TD, LeT also employs a representation that requires as many units as time-steps, making it a resource-intense model.

Modular Theory (MoT, Guilhardi, Yi and Church, 2007; Kirkpatrick, 2002) is a timing model which because of its explicit goal of integrating timing and learning may be called a hybrid theory. MoT has introduced novelties that allow it to account for some aspects of the dynamics of classical conditioning that LeT cannot. Its architecture is different than the connectionist one (states or units connected by modifiable links) assumed by RW, TD and LeT. Instead, it uses a more cognitive architecture, with separate information processing stages that deal with perception, memory and decision. It postulates two separate memories: a pattern memory which stores CS durations, and a strength memory which stores the associative strength between each pattern memory and the US. This separation allows MoT to deal with more complex situations involving the dynamics of learning during acquisition and extinction. However, MoT also relies on the linear operator to update its strength memory, which, like LeT, prevents it from accounting for cue-competition phenomena.

Although the models mentioned above, namely TD, LeT and MoT, have accomplished a great deal in terms of bringing together timing and conditioning, they each have their different strengths and weaknesses as I have touched above. In this thesis I introduce a model that tries to address some of these weaknesses while preserving the strengths. More specifically, the model has the following strengths. It represents

time in real-time. Like MoT and unlike LeT and TD, its time representation does not require an arbitrary large number of units or states. Similarly to TD but unlike LeT and MoT, it uses a learning rule that preserves the main features of RW which allow it to account for compound phenomena. It can time the onset and offset of all stimuli, not only of rewards, and store a memory for each. It includes two update rules: one for timing that is updated by time-markers (such as stimulus onset/offset), and another for associations that is updated by the US. Hence, simple stimulus exposure causes the model to learn and store its duration. This capability is not present in models that depend only on an associative learning rule to also learn about time, such as TD and LeT.

This new model is essentially a way to connect one of the most influential classical conditioning theories, the Rescorla-Wagner model (Rescorla and Wagner, 1972), with a recently developed timing theory called Timing Drift-Diffusion Model (TDDM, Rivest and Bengio, 2011; Simen et al., 2011). The TDDM is based on the drift-diffusion model, widely used in decision making theory, and it provides an adaptive time representation that has commonalities with pacemaker-based models like SET and LeT (Simen et al., 2013). These models postulate the existence of a pacemaker that emits pulses at a regular rate, which are then counted to mark the passage of time. To preserve timescale invariance they either postulate a specific type of noise in the memory saved for intervals and a ratio-based decision process (SET) or adapt the rate of pulses (LeT). The TDDM takes the latter route but sets a fixed threshold on pulse counting. To emphasize the unification of these two theories I call our proposal the Rescorla-Wagner Drift-Diffusion Model (RWDDM).

I evaluate RWDDM based on how well it can simulate the behaviour of animals in a number of experimental procedures. Many classical conditioning phenomena have been identified which collectively represent a significant challenge for any single model to explain. A recent list (Alonso and Schmajuk, 2012) has compiled 12 categories, which include acquisition, extinction, conditioned inhibition, stimulus competition, preexposure effects, temporal properties, among others. Of particular interest to a theory of timing and conditioning are phenomena that involve elements of both timing and conditioning. As I detail later, I have searched the literature for documented effects that can challenge the main mechanisms embodied in RWDDM.

I proceed by first reviewing the literature related to learning and timing theories. I then introduce the new model and compare its formalism with four models that have similar scope, namely CSC-TD, MS-TD, MoT and LeT. In the results section I present the phenomena I will simulate, followed by the results of our simulations, and compare them to the current explanations given by LeT, MoT and TD.

1.1 Outputs

1.1.1 Conference Presentations

- Presentation at the XXVI (2014) International Congress of the Spanish Society for Comparative Psychology (Braga, Portugal).
- Presentation at the XXVII (2015) International Congress of the Spanish Society for Comparative Psychology (Seville, Spain).
- Presentation at the 2016 Associative Learning Symposium (Gregynog Hall, Wales).

1.1.2 Publications

- Luzardo, A., Alonso, E., & Mondragón, E. (2017). A Rescorla-Wagner Drift-Diffusion Model of Conditioning and Timing. PLOS Computational Biology, vol: 13 (11) pp: e1005796.
- Luzardo, A., Rivest, F., Alonso, E., & Ludvig, E. A. (2017). A drift–diffusion model of interval timing in the peak procedure. Journal of Mathematical Psychology, 77, 111–123.

1.1.3 Code

The Matlab code to generate the results of RWDDM published in this thesis can be found at https://github.com/ndrluzardo/PhDThesis.

Chapter 2

Literature Review

2.1 Learning Theories

The ability to learn relationships between different stimuli and events is an important adaptive mechanism. In order to survive and reproduce an organism must be able to search for food and mates in an environment that is constantly changing albeit with certain regularities. An organism that is able to learn from these regularities will maximise its chances of survival and reproduction.

Traditionally, two distinct types of associative learning have been recognized: classical (or Pavlovian) and operant (or instrumental) conditioning. In classical conditioning an association is believed to be formed between a stimulus (S) and a response (R), or between a stimulus and another stimulus. In operant conditioning an association is believed to be formed between R and an outcome (O). However the current tendency in the study of learning is to regard the associative structure underlying both S-R, S-S and R-O as fundamentally similar (Gallistel and Gibbon, 2000; Hall, 2002). A strict distinction therefore will not be made between these two procedures here.

This section will review the main learning models, with a focus on their formalisms. These are all connectionist models in the sense that they consist of nodes (or units) which represent the CS and US, and associative links (connections) between these nodes (see figure 2.1 for a generic scheme). As Brandon, Vogel and Wagner, 2002 remarked,

> in such theories, the major theoretical options are centred around three questions. How shall the CSs and USs be represented? How shall the links between stimulus representations be construed to change during conditioning? How do the measures of conditioned responding depend on the current values of the stimulus representations and their associative links? (p. 233)

I will address each of these questions first as they will help contextualise the analysis of learning models that follows.

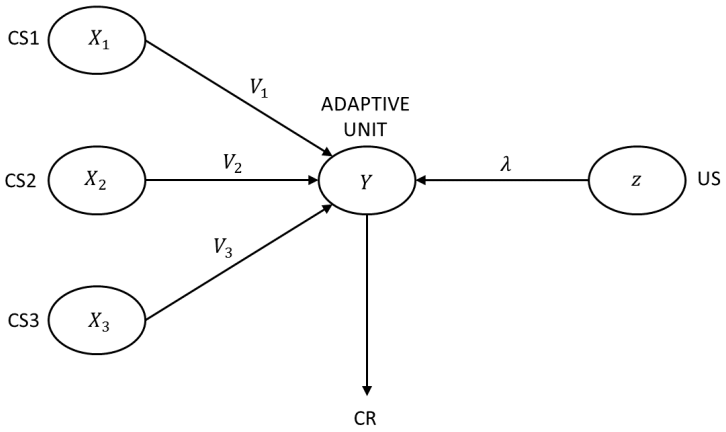


FIGURE 2.1: Schematic representation of classical conditioning theories. When a CS_i is present it activates its respective neuron-like unit X_i. Each CS unit is connected to the main response unit Y by modifiable links V_i. The US unit z is also connected to Y but by an unmodifiable link λ. Adapted from Vogel, Castro and Saavedra, 2004.

2.1.1 Stimulus Representation

Stimulus representation is the problem of finding how the brain codes external physical stimuli. This has a long history, going back to the beginning of experimental psychology with early behavioural theorists like Pavlov and Hull. Hull, 1943 adopted the *stimulus-trace* hypothesis of Pavlov, the idea that an external stimulus generates an internal representation that grows in strength at first, decays slowly until the physical stimulus is gone and then persists for a while as a rapidly decaying trace. Figure 2.2 shows a theoretical example. By and large, similar versions and variations of this simple concept have been adopted by every major learning theory to this day.

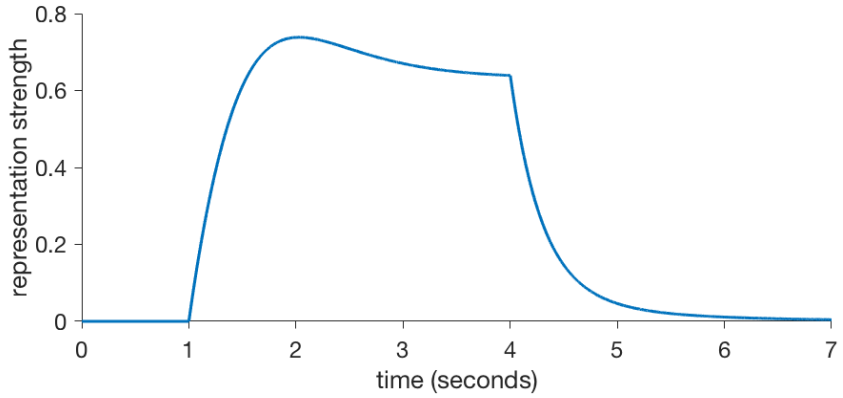


FIGURE 2.2: An idealized stimulus representation as it would be produced if a physical stimulus with constant intensity was presented from 1 to 4 seconds.

A theoretical question arises: what is the fundamental unit of stimulus representation? Is every CS and US to be independently represented or should they share common elements? Learning theories that take the first approach are known as *elemental* and the latter as *configural*. Configural models are broader in scope than elemental ones, as they are able to handle stimulus generalisation and discrimination effects.

2.1.2 Learning Rules

How are the changes in learning actually instantiated? The vast majority of models are based on a theoretical construct called *associative strength* represented by the letter V. V is conceptualized as a modifiable link between the CS and US representations (or, as in figure 2.1, between the CS and an adaptive unit which is also connected to the US).

Historically, temporal contiguity was considered to be a necessary and sufficient cause of learning. It was thought that V, to use modern terminology, increased only when CS and US occurred together. This view was shown to be too simplistic. Based on the work of Rescorla, 1967, Balsam, Drew and Gallistel, 2010 and Balsam and Gallistel, 2009 have argued that temporal contiguity is neither necessary nor sufficient and give two examples. Experiments with conditioned inhibition, where the US is presented only in the absence of the CS (see for example Rescorla, 1969), can turn the CS into an inhibitor and hence demonstrate that temporal contiguity is not necessary. Experiments with blocking (Kamin, 1968), where a CS fails to acquire a CR when it is paired both with a US and with another CS that has already been conditioned, shows that simple temporal contiguity is not sufficient. What these authors (Rescorla, 1967; Balsam and Gallistel, 2009; Balsam, Drew and Gallistel, 2010; Gallistel, Craig and Shahan, 2014) argue is that contingency (correlation), and not contiguity, is the critical basis of learning. As it will be demonstrated later, one of the most successful learning rules, the Rescorla-Wagner rule, was devised as an attempt to capture contingency.

2.1.3 Response Rules

The ongoing difficulty and uncertainty in identifying neural substrates has meant that the most reliable way to assess learning continues to be through behaviour. However, this approach necessarily introduces confounds. Motivation or other factors can interfere with behaviour, thus distorting the expression of learning. Because of this theorists commonly make a distinction between learning and performance.

Learning models generally incorporate a rule that translates associative strength V (a measure of learning) into response Y (a measure of performance). This rule may take the form of a simple multiplication, for example

$$Y = X \cdot V \tag{2.1}$$

where X stands for the CS representation. Variations are of course possible such as the inclusion of a threshold on the value of V below which learning is assumed to be too low to be translated into performance.

2.1.4 Trial-based Models

Trial-based learning models consider the trial as the fundamental unit of time. What constitutes a trial for these models will sometimes vary but it is usually defined as the period between CS onset and US onset.

One of the first, and still most influential, trial-based theories is the Rescorla-Wagner Model (RW, Rescorla and Wagner, 1972). It consists of an error-correction rule that adjusts the level of associative strength V between the internal representations of the CS and US:

$$\Delta V_i = \alpha_i \beta_j \left(z_j \lambda_j - \sum_i X_i V_i \right) X_i. \tag{2.2}$$

with

$$X_i \text{ or } z_j = \begin{cases} 1 & \text{if CS}_i \text{ or US}_j \text{ is present} \\ 0 & \text{if CS}_i \text{ or US}_j \text{ is absent.} \end{cases}$$

Constants α_i and β_j are parameters related with the CS_i and US_j respectively, which together determine the learning rate.

RW's main achievement is in the assumption that all CSs currently present compete for a limited associative strength set by the US. This is what allows RW to explain some cue-competition phenomena such as the blocking effect (Kamin, 1968). In blocking, a CS_1 is first conditioned alone and then it is presented in compound with a CS_2. Despite receiving reinforcement during the compound phase, when the CS_2 is presented alone it does not elicit a response. RW can readily explain this with its competition for association assumption. V_1, the associative strength of CS_1, is at asymptotic level ($V_1 = 1$) at the start of the compound phase, i.e. it has acquired all the available associative strength carried by the US, and hence V_2 does not change:

$$\begin{aligned} \Delta V_2 &= \alpha_2 \beta (1 - X_1 V_1 - X_2 V_2) \\ &= \alpha_2 \beta (1 - 1 - 0) \\ &= 0. \end{aligned}$$

This competition mechanism also explains summation. Here two CSs are first paired separately with the same US until acquisition is complete (it is assumed that $V_1 = V_2 = 1$ at the end of training). In the next phase they are presented as a compound CS_{1+2} and also paired with the US. Initially the compound is able to elicit a stronger CR than either CS did when presented alone, but then soon settles to a lower level. Crucially, when tests are made with the CSs presented separately, CR levels are significantly lower than before the compound phase (Kehoe and Macrae,

2002, p. 207). Equation (2.2) predicts this will happen, since in the beginning of second phase:

$$\begin{aligned}\Delta V_1 = \Delta V_2 &= \alpha\beta(1 - X_1V_1 - X_2V_2),\\ &= \alpha\beta(1 - 1 - 1),\\ &= -\alpha\beta,\end{aligned}$$

and hence both V_1 and V_2 will decrease until they equal each 0.5, half of the US's associative strength. Using a similar reasoning it is easy to see that RW also predicts superconditioning, an increased CR strength that occurs when a CS is paired with a US and another inhibitory CS (Williams and McDevitt, 2002).

Another one of RW's successes over its predecessors is that it can explain correlational experiments. Conditioning strength is positively correlated to the probability of the US in the presence of CS and negatively correlated to the probability of US in the absence of CS (Rescorla, 1968). Here CS_2 stands for situational cues (the context) which is alone present during the intertrial interval and in compound with CS_1 during the trial. The asymptotic associative strength of CS_1, CS_2 and the compound CS_{1+2} are (Rescorla and Wagner, 1972):

$$V_1 = V_{1+2} - V_2, \tag{2.3}$$

$$V_2 = \frac{\pi_2\beta_p}{\pi_2\beta_p - (1-\pi_2)\beta_a}, \tag{2.4}$$

$$V_{1+2} = \frac{\pi_{1+2}\beta_p}{\pi_{1+2}\beta_p - (1-\pi_{1+2})\beta_a}, \tag{2.5}$$

where π_2 and π_{1+2} are the probabilities of reinforcement during CS_2 and CS_{1+2} respectively, and β_a and β_p are the US parameters in the absence and presence of reinforcement respectively. If for example $\pi_2 = \pi_{1+2}$ we have $V_2 = V_{1+2}$ (by (2.4) and (2.5)) and therefore by (2.3)

$$V_1 = 0,$$

in other words, when the probability of US is the same both in the presence and absence of the CS, this CS does not acquire any associative strength, a prediction that is matched by experimental results (Rescorla, 1968).

Conditioned inhibition is also readily explained. Here the US is only present in the absence of the CS, hence $\pi_2 = 1$ and $\pi_{1+2} = 0$. By (2.4) and (2.5), $V_2 = 1$ and $V_{1+2} = 0$, hence by (2.3) $V_1 = -1$, i.e. the CS acquires negative strength. This is verified to be true experimentally by introducing a subsequent phase where the CS is now paired with US. In this phase the CS acquires a CR much more slowly than another CS that had not been inhibited (Rescorla, 1969).

RW has some known limitations. Its main problem is with extinction of conditioned inhibitors. A prediction of the theory says that when an inhibitory CS is repeatedly presented alone its associative strength should increase towards zero and hence lose its inhibitory characteristics. This is because an inhibitory CS has negative V and so $(0 - V)$ will be positive, increasing V towards zero. This prediction however fails to hold experimentally (Zimmer-Hart and Rescorla, 1974). A common solution is to consider conditioned excitation and inhibition as two distinct processes, each with their own associative strength: V_E and V_I. V_E is updated only when $(z_j\lambda_j - \sum X_iV_i) > 0$ and V_I only when $(z_j\lambda_j - \sum X_iV_i) > 0$ (see for example equation 11 in Brandon, Vogel and Wagner, 2003). Finally, a single-unit RW model, as depicted in figure 2.1, cannot explain negative patterning. In this conditioning phenomena, two different CSs signal reward but a compound formed by both of them does not. Animals learn to discriminate appropriately, but a single-unit RW fails to reproduce this behaviour. The answer is to add a second, hidden, unit whose outputs will serve as the input to an output unit (see figure 2.8). I will have more to say about negative patterning and its importance to learning when I cover Artificial Neural Networks in section 2.1.6.

Looking at the problem from a different angle, Mackintosh, 1975a proposed an attention-based theory of learning. It is constructed from an established assumption in theories of selective attention, namely that an organism will pay more or less attention to a stimulus to the extent that this stimulus is a better or worse predictor of changes in reinforcement than the other stimuli available. Attention therefore is assumed here to be not just an intrinsic property of the CS but to also change with experience. This should be contrasted with the assumption in RW that α, a learning rate that is related to the attentional properties of the CS, is constant. Mackintosh's model operates by adjusting α so as to increase it when the CS becomes a better predictor of the US, and decrease otherwise. Formally, $\Delta\alpha_i > 0$ if

$$|z\lambda - X_iV_i| < \left|z\lambda - \sum X_jV_j\right| \tag{2.6}$$

and $\Delta\alpha_i < 0$ if

$$|z\lambda - X_iV_i| \geq \left|z\lambda - \sum X_jV_j\right| \tag{2.7}$$

where $\sum X_jV_j$ is the sum of associative strength of all CSs present in the trial except CS_i. The change in α is made proportional to the discrepancy S between $|z\lambda - X_iV_i|$ and $|z\lambda - \sum X_jV_j|$,

$$\Delta\alpha_i = S \cdot \alpha_i. \tag{2.8}$$

The change in associative strength is given by

$$\Delta V_i = \alpha_i(\lambda - V_i). \tag{2.9}$$

It can be seen that Mackintosh's model departs from RW in two ways: it turns an attention parameter from a constant into a variable, and it makes this parameter, and not RW's CS competition for associative strength, capture CS contingency. Mackintosh's model provides an alternative explanation to blocking and overshadowing (where two equally salient CSs compete for associative strength). Whilst RW explains these phenomena by invoking CS competition for a limited amount of associative strength, Mackintosh's theory uses a principle of learned irrelevance. In both cases, one of the CSs becomes a redundant signal of US, either because it doesn't predict anything new (blocking) or because it is less salient and hence conditions more slowly (overshadowing).

Following Mackintosh, Pearce and Hall, 1980 proposed a model that is also based on the idea that experience can produce changes in CS effectiveness. They point out RW's failure in explaining two variations in the classical blocking experiment. In the first, blocking was attenuated when in the compound phase the CSs were paired with a milder US than in the CS alone phase (Dickinson, Hall and Mackintosh, 1976). In the second, blocking was not observed after only one trial was given with the compound (Mackintosh, 1975b). These phenomena, the authors argue, support Mackintosh's theory that learning involves a change in CS effectiveness. But as the authors also point out, Mackintosh's model, like RW, does not effectively deal with latent inhibition, the retardation in learning that occurs when a CS is first presented alone before conditioning. The solution proposed by Pearce and Hall (PH) is to abandon the idea that conditioning depends on CS competition for limited associative strength set by US (which can also be stated, as the authors do, as a change in US effectiveness) adopted by both RW and Mackintosh, and to rely solely on Mackintosh's idea of changes in CS effectiveness. Unlike Mackintosh, which assumed that a CS is more effective if its a good predictor of its consequences, PH assumes the opposite: a CS is more effective to the extent that it is not an accurate predictor of its consequences. This is formalized as follows:

$$\alpha_i^n = \left| z\lambda^{n-1} - \sum X_i V_i^{n-1} \right|, \tag{2.10}$$

or in words, the CS associability parameter (or learning rate as in RW) in the present trial n is a consequence of how well the CS predicted reinforcement in the last trial $n-1$. The authors acknowledge that such a one-trial change may be too fast and suggest that a moving average could be used instead, but the principle is the same. Change in associative strength is given by

$$\Delta V_i = S_i \cdot \alpha_i \cdot \lambda \tag{2.11}$$

where S_i is a constant that depends on CS salience. Consider as an example how the model explains latent inhibition. It is assumed that the starting value of α for any CS is non-zero. When the CS is presented without reinforcement, $\lambda = 0$ and so α goes to zero in the very first trial by equation (2.10). When reinforcement is

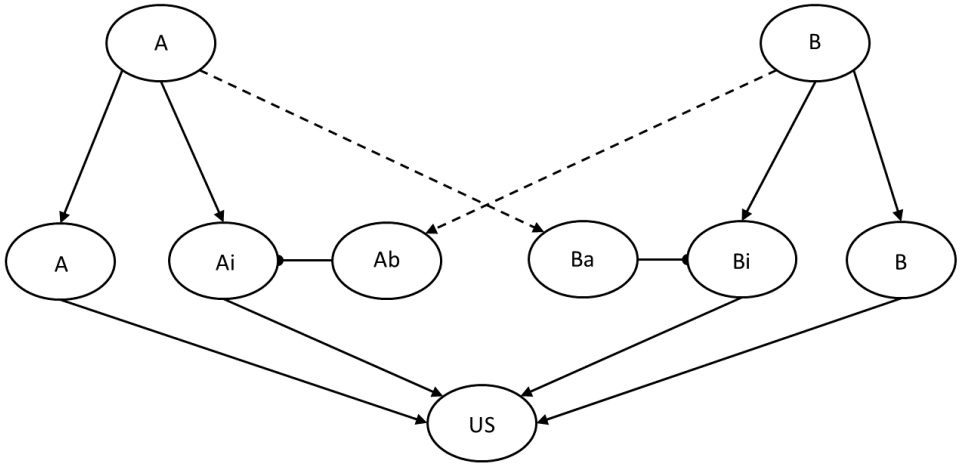
introduced it will therefore take longer for this CS to acquire associative strength than a CS that has not been pre-exposed. A few limitations have been identified since the model has first been proposed. Hall, 2008 notes that the results obtained in Mackintosh, 1975b blocking experiment have not been replicated, and that more research showed some blocking does occur even in the very first trial. This lends support to RW's interpretation.

The models described so far are called *elemental*, i.e. they treat each stimulus representation as an independent element. From this assumption it follows that if two CSs, say A and B, were conditioned separately (A+, B+ where '+' means the US is present) and then presented together as the AB compound, their associative strengths should simply add up. This assumption was challenged by the negative patterning experiment. In this preparation, the subject is asked to discriminate between conditions A+, B+ and a compound AB− where '−' means the US is not present. Studies show that the subject is able to make such a discrimination, albeit not quickly, demonstrating that CSs interact in a more sophisticated fashion than the simple summation of RW.

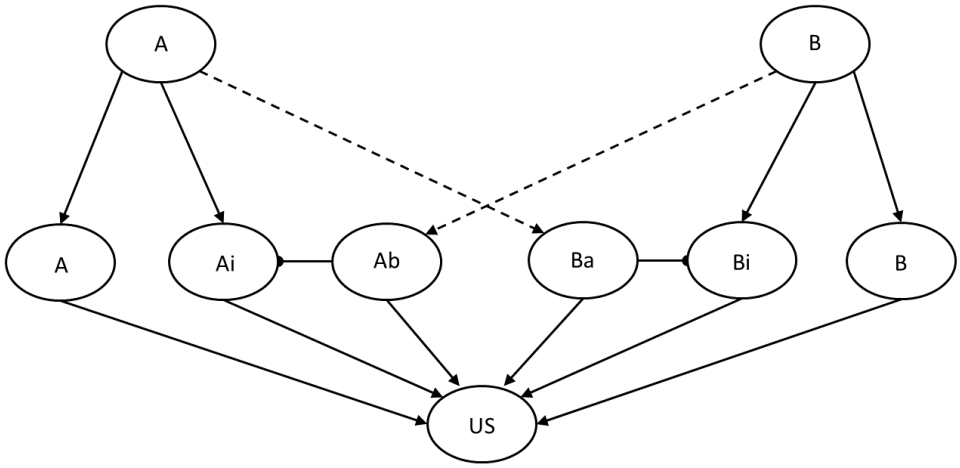
An alternative way is to conceive of a compound as a *configuration*. Pearce, 1987 developed a theory where configuration and generalization play a central role. In this theory A, B, and AB are all taken as configurations, each represented by a distinct neuron-like unit, with AB sharing half of its elements with each A and B. Pearce's theory makes a further assumption, that the salience of all configurations is equal. This model solves negative patterning because AB is now an independent unit which enters into association with the US.

Wagner and Brandon, 2001 have proposed an equivalent componential version of Pearce, 1987 model and called it inhibited elements. They have also advanced their own componential stimulus representation called replaced elements (Brandon, Vogel and Wagner, 2000). See figure 2.3 for diagrams of these two stimulus representation theories. These theories will not be evaluated further here since they do not add much to the question of time in learning central to this thesis, except to say that they propose to address issues that the simple adding of stimulus in RW does not contemplate.

As we have seen, trial-based models represented the first attempt to understand conditioning phenomena. They are still widely used because of their relative simplicity and usefulness in explaining 'static' conditioning phenomena, that is, where time is not a variable. Their relative simplicity allows for easier abstraction of general learning principles. However, when the duration of stimuli is of experimental concern, or when one wishes to reproduce behaviour in real-time, another class of models is needed. I turn to them next.



(A) Inhibited elements



(B) Replaced elements

FIGURE 2.3: Two theories of stimulus representation. Dashed lines are only active when A and B are presented as a compound. Arrowheads represent excitatory connections whilst dotted ends are inhibitory. Note that the only structural difference between the two models is that in the replaced elements units Ab and Ba are also connected to the US. Adapted from Williams, 2014.

2.1.5 Real-time Models

Because trial-based models compress the whole duration of a trial to one single time point they cannot account for real-time changes in behaviour, the core of many timing phenomena.

Real-time models attempt to describe behaviour as it unfolds. They will be described here either in differential or difference equations.

TD

Arguably, the most influential real-time learning model to date is the Temporal Difference (TD; Sutton and Barto, 1990; Sutton and Barto, 1998). TD builds on the idea of expectation of US as a time derivative computation, first introduced by the same authors in an earlier model known as S-B (Sutton and Barto, 1981). S-B will be discussed first as it provides an introduction to TD.

Sutton and Barto, 1981 made an attempt to update adaptive neural networks with the principles discovered in animal learning theory. These two areas had been developing in parallel and with little contact. Sutton and Barto theorized that some internal processing analogous to cellular activity inside a neuron must take place inside an adaptive unit (Y in figure 2.1). This activity is described as a trace generated by the weighted average $\bar{Y}$ of ongoing activity Y,

$$\bar{Y}(t+1) = \beta\bar{Y}(t) + (1-\beta)Y(t). \tag{2.12}$$

If a US is present $Y(t) = 1 + \sum_i X_i V_i$, if US is absent $Y(t) = \sum_i X_i V_i$. CS_i is represented by another trace, called eligibility trace:

$$\bar{X}_i(t+1) = \alpha\bar{X}_i(t) + X_i(t), \tag{2.13}$$

where α is a trace decay constant. It is assumed that $X_i = 1$ when CS_i is present and 0 otherwise. Change in associative strength is computed by

$$\Delta V_i(t+1) = c[Y(t) - \bar{Y}(t)]\bar{X}_i(t), \tag{2.14}$$

where c is a learning rate.

The model produces results that are very similar to RW, to the extent that Sutton and Barto, 1981 consider it a 'temporally refined extension of the Rescorla-Wagner model'. They demonstrate by simulations that S-B can reproduce roughly the interstimulus effect, a conditioning-timing phenomena where the magnitude of the conditioned response decreases with increasing interstimulus (CS onset-US onset) interval (ISI). They also show that S-B can correctly assign associative strength to a CS_1 that starts earlier and then overlaps with a CS_2 which becomes redundant in this situation. Because its learning rule is based on the term $[Y(t) - \bar{Y}(t)]$, learning can occur by the activity of CSs only, allowing the model to also explain second order

conditioning, where a CS_1 previously conditioned with a US can make a CS_2 also acquire some associative strength when it is paired with CS_1. Barto and Sutton, 1982 demonstrated that S-B can produce conditioned inhibition.

Sutton and Barto, 1990 identified problems with S-B's account of the ISI effect and proposed a solution. Under more realistic simulations S-B predicted that a CS would become strongly inhibitory if simultaneously presented with US (ISI=0). This contradicts the data which shows that in this condition the CS actually becomes excitatory. Another problem was that S-B predicted equal conditioning strength for almost all ISIs in delay conditioning, whilst the ISI effect predicts a decay in strength with ISI. The solution they proposed was the Temporal Difference (TD) model.

TD introduces the idea of discounted future rewards into the time derivative S-B model. TD assumes that future rewards λ are less valuable and so should be discounted in proportion to their distance from the present moment. It then creates a prediction $\bar{V}$ of the sum of these discounted rewards,

$$\bar{V}(t) = \lambda(t+1) + \gamma\lambda(t+2) + \gamma^2\lambda(t+3) + ... \tag{2.15}$$

where γ is a discount factor. The authors then derive a replacement for S-B's reinforcement term $[Y(t) - \bar{Y}(t)]$ as follows. First they note that:

$$\begin{aligned}\bar{V}(t) &= \lambda(t+1) + \gamma\lambda(t+2) + \gamma^2\lambda(t+3) + ... \\ &= \lambda(t+1) + \gamma(\lambda(t+2) + \gamma^2\lambda(t+3) + ...) \\ &= \lambda(t+1) + \gamma\bar{V}(t+1).\end{aligned}$$

The time derivative of prediction therefore is

$$\lambda(t+1) + \gamma\bar{V}(t+1) - \bar{V}(t). \tag{2.16}$$

Substituting this for $[Y(t) - \bar{Y}(t)]$ into (2.14) yields the TD learning rule:

$$\Delta V_i(t+1) = c[\lambda(t+1) + \gamma\bar{V}(t+1) - \bar{V}(t)]\bar{X}_i(t). \tag{2.17}$$

Figure 2.4 shows a diagram of TD.

Sutton and Barto, 1990 showed that TD can fix the problems in S-B and provide a better account of ISI effect in both delay and trace conditioning. They also demonstrate how it can exhibit blocking and second-order conditioning. It can also reproduce the resistance to extinction of a conditioned inhibitor if the following condition is made:

$$\bar{V} = \begin{cases} \sum_i X_i V_i & \text{if } \sum_i X_i V_i > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{2.18}$$

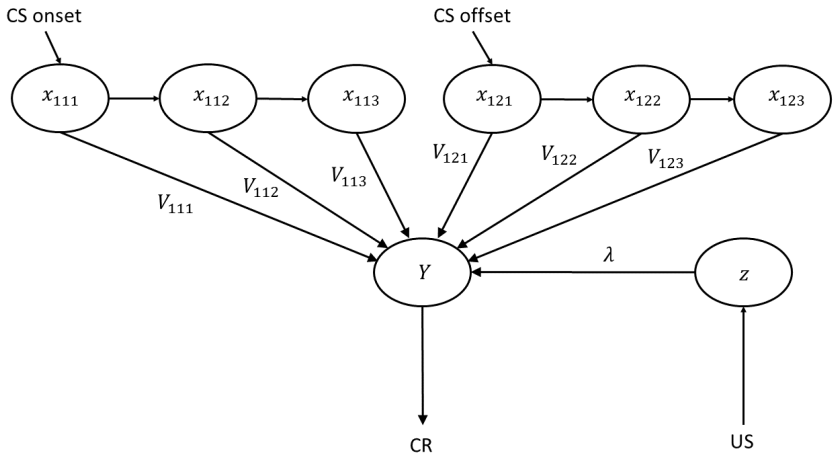


FIGURE 2.4: TD learning model with a complete serial compound (CSC) stimulus representation. In this representation both the onset and offset of the CS are instantiated by different x units. The subscripts ijk represent the CS(i), the onset or offset (j) and the order of activation (k). Adapted from Vogel, Castro and Saavedra, 2004.

TD has found many applications in the field of artificial intelligence and adaptive control (Sutton and Barto, 1998) and constitutes the basis of modern deep learning algorithms (Mnih et al., 2015). It has received a particular boost in popularity in the field of neuroscience. Schultz, Dayan and Montague, 1997 have shown that dopaminergic neurons fire in anticipation of a reward in a manner resembling TD predictions.

Ludvig, Sutton and Kehoe, 2008 proposed a stimulus representation to be used with TD with the aim of refining its timing predictions. Even though CSC-TD could explain some timing phenomena, it was not able to explain timescale invariance. This timing phenomena refers to the finding that the timing of the CR tends to scale with stimulus duration. Microstimuli (MS), as the representation is called, formalises the CS as a series of Gaussians

$$f(y, \mu, \sigma) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right), \tag{2.19}$$

where y is an exponentially decaying trace set at 1 at CS onset. The ith microstimulus is given by:

$$X_i(t) = f(y(t), i/m, \sigma)y(t), \tag{2.20}$$

where m is the total number of microstimuli. The right panel in figure 2.5 shows the structure of microstimuli.

Ludvig, Sutton and Kehoe, 2012 compared microstimuli against two other previous CS representations used with TD. The first, known as *presence*, assumes only one representation, or only one neuron-like unit X, of the CS (see left panel in figure

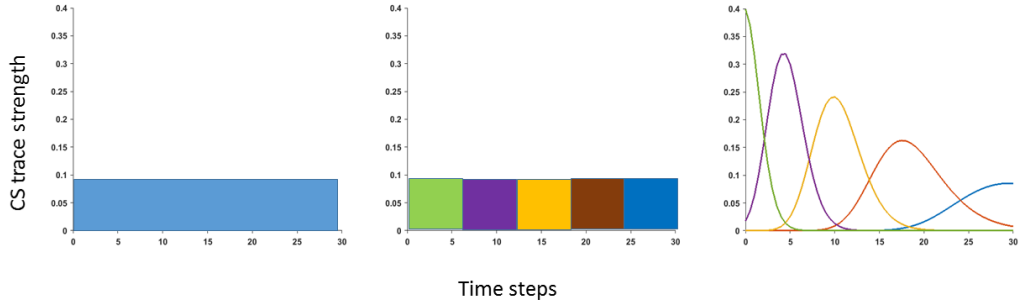


FIGURE 2.5: The choices of stimulus representation in TD. Left panel: presence. Middle panel: complete serial compound. Right panel: microstimuli.

2.5). It is arguably the most basic representation but it is able to explain the ISI effect on conditioning (Sutton and Barto, 1990). But since it represents absolute temporal generalization, it cannot account for the intertrial temporal dynamics. The other TD stimulus representation is the *Complete Serial Compound* (CSC) first proposed by Moore, Choi and Brunzell, 1998. In CSC the CS is composed of a series of neuron-like units, one for each time-step (see middle panel in figure 2.5). Such a representation is not very realistic in terms of neurophysiology as it postulates the existence of an arbitrarily large number of neuron-like units, but it is a clear improvement on the presence representation as shown by Sutton and Barto, 1990. Ludvig, Sutton and Kehoe, 2012 demonstrated that microstimuli fared the same or better than the other representations on the ISI effect, CR timing, CR scalar invariance, blocking, blocking with a change in ISI and overshadowing. The authors point out that TD with MS cannot account for certain phenomena such as discrimination, preexposure and recovery.

SOPs

Wagner, 1981 developed a model intended to describe standard operating procedures (SOP) in memory. In this model (see figure 2.6) CS and US are conceptualized as a set of individual units or elements. Elements transition between three states. CS onset excites elements into state A1 with probability $p1$. Elements in A1 state gradually decay into state A2 with probability $pd1$. Elements in state A2 gradually decay into an inactive state I with probability $pd2$. From state I an element can either stay inactive, if CS is no longer present, or transition back to state A1 with probability $p1$,

if the CS is still present. The change in the proportion of elements in states A1, A2 and I are given by:

$$\Delta p_{A1} = p1(p_I) - pd1(p_{A1}) \tag{2.21}$$

$$\Delta p_{A2} = pd1(p_{A1}) - pd2(p_{A2}) \tag{2.22}$$

$$\Delta p_I = pd2(p_{A2}) \tag{2.23}$$

where p_{A1}, p_{A2} and p_I are the proportion of elements in states A1, A2 and I respectively. These equations produce a stimulus representation for A1 activation similar to the one in figure 2.2.

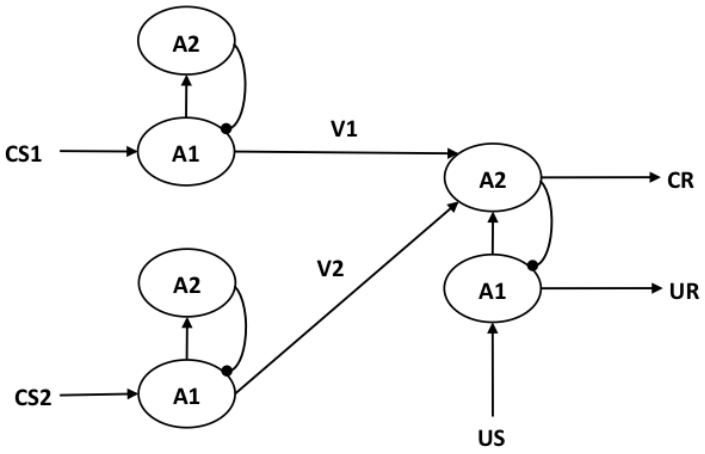


FIGURE 2.6: SOP model network. Arrows represent excitatory connections and circles inhibitory. Adapted from Brandon, Vogel and Wagner (2002, p. 237).

The stimulus representation in SOP allows the model to reproduce habituation effects. Brandon, Vogel and Wagner, 2003 simulated repeated presentations of the US, demonstrating that the model produces weaker URs with massed US training, an effect well-known empirically.

SOP's learning rules are:

$$\Delta V_i^+ = L^+ \sum_i p_{A1,CSi} \times p_{A1,US_j} \quad \text{for excitatory learning,} \tag{2.24}$$

$$\Delta V_i^- = L^- \sum_i p_{A1,CSi} \times p_{A2,US_j} \quad \text{for inhibitory learning,} \tag{2.25}$$

$$\Delta V_i = \Delta V_i^+ - \Delta V_i^-, \tag{2.26}$$

where L^+, L^- are constants. The modifiable associative strength V is the net result of inhibitory V^- and excitatory V^+ strengths.

Conditioning is assumed to change the dynamics of elements in the US node by establishing a link between the inactive I state and $A2$. Hence, after learning is established, CS presence will cause elements in the US node to transition from the inactive state into state $A2$. This activity is mediated by the modifiable weight $p2$

given by

$$p2_{US/\sum CS} = \sum^{i} V_{CS_i US}(r1pA1_{CS_i} + r2pA2_{CS_i})$$

where r_1, r_2 are constants and $p2$ is restricted to the interval (0,1).

Responses R in the model are generated by:

$$R = f(w_1 p_{A1,US} + w_2 p_{A2,US}), \tag{2.27}$$

where w_1, w_2 are constants. The precise function f is left to be defined by data fits.

Brandon, Vogel and Wagner, 2003 showed that SOP can produce changes in associative strength compatible with acquisition, extinction and cue competition, in a similar manner as Rescorla-Wagner. They also argue SOP is the only model to predict inhibition with backwards conditioning.

In its original formulation SOP does not adequately predict timing phenomena. Brandon, Vogel and Wagner, 2003 attributes this failure to SOP's unitary stimulus representation. Their simulations showed that SOP predicts optimal conditioning with simultaneous CS and US presentation, whilst the data show a minimum ISI is required for optimal conditioning. It does not predict the correct CR timing; its associative strength peaks too early into the trial at the end of training.

Two other variations of SOP have been proposed. The first, called AESOP for affective extension of SOP (Brandon, Vogel and Wagner, 2003), postulates two separate units for the US representation, one emotive and the other sensory-perceptual. The emotive unit gives rise to a conditioned emotive response (CER), a diffuse response that is believed to regulate the CR. The sensory-perceptual unit supports the CR. Apart from this novel US representation, AESOP maintains the assumptions of SOP.

The main advantage of AESOP over SOP is in dealing with CERs. Indeed, the motivation behind its development was to explain the phenomenon called 'divergence of response measures'. It refers to the observation that different measures of conditioned response may yield contrary, or uncorrelated, results. For example, backward eyelid conditioning produces inhibitory learning when eyeblinks are the measure of conditioning but produces excitatory learning when suppresion of drinking is measured. Such phenomena can be accounted by the two distinct US nodes in AESOP. By using different decay constants for these units, AESOP can also correctly account for the ability of CERs to develop at longer ISIs than CRs.

The second SOP variation, called Componential-SOP (C-SOP, Brandon, Vogel and Wagner, 2003), was developed with the specific aim of overcoming SOP's timing limitations. In particular, it was inspired by data on occasion setting and CR timing.

Occasion setting occurs during a conditioning procedure where a CS, usually called a feature, precedes another CS, called a target, which is then reinforced or not. If the target CS is only reinforced when preceded by the feature CS (and not

reinforced when presented alone) the procedure is called a feature positive discrimination. Conversely, if the target CS is not reinforced when preceded by the feature CS the procedure is called a feature negative discrimination. If feature and target do not overlap, i.e. are presented serially, then the feature CS is said to act as an 'occasion setter' for the target CS. The occasion setter is seen as conveying information about the impending target. Of particular interest here is temporal information. A feature CS may indicate that the target CS will come after a certain time interval, what is called a feature-target interval (FTI).

Holland, 1998 performed a series of experiments with FTIs ranging from 5 to 50 seconds. Their results were in line with data obtained in the peak procedure (a common experiment in the timing literature) in that the response curves obtained with different FTIs superimposed when appropriately scaled. Of particular interest are compound features experiments. In a particularly interesting mix of cue competition and timing, Holland, 1998 presented a 10/30 second feature compound and then either the 10 or 30 target CS. The response curves looked very similar to each other (Figure 2 in Holland, 1998), with two apparent peaks, the first higher than the second and both shifted to the right of the target intervals. This points to a kind of time subtraction in the compound feature cue. One way to interpret this result is that the internal clock was slowed down by the compound cue.

As mentioned above, C-SOP is an attempt at providing an explanation for the type of timing phenomena seen in occasion setting and CR timing. SOP represented the CS as a set of elements that can be in one of three states, and its learning rules (equations (2.24) and (2.25))are applied to the whole sets. C-SOP applies these rules directly to the elements. These are assumed to have value 1 when active and 0 when inactive. C-SOP also introduces the assumption that some elements are temporally correlated, showing consistency from trial to trial. Hence, in C-SOP the CS elements belong to one of two classes, one temporally correlated and another randomly distributed.

Brandon, Vogel and Wagner, 2003 argue that C-SOP treats the question of CR timing as one analogous to an AX+, BX− discrimination. Elements that are active during US presentation become excitatory, the ones active only in the absence of the US become inhibitory and the ones active at both times become moderately excitatory. The CS trace is built by adding them at each time step. It is also assumed that each element can carry a limited amount of inhibitory or excitatory associative strength, so that $l^- \leq v_i \leq l^+$. Finally, a constrained version of the RW learning rules is used:

$$\Delta V_i = \begin{cases} \alpha\beta^+(\lambda - \sum v_i)(l^+ - v_i) & \text{if } (\lambda - \sum v_i) > 0, \\ \alpha\beta^-(\lambda - \sum v_i)(v_i - l^-) & \text{if } (\lambda - \sum v_i) < 0, \end{cases} \tag{2.28}$$

where $|l| < \lambda$.

With these modifications Brandon, Vogel and Wagner, 2003 and Vogel, Brandon

and Wagner, 2003 showed that the model can reproduce the ISI effect, CR timing and timescale invariance of CR timing.

Harris

Harris and Livesey, 2010 presented a model that relies heavily on a neural computation known as divisive normalization. This is accomplished by dividing the responses of an individual neuron by the summed activity of a pool of neurons. Normalization is considered an ubiquitous neural computation that may underlie the modulatory effects of visual attention, the encoding of value and the integration of multisensory information (Carandini and Heeger, 2012).

Because the normalization equation is so central to the model, it will be useful to explain it in detail first. The basic idea is that the normalized response of a neuron R_j is given by (Carandini and Heeger, 2012):

$$R_j = \frac{D_j^n}{\sigma^n + \sum_k D_k^n} \tag{2.29}$$

where D_j is the neuron's input, σ a constant and D_k the input from the normalization pool which is considered not normalized. Informally, this equation means that in a background of strong activation it will take a stronger input to reach the same normalized response that a weaker input would produce in a background of weak activity.

Figure 2.7 shows a diagram of the model. A CS is thought to activate elements E. Each element is sparsely connected to other elements (Harris and Livesey, 2010, made each E connected with half of the rest of elements in the network). Each E activates one inhibitory unit I and weakly activates another nearby I unit. Unit I then normalizes E activity (via inhibition) according to the background E activity. The CS also activates an attention network A which inhibit units I. A second normalization occurs between A units via their own inhibitory connections. Associative strength V is carried by the connections between E units.

Formally, the model describes the changes in activity strength of units E, I and A. A distinction is made between two types of unit responses: a potential response R_{pot} and an actual response R. Their relationship is described by:

$$\frac{dR}{dt} = \sigma(R_{\text{pot}} - R), \tag{2.30}$$

with σ a constant. Accordingly, activation of unit E is given by:

$$\frac{dE_x}{dt} = \delta(E_{\text{pot}} - E_x) \tag{2.31}$$

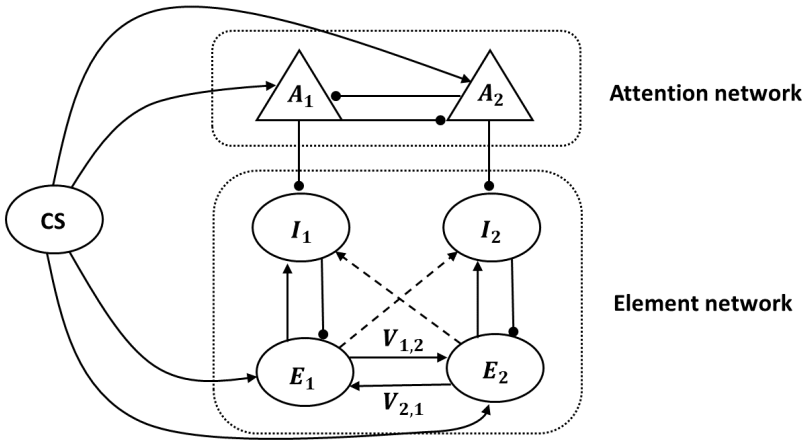


FIGURE 2.7: The attention-modulated associative network. Arrows indicate excitatory connections and dots inhibitory. Adapted from Harris and Livesey, 2010.

where

$$E_{\text{pot}} = \frac{\text{Input}(E_x)^p}{\text{Input}(E_x)^p + I_x^p + D}, \tag{2.32}$$

$$\text{Input}(E_x) = \begin{cases} S_x + \sum_{i=1}^{n} V_i E_i & \text{if } (S_x + \sum_{i=1}^{n} V_i E_i) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.33}$$

where p, D are constants and S_x external input (CS).

Activation of unit I is given by

$$\frac{dI_x}{dt} = \delta(I_{\text{pot}} - I_x), \tag{2.34}$$

where

$$I_{\text{pot}} = \frac{\text{Input}(I_x)^p}{\text{Input}(I_x)^p + (k_a A_x)^p + D}, \tag{2.35}$$

$$\text{Input}(I_x) = \sum_{i=1}^{n} z_{i,x} E_i. \tag{2.36}$$

Here $z_{i,x}$ is the weight between E_i and I_x. This is taken to reflect the degree of similarity between the receptive fields of E_i and E_x. $z_{x,x}$ is set to 1, whilst every other $z_{i,x}$ is less than 1.

Activation of unit A is given by

$$\frac{dA_x}{dt} = \delta(A_{\text{pot}} - A_x), \tag{2.37}$$

where

$$A_{\text{pot}} = \frac{S_x^p}{S_x^p + A_x^{'p} + D}, \tag{2.38}$$

$$A'_x = w \left[\left(\sum_{i=1}^{n} A_i \right) - A_x \right], \tag{2.39}$$

with w a constant.

Finally, associative strength is calculated in a manner that resembles S-B model, but the authors make one modification. Their rule intends to formalize the idea that associative change of the recipient element is proportional to the difference between the change in excitation and the change in inhibition. It also incorporates the idea of a change in salience or associability α common to the trial-based attention models. First, a function that determines the co-activation of elements E_x and E_y is calculated by:

$$\Delta_{x,y} = \alpha_x \left(\beta_E \frac{dE_y}{dt} - \beta_I \frac{dI_y}{dt} \right), \tag{2.40}$$

where

$$\beta_E = \begin{cases} 0.02 & \text{if } \frac{dE_y}{dt} > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{2.41}$$

$$\beta_I = \begin{cases} 0.1 & \text{if } \frac{dI_y}{dt} > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{2.42}$$

Then to make the change in associative strength a gradual process the authors use a rule that changes the acceleration of V:

$$\frac{d^2 V_{x,y}}{dt^2} = k_\nu \left(\Delta_{x,y} - \frac{dV_{x,y}}{dt} \right), \tag{2.43}$$

with k_ν a constant. Also, the change in associability is given by:

$$\frac{d\alpha_x}{dt} = k_\alpha (E_x - \alpha_x), \tag{2.44}$$

where k_α is a constant.

The authors ran simulations using 20 elements for each CS including US and context. The model was created with the intention of explaining stimulus discrimination effects so a number of simulations were run to test this. The model was able to reproduce negative patterning and biconditional discrimination. Because of normalization, when a CS is presented in compound with another, they both activate many elements with overlapping receptive fields. The normalised response is therefore not a summation of associative strength as Rescorla-Wagner would predict, and

so the model can cope well with non-linear cue competition phenomena. The authors also demonstrate that the model can produce latent inhibition because of its attentional network and its rule for associability modification. Crucially, the latent inhibition produced by the model is not due to US absence, as models like Mackintosh, 1975a would require, but simply because of a decrement in CS processing, in a manner similar to the model by McLaren and Mackintosh, 2000.

It is worth noting at this point that this model's formalism has more biological realism than it is usually seen in learning models. Its timing properties, however, remain untested.

Schmajuk

Schmajuk and colleagues have proposed several connectionist models, one of which (Buhusi and Schmajuk, 1999) is of particular interest here due to its timing features. It is a blend of two previous models: the Schmajuk-DiCarlo (S-D, Schmajuk and DiCarlo, 1992) and the Spectral Timing (STM, Grossberg and Schmajuk, 1989) models. STM is primarily a timing model and will be covered in more detail in the next chapter. It suffices to state here that its main timing engine is a cascade of traces set off by the CS, each trace with its own timing characteristics. S-D is a classical conditioning connectionist model, with a hidden-unit layer that allows it to account for certain stimulus configuration phenomena such as negative patterning that two-layer networks cannot explain.

The model builds on the idea that CSs compete not only for associative strength, the assumption behind Rescorla-Wagner, but also for US timing. Figure 2.8 shows the main components of the model. Each CS_i sets off k traces τ_{ik}, including the context CX. These traces are formed by activities x_{ik} and y_{ik} given by[1]:

$$\Delta x_{ik} = \frac{k_1}{k}(1 - x_{ik})(CS(i) + k_2 f_1(i)) - k_3 x_{ik}, \tag{2.45}$$

$$\Delta y_{ik} = k_4(1 - yik) - k_5 f_2(x_{ik}) y_{ik}, \tag{2.46}$$

where $CS(i)$ is 0 or 1 depending on CS absence or presence respectively, k_i are constants controlling decay or growth, and f_i are sigmoid functions. These two activities are combined by the following rule:

$$\tau_{ik} = y_{ik} f_2(x_{ik}). \tag{2.47}$$

Equation (2.47) produces gaussian-looking traces, with higher peaks and smaller widths near CS onset. These traces are very similar in shape to the ones postulated by the Microstimuli model (see right panel in figure 2.5).

[1]The following is an abridged description where some of the equations that do not form the core of the model, and hence are not crucial for the understanding of its main mechanisms, have been omitted. That explains why there are jumps in the numbering of the constants. For the complete set of equations see Appendix A in Buhusi and Schmajuk, 1999.

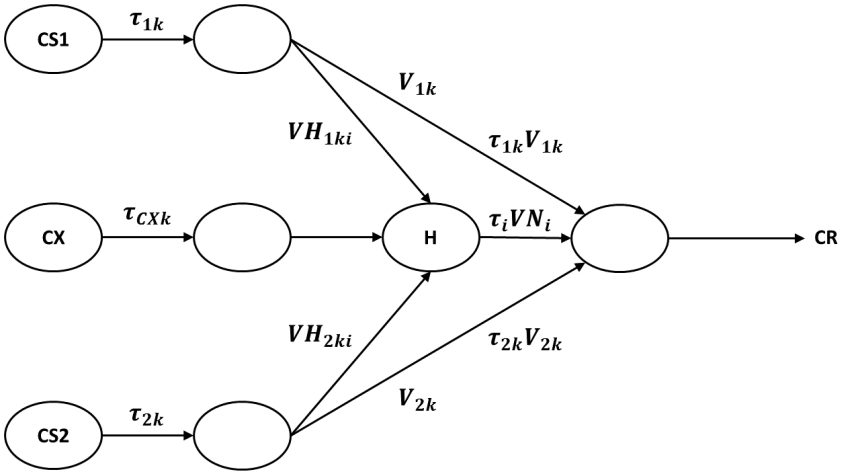


FIGURE 2.8: The S-D model. Its network architecture includes a 'hidden' unit H. This hidden unit is what allows the S-D model to explain negative patterning.

These traces are fed into the second network layer seen in figure 2.8, which in turn is connected to the hidden-unit H_j and directly to the output layer by modifiable links VH_{ikj} and V_{ik} respectively (except context CX which is only connected to the hidden unit). The learning rule for V_{ik} is

$$\Delta V_{ik} = k_{11}\tau_{ik}(US - B_{US})(1 - |V_{ik}|) \tag{2.48}$$

where B_{US} is the aggregate US prediction produced by the second and hidden layers

$$B_{US} = \sum_i \sum_k \tau_{ik} V_{ik} + \sum_j \tau_j VN_j. \tag{2.49}$$

Note the term $(1 - |V_{ik}|)$ in equation (2.48). This term prevents V_{ik}, the direct associative link between CS and output layer, from changing when $V_{ik} = \pm 1$. It also modulates the learning rate so as to make it faster when V_{ik} is further from its asymptotic value. The learning rule for the hidden-unit link VH_{ikj} is

$$\Delta VH_{ikj} = k_{13}\tau_{ik}\tau_j(US - B_{US})VN_j. \tag{2.50}$$

The output of the hidden unit is the trace τ_j computed as follows:

$$\tau_j = k_8 f_4 \left(\sum_i \sum_k \tau_{ik} VH_{ikj} \right). \tag{2.51}$$

Finally, the hidden-unit link VN_j with the output unit is updated by

$$\Delta VN_j = k_1 2\tau_j(US - B_{US})(1 - |VN_j|), \tag{2.52}$$

and model responses are produced by $CR = k_{10}B_{US}$.

As mentioned earlier, this model assumes competition between traces τ for the timing of US. Traces that peak closer in time to reinforcement receive more strength and come to dictate response topography. It is this feature that allows the model to reproduce the basic ISI effect. Another important feature is the two independent CS connections, one directly to the US representation and another indirectly via the hidden unit. According to Buhusi and Schmajuk, 1999 this independence allows the CS to act both as a simple CS (via direct link) and as an occasion setter (via hidden unit). It is this last property that allows the model to explain stimulus configuration phenomena.

Adding to the list of successful conditioning paradigms simulated in Schmajuk and DiCarlo, 1992 with its successor S-D model, which include acquisition of delay and trace conditioning, extinction, blocking, overshadowing, generalisation, and negative and positive patterning, Buhusi and Schmajuk, 1999 showed that this model can reproduce temporal and associative properties of blocking and serial feature positive discrimination.

McLaren

McLaren and Mackintosh, 2000 and McLaren and Mackintosh, 2002 set out to show that an elemental model is also capable of explaining generalization and discrimination, a class of phenomena frequently thought to require a configural stimulus representation. The inclusion of a weight decay in its associative links allows the model to also explain some timing properties.

Their starting point is a distributed network model of information processing and memory (McClelland and Rumelhart, 1985). This type of model consists of highly interconnected units, activated both by external stimuli or internal stimuli originating from other units. A CS is therefore here represented as a pattern of activation distributed over the network. As with other connectionist models, learning occurs through a change in connection weights. This change is guided by the delta rule:

$$\Delta_i = e_i - i_i,$$

where e and i are external and internal activations respectively. Learning therefore is interpreted here as the network trying to match (or equalize) external activity with internal activity. When these are equal, learning is complete.

McLaren and Mackintosh, 2000 make two main modifications on McClelland and Rumelhart, 1985. First, they include a positive feedback mechanism in the delta rule, which has a modulatory effect on the salience of external stimuli. Second, they introduce the idea of weight decay; weights change with experience but also decay with time. This modification allows the model to reproduce ITI effects.

Figure 2.9 shows the basic model architecture. Nodes are all interconnected, with weight w_{ij} representing the weight from unit i to j. Inputs are either internal or

external. The sum of internal inputs arriving at unit i from unit j is given by:

$$i_i = \sum w_{ji}\Omega_j. \tag{2.53}$$

Activation in node i is symbolized by Ω_j and changes according to:

$$\frac{d\Omega_i}{dt} = \begin{cases} E(e_i + i_i)(1 - \Omega_i) - D\Omega_i & \text{for } \Omega \geq 0 \\ E(e_i + i_i)(1 + \Omega_i) - D\Omega_i & \text{otherwise} \end{cases} \tag{2.54}$$

where D and E are decay and excitation constants respectively. The delta rule is modified to include the modulation term $r\Delta$, r a constant, as follows:

$$\Delta_i = (e_i + r\Delta_i) - i_i. \tag{2.55}$$

The modulation constant r is taken to be 0 when $\Delta_i \leq 0$. Finally, the weights are adapted according to the following system:

$$\frac{dw_{ij}}{dt} = S\Delta_i\Omega_j - Km_{ij}, \tag{2.56}$$

$$\frac{dm_{ij}}{dt} = S\Delta_i\Omega_j - Lm_{ij}, \tag{2.57}$$

with constants $L > K$. Equation (2.57) is a negative feedback that implements weight decay.

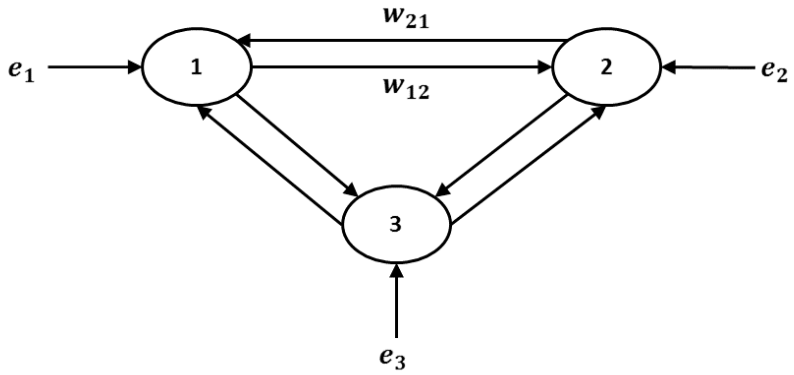


FIGURE 2.9: A three-node network used by the McLaren model. The nodes are fully interconnected by the links $w_{i,j}$, and also receive external inputs e_i.

McLaren and Mackintosh, 2000 applied the model to latent inhibition and perceptual learning, and McLaren and Mackintosh, 2002 to generalization and discrimination. The ITI effect was not tested directly but as a factor influencing latent inhibition. The model predicts that spaced (longer ITI) CS preexposure has a stronger effect in latent inhibition than massed (shorter ITI) preexposure, and the authors find evidence in the literature for just such an effect. The authors also acknowledge

a resemblance to SOP in this account of latent inhibition. Nonetheless, the model is an interesting effort in the direction of unifying distributed theories of memory and learning.

In this section we have seen the main real-time models of conditioning. Of all models analysed here, TD is the one that found the largest applicability outside experimental psychology. Within experimental psychology, Schmajuk's model is by far the most complete, but also one of the most complex which perhaps explains why it has not been widely adopted. As regards timing, none of the models here achieves the degree of sophistication found in dedicated timing theories, which will be reviewed later. For now, it suffices to note that the models above cannot offer a good account of timescale invariance, which precludes their use as hybrid timing-learning theories.

Next, I will review how learning and timing have been studied in the field of computer science.

2.1.6 Artificial Neural Networks

The theory of adaptive neural networks in computer science has been developed largely in parallel from the theories of learning in classical conditioning. Yet they show considerable overlap, with each serving as source of inspiration for the other at some point or another in time. Both theories are said to be 'loosely' inspired by real neurons and their connections. The analogy should however not be taken very far, as real neurons are much more complex than artificial neural nets, and some of the properties of neural nets (such as the weight update procedure called backpropagation) are not thought to be present in real neural networks.

The Perceptron

The first algorithm describing a neural network that could learn with the help of a teacher (supervised learning) was the *perceptron* (Rosenblatt, 1958). The perceptron is an idealized 'neuron' that is still used to perform simple binary classification. It receives excitatory or inhibitory inputs either from sensory units or other perceptrons, adds them up and outputs a value.

As with all connectionist theories, the information is stored in the connections between the perceptrons. These connections have adaptive weights which can change according to experience. In figure 2.10 the inputs to the perceptron are denoted by $x_1, x_2, ..., x_n$ and their weights[2] by $V_1, V_2, ..., V_n$. The set of inputs $x_1, x_2, ..., x_n$ is fixed

[2]It is customary in the neural network literature to denote connection weights by w. However, I will use V here to emphasize the link with computational models of classical conditioning.

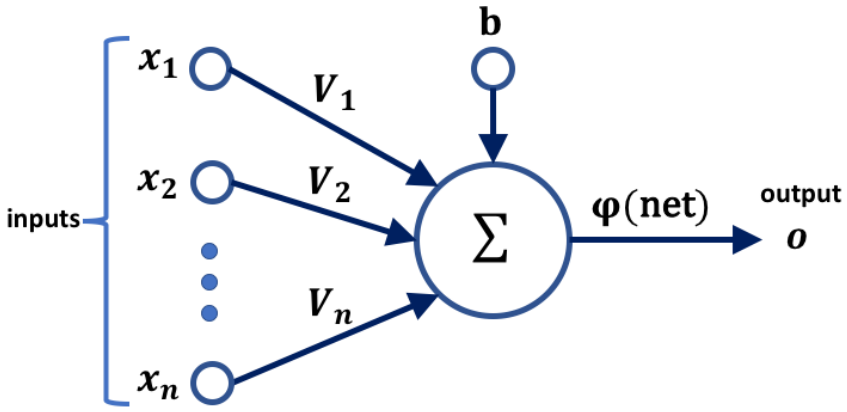


FIGURE 2.10: The perceptron. Each feature x_i of the input is directly connected to a summing unit via modifiable links V_i. A bias unit of arbitrary value b which can be adjusted manually to improve the prediction is also connected to the summing unit. Note the similarity to the diagram of figure 2.1.

for each input pattern or example, and may be regarded as the dimensions or features of the input. The computation performed by the perceptron is the linear combination or net sum of its inputs,

$$\text{net} = \sum_{i=1}^{n} V_i x_i + b,$$

where b is a bias term which is fixed and is used to increase or decrease the net value. Classification is accomplished by passing the net output through an *activation function*,

$$o = \varphi(\text{net}).$$

There are different activation functions, and two common examples are the Heaviside or step function and the sigmoid. For simple binary classification the Heaviside is a natural choice:

$$\varphi(\text{net}) = \begin{cases} 1 & \text{if net} \geq 0 \\ 0 & \text{if net} < 0 \end{cases}$$

One disadvantage of the Heaviside activation function is that it is not continuous, hence not differentiable. The backpropagation algorithm used to train modern neural networks requires a differentiable activation function, as it will be seen later. However, for training the perceptron this is not an issue, as it will become apparent in what follows.

We desire to find a set of weights $V_1, ..., V_n$ such that the perceptron can classify

correctly the largest possible number of input patterns of the form $x_1, ..., x_n$. The first algorithm for performing this adaptation was put forward by Widrow and Hoff, 1960. Let the kth input pattern be the vector $\mathbf{x}(k) = x_1, ..., x_n$, its desired classification $d(k)$, and its respective neuron output $o(k)$. A simple measure of error is given by

$$e(k) = \frac{1}{2}\left[d(k) - o(k)\right]^2. \tag{2.58}$$

The above expression is the error for a single instance of classification. In general, we would like to minimize the mean squared error

$$E\left[\sum_k^m e(k)\right] = E\left\{\frac{1}{2}\sum_k^m [d(k) - o(k)]^2\right\} \tag{2.59}$$

over a large number m of input patterns. Since we are minimizing the mean squared error, the rule for doing so is also known as the *Least Means Squares* (LMS). The minimum of the error function in (2.59) can be found by using *gradient descent*. This method works by finding the gradient of the error as a function of the weights V_i for an initial set of weights, then taking a small step in the opposite direction of the gradient, which will lead to a new set of weights from which the procedure is repeated until the new set of weights equals the previous set. Widrow and Hoff, 1960 realized that minimization of (2.59) is equivalent to consecutive minimizations of the single classification error (2.58). Also, in the case of the perceptron minimization of the error using the output of the activation function is equivalent to minimization using the net output, before it is passed through the activation function. Therefore, we need to find the gradient of

$$\mathcal{E}(\mathbf{V}) = \frac{1}{2}[d(k) - \text{net}(k)]^2 = \frac{1}{2}\left[d(k) - \sum_{i=0}^{n} V_i x_i\right]^2. \tag{2.60}$$

Note that the bias term b has been incorporated as another weight V_0 that has a constant input $x_0 = 1$. The gradient of the error is

$$\nabla\mathcal{E}(\mathbf{V}) = \left(\frac{\partial\mathcal{E}}{\partial V_1}, ..., \frac{\partial\mathcal{E}}{\partial V_n}\right), \tag{2.61}$$

and the ith partial derivative is

$$\frac{\partial\mathcal{E}}{\partial V_i} = -\left[d(k) - \mathbf{x}^T(k)\mathbf{V}(k)\right] x_i. \tag{2.62}$$

Next, we update the weights by taking a small step α in the direction opposite the gradient

$$V_i(\text{new}) = V_i(\text{old}) + \alpha\left[d(k) - \mathbf{x}^T(k)\mathbf{V}(k)\right] x_i. \tag{2.63}$$

This completes the LMS algorithm for a perceptron unit.

Note that equation (2.63) may be equivalently written as

$$\Delta V_i = \alpha \left(\lambda - \sum_{i=0}^{n} V_i x_i \right) x_i \tag{2.64}$$

which is the Rescorla-Wagner learning rule (2.2) introduced earlier in section 2.1.4. Hence, the RW rule is equivalent to the LMS. As Widrow and Hoff, 1960 noted, the LMS/RW can only find the appropriate weights if the input patterns being classified are linearly separable. This means that the two classes must be able to be separated by a hyperplane, i.e. a plane that is one dimension less than the input pattern dimension. The classic example of a linearly inseparable classification problem is the XOR or exclusive OR function. Minsky and Papert, 1988 analysed this and a number of other limitations of the perceptron. They pointed out that if an extra perceptron layer was added, the resulting network would be able to overcome these limitations. Figure 2.11 shows a network of perceptrons with a hidden layer. The hidden layer would act by recoding the input in a way that it becomes linearly separable, allowing the output to be classified correctly.

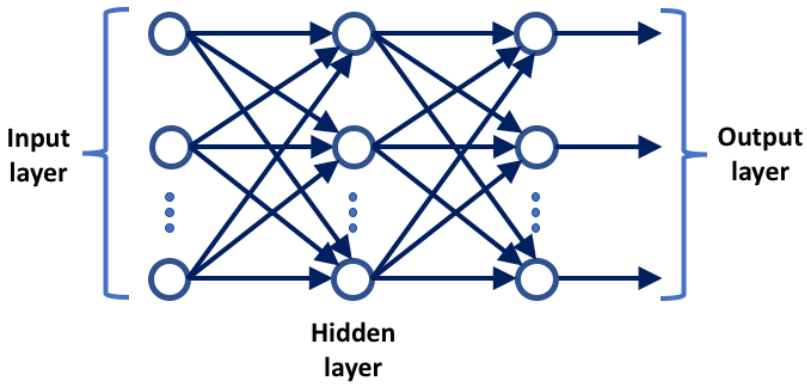


FIGURE 2.11: A fully-connected hidden layer perceptron network.

The limitations of the perceptron also constrained its use in modelling classical conditioning phenomena. The XOR problem that is unsolvable by the perceptron is readily solved by animals in a conditioning procedure called *negative patterning*. In this procedure, two different stimuli, a tone and a light for example, signal reward, whilst their compound does not. Animals can learn to respond to the tone and the light and not to the compound. Due to its equivalence to the LMS, the RW rule cannot model negative patterning.

The hidden layer solution proposed by Minsky and Papert, 1988 had the disadvantage of lacking a straightforward procedure like the LMS for adapting the

weights. Unlike the perceptron case, where the unit's output was immediately compared with the desired class, the output of the hidden layer is not apparent. A solution was finally found by Rumelhart, Hinton and Williams, 1986 with a procedure called *backpropagation*, to which I now turn.

Backpropagation

Backpropagation is a generalization of the LMS algorithm. Assume a network of the kind illustrated in figure 2.11, but with any number of hidden layers. This network is sometimes called a *multilayer perceptron* or MLP. Unlike the LMS, here the use of a nonlinear activation function is required as 'hidden units with linear activation functions provide no advantage' Rumelhart, Hinton and Williams, 1986. The network in figure 2.11 is also called a *feedforward* network, because the input pattern information propagates forward through the layers, generating the output that is compared with the desired output in the error function. The name backpropagation alludes to the information flow in the opposite direction, from the error function to the hidden layers, which allows the computation of the gradient. Once the gradient is computed, the weight correction can be performed following gradient descent, the same procedure as LMS.

Similar to the previous section, I assume the input to be an n-dimensional vector $\mathbf{x}(k) = (x_1(k), x_2(k), ..., x_n(k))$. Unlike in the perceptron, the desired output is also a vector of the same dimension $\mathbf{d}(k) = (d_1(k), d_2(k), ..., d_n(k))$. Each dimension corresponds to one neuron in the input layer, and one neuron in the output layer. The error or cost function for the kth pattern input generated by the jth output neuron is

$$\mathcal{E}_j(\mathbf{V}) = \frac{1}{2}\left[d_j(k) - o_j(k)\right]^2. \tag{2.65}$$

The *total* error of the network is obtained by summing over the n errors generated by each output neuron

$$\mathcal{E}(\mathbf{V}) = \sum_{j=1}^{n} \mathcal{E}_j(\mathbf{V}) = \frac{1}{2}\sum_{j=1}^{n}\left[d_j(k) - o_j(k)\right]^2. \tag{2.66}$$

The output of the jth neuron is

$$o_j = \varphi(\text{net}_j) = \varphi\left(\sum_{i=1}^{n} V_{i,j}(k)x_i(k)\right), \tag{2.67}$$

where φ is an activation function.

Backpropagation works first on the weights connecting the output layer, then on the weights connecting the previous layer, and so on. Following the LMS, we correct the weight $V_{i,j}$ by taking a small step in the direction opposite the partial derivative

$\frac{\partial \mathcal{E}}{\partial V_{i,j}}$. To find this partial derivative we use the chain rule:

$$\frac{\partial \mathcal{E}}{\partial V_{i,j}} = \frac{\partial \mathcal{E}}{\partial \mathcal{E}_j} \frac{\partial \mathcal{E}_j}{\partial o_j} \frac{\partial o_j}{\partial \text{net}_j} \frac{\partial \text{net}_j}{\partial V_{i,j}}. \tag{2.68}$$

Note that the partial derivative $\frac{\partial \mathcal{E}_j}{\partial o_j}$ on the right hand side of equation (2.68) is straight forward when the neuron j is in the output layer. There is only one output $o_j(k)$ and one desired output $d_j(k)$, so all we need to do is take the derivative of equation (2.65). However, if neuron j is in a hidden layer the desired output is not so clear.

Consider first the case when neuron j is in the output layer. The partial derivatives are:

$$\begin{aligned}
\frac{\partial \mathcal{E}}{\partial \mathcal{E}_j} &= 1, \\
\frac{\partial \mathcal{E}_j}{\partial o_j} &= -(d_j(k) - o_j(k)), \\
\frac{\partial o_j}{\partial \text{net}_j} &= \varphi'(\text{net}_j), \\
\frac{\partial \text{net}_j}{\partial V_{i,j}} &= x_i(k).
\end{aligned}$$

Putting all together we get

$$\frac{\partial \mathcal{E}}{\partial V_{i,j}} = -(d_j(k) - o_j(k))\varphi'(\text{net}_j)x_i(k), \tag{2.69}$$

where the dash sign indicates a differentiation with respect to the argument of the function. Also note that because neuron j is in the output layer, $x_i(k)$ represents one of the inputs coming from the layer immediately to the left, not the initial input pattern. Applying the gradient descent correction, the weight update is

$$V_{i,j}(k+1) = V_{i,j}(k) + \alpha(d_j(k) - o_j(k)).\varphi'(\text{net}_j)x_i(k) \tag{2.70}$$

Consider the case when neuron j is in a hidden layer. As mentioned before, the problem here is finding the partial derivative $\frac{\partial \mathcal{E}_j}{\partial o_j}$ in equation (2.65). The solution is to regard $\mathcal{E}_j$ as a function of the outputs of all neurons, say $L = \{l, m, ..., w\}$, that receive input from neuron j,

$$\mathcal{E}_j(o_l, o_m, ..., o_w) = \mathcal{E}_l + \mathcal{E}_m + ... + \mathcal{E}_w. \tag{2.71}$$

The equation above establishes the recursion involved in backpropagation; the error of a hidden neuron can be calculated based on the total error produced by the layer immediately to the right.

The partial derivative of this function is

$$\begin{aligned}\frac{\partial \mathcal{E}_j}{\partial o_j} &= \frac{\partial \mathcal{E}_j(o_l, o_m, ..., o_w)}{\partial o_j} \\ &= \frac{\partial \mathcal{E}_j}{\partial o_l}\frac{\partial o_l}{\partial \text{net}_l}\frac{\partial \text{net}_l}{\partial o_j} + ... + \frac{\partial \mathcal{E}_j}{\partial o_w}\frac{\partial o_w}{\partial \text{net}_w}\frac{\partial \text{net}_w}{\partial o_j} \\ &= 1\varphi'(\text{net}_l)V_{j,l} + ... + 1\varphi'(\text{net}_w)V_{j,w} \\ &= \sum_{l \in L} \varphi'(\text{net}_l)V_{j,l}. \end{aligned} \tag{2.72}$$

It is customary to define

$$\delta_j = -\frac{\partial \mathcal{E}_j}{\partial \text{net}_j} \tag{2.73}$$

as the *local gradient* for a neuron j (see Haykin, 2009, p. 131). We can then write equation (2.72) more succinctly as

$$\frac{\partial \mathcal{E}_j}{\partial o_j} = -\sum_{l \in L} \delta_l V_{j,l}, \tag{2.74}$$

where δ_l are the local gradients for the neurons in the layer to the right of neuron j.

Putting it all together we have

$$\frac{\partial \mathcal{E}}{\partial V_{i,j}} = -\sum_{l \in L} \delta_l V_{j,l} x_i. \tag{2.75}$$

And the weight update is given by taking a small step in the direction opposite the gradient:

$$V_{i,j}(k+1) = V_{i,j}(k) + \alpha \left(\sum_{l \in L} \delta_l V_{j,l} \right) x_i. \tag{2.76}$$

In summary, the complete update rule for weight $V_{i,j}$ is

$$V_{i,j}(k+1) = V_{i,j}(k) + \begin{cases} \alpha \delta_j x_i(k) \text{ if } j \text{ is in the output layer} \\ \alpha \left(\sum_{l \in L} \delta_l V_{j,l} \right) x_i(k) \text{ if } j \text{ is in a hidden layer} \end{cases} \tag{2.77}$$

The XOR problem in classical conditioning

As previously mentioned. the conditioning analogue of the XOR function is negative patterning. Schmajuk and DiCarlo, 1992 was the first to propose a conditioning model capable of reproducing negative patterning. This is the S-D (Schmajuk-DiCarlo) model introduced in section 2.1.5. We can now see that S-D is multilayer neural network with a single hidden layer. The difference from the conventional machine learning MLPs is that in S-D the input layer also makes direct connections with the output layer. The hidden unit is also known as a *configural unit*, and it is

still a widely used solution for modelling negative patterning and other configural cue phenomena (Mondragón et al., 2014).

Long Short-Term Memory Networks

Certain data patterns are defined only by their temporal information, such as the rhythmic beats of a drummer. Standard neural nets are not equipped to recognize such temporal patterns. In the case of conditioning models, when timing is taken into account it is usually through a hardcoded representation such as a tapped delay line. Such a representation assumes that time is discretized into successive states, each state acting as an independent representation that can come into association with the US. The temporal representation is therefore hardcoded and not learned. This has also the disadvantage of requiring a high number of states, as many as the time resolution and interval requires. For example, if the interval being timed is 10 minutes and we require millisecond resolution we will need 600000 states per stimulus.

A better solution would be to learn directly from the data the adequate temporal dependencies. The only machine learning framework capable of learning long-term dependencies analogous to the ones faced by an animal in a timing experiment is the Long Short-Term Memory (Hochreiter and Schmidhuber, 1997, LSTM). LSTMs are a type of Recurrent Neural Network widely used in speech recognition, machine translation, and image-to-caption generation.

An LSTM consists of a cell which receives an input, process it through multiple memory blocks and outputs a value. Each memory block is equipped with three gates that control the flow of information (see figure 2.12); they are the input gate, the forget gate and the output gate. By modulating the information flow through these gates, the cell is capable of maintaining or ending recurrent activation. The challenge faced by LSTMs is deciding which parts of the information it has already seen are relevant to predict what is coming next, hence to maintain activation, and what parts are irrelevant, hence to end it. The further in the past the relevant information is located, the bigger the challenge for an LSTM.

Here we will briefly describe one version of a LSTM that was used in a simulation of classical conditioning phenomena (Rivest, Kalaska and Bengio, 2014). In this simulation the inputs to the memory block were presence or absence of CS and US, x_{CS} and x_{US}, and a bias term b. The memory block also has a recurrent input, y_1, given by its time delayed output. Let the inputs be represented by a vector

$$\mathbf{x}_t = (b, x_{\text{CS},t}, x_{\text{US},t}, y_{1,t-1}).$$

The three gates, input, forget and output, also received the same input vector $\mathbf{x}_t$, but also received the output of the recurrent memory cell c_1. The output of each gate is

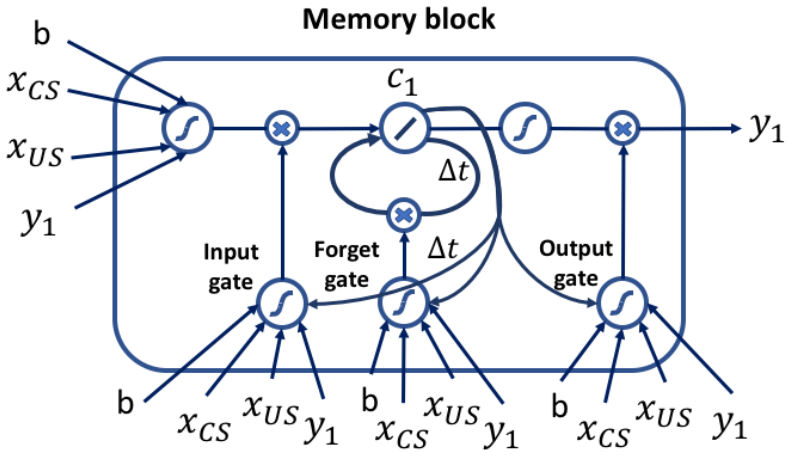


FIGURE 2.12: LSTM. The memory block is the basic unit of processing in an LSTM. Its inputs are the bias b, the CS and US values and the one time-step delayed output y. These are weighted and passed through a sigmoid function, before being multiplied by the output of the Input gate. The recurrent memory cell c_1 adds this signal to the product of its own time delayed signal and the forget gate signal. The c_1 output is passed through another sigmoid function and multiplied by the signal coming from the output gate, and the result is the memory block output y. The gates the same inputs as the memory block, and also the output of c_1 which is delayed in the case of input and forget gates but not for the output gate.

given by

$$y_{\text{in},1,t} = \text{asig}(\mathbf{w}_{\text{in},1,t}[\mathbf{x}_t, c_{1,t-1}]),$$
$$y_{\text{fgt},1,t} = \text{asig}(\mathbf{w}_{\text{in},1,t}[\mathbf{x}_t, c_{1,t-1}]),$$
$$y_{\text{out},1,t} = \text{asig}(\mathbf{w}_{\text{in},1,t}[\mathbf{x}_t, c_{1,t}]),$$

where asig() is a sigmoid function with range [0,1], $\mathbf{w}_{g,1,t}$ is the weight vector for gate g (input, forget and output) and [] is the concatenation operator. The memory cell c_1 acts as a recurrent unit with an activation output given by

$$c_{1,t} = y_{\text{in},1,t}\text{sig}(\mathbf{w}_{1,t}\mathbf{x}_t) + y_{\text{fgt},1,t}c_{1,t-1},$$

where sig() is a sigmoid function with range [-1,1], $\mathbf{w}_{1,t}$ is the input weight vector of memory block 1. The output of the memory block 1 is given by

$$y_{1,t} = y_{\text{out},1,t}\text{sig}(c_{1,t}).$$

A whole LSTM cell may have many memory blocks like the one in figure 2.12, so their outputs are weighted by the LSTM weight $\mathbf{w}_{\text{US},t}$ and passed through another

sigmoid function, yielding the LSTM output:

$$y_{\text{US},t} = \text{asig}(\mathbf{w}_{\text{US},t}[b, x_{\text{CS},t}, x_{\text{US},t}, y_{1,t}]).$$

All LSTM weights are updated according to the backpropagation rule

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \alpha \nabla_{\mathbf{w}} e^2_{\text{US},t},$$

where the error function is

$$e^2_{\text{US},t} = (y_{\text{US},t} - x_{\text{US},t+1})^2. \tag{2.78}$$

As an example consider time-series prediction. Here the LSTM must learn to predict when the next event will happen. The events are separated by an interval that consists of a fixed interval plus another variable interval. This is analogous to a delay conditioning task where the animal learns that the start of a stimulus signals that a reward is coming after a variable interval. The LSTM does not have any pre-existing time representation, so it must build one from scratch. This simple task is extremely challenging to existing LSTMs: it can take millions of trials for an LSTM to learn to predict events separated by only 50 time-steps, and even then the LSTM might not learn the task at all (Gers, Schraudolph and Schmidhuber, 2002). Animals, in contrast, can learn to associate events separated by intervals of minutes or even hours in tens or at most hundreds of trials.

Rivest, Kalaska and Bengio, 2014 coupled a LSTM with the TD model in order to see if this LSTM-TD could reproduce conditioning behaviour. The LSTM was responsible for learning to predict the stimuli at the next time step Δt, whilst the TD model learned to estimate the sum of future rewards. Here the only inputs to the LSTM were the CS and US presence or absence (1 or 0). The LSTM outputted a sigmoidal value between 0 and 1 as the prediction for the US at the next time step. The temporal resolution used (time-step) was 100 ms, and the interstimulus interval was 1 second. The LSTM-TD model was able to learn a temporal representation in the form of a ramping activity that started at CS onset and ended at US onset. It was able to reproduce delay, trace and embedded or extended conditioning (where the CS continues past the US presentation). Furthermore, in trace conditioning the LSTM learned to time the US from CS offset, and not CS onset, matching with experimental data (Buhusi and Meck, 2000).

Learning a temporal representation from scratch is still very challenging for a LSTM. The LSTM-TD model by Rivest, Kalaska and Bengio, 2014 was only tested on delay trace and embedded conditioning with a very short interstimulus interval, and took much longer to train than animals. Also, the weights of a LSTM are trained using as cost function the squared error e^2 between the prediction and the output at each time step. This cost function assigns the same error magnitude to a prediction that is off by one time-step as one that is off by n time-steps. A more adequate cost

function would assign an error value that is proportional to the size of the timing error. Finding such a cost function is not trivial, as criteria such as differentiability and real-time computability, also need to be met.

In this section we saw the equivalence between RW and the perceptron, and how perceptron can be connected together as a network to create a powerful learning architecture. We have also seen that the timing performance of the best architecture, LSTM, is still quite far behind animal performance. Computer science may thus benefit considerably from the body of theories developed for the study of psychological timing. We turn to that study next.

2.2 Timing Theories

The way animals experience the passage of time has implications to almost every cognitive capacity. Although timing was recognized by Pavlov as an important component of learning in classical conditioning, dedicated timing models did not start to appear until the 1970's. Influenced by psychophysics, these models focused on precise measurements of behaviour obtained after learning is consolidated at steady state. They also used a paradigm known as *information processing*, a computer-inspired metaphor with origins in cognitive science. It goes beyond the stimulus-response or the stimulus-outcome-response paradigms in postulating mental processes analogous to the workings of a computer.

The theoretical questions that timing models try to address concern how time is encoded (linearly, logarithmically), how it is stored in memory and how it gets translated into behaviour. The notion of a *pacemaker* is central to most theories, although some models do not use it. The models that use a pacemaker tend to treat it as a stochastic process and make assumptions on the probability distribution underlying it.

Because timing models tended to focus on steady-state behaviour, learning processes are usually ignored. A few efforts have been made to include the type of learning involved in classical and operant conditioning but these have only covered some of the most basic learning phenomena.

2.2.1 Scalar Expectancy Theory

The first, and still most influential, timing theory began with a formal description of a basic property of the timing of responses in operant avoidance procedures. Gibbon, 1971 proposed that the steady-state (or asymptotic) behaviour seen in these procedures is driven by an estimate of the time of the next shock. Crucially, this estimate is claimed by Gibbon to be a scale transform of a stochastic process, hence the *scalar property*.

Gibbon's scalar property can be formally stated as follows. Let X be the random variable representing the expected time to the next US with cdf $F_X(x)$ and pdf $f_X(x)$.

Let $Y = aX$, where a is a constant greater than zero. The cdf of Y, i.e. the cdf of the scale transform of X, can be stated in terms of the cdf of X:

$$F_Y(y) = P(Y \leq y) = P(aX \leq y) = P\left(X \leq \frac{y}{a}\right) = F_X\left(\frac{y}{a}\right),$$

hence $F_Y(y) = F_X(x)$ since $y = ax$. The pdf of Y can also be stated in terms of the pdf of X:

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{dF_X(x)}{dy} = f_X(x)\frac{dx}{dy} = \frac{f_X(x)}{a}.$$

Accordingly, we can find the mean and variance of Y in terms of X:

$$\begin{aligned} E(Y) &= \int_{-\infty}^{\infty} axf_X(x)dx = aE(X), \\ \text{Var}(Y) &= E(Y^2) - [E(Y)]^2 = E(a^2X^2) - [aE(X)]^2 \\ &= a^2E(X^2) - a^2[E(X)]^2 = a^2\text{Var}(X). \end{aligned}$$

Hence, according to the scalar property the mean and standard deviation of estimates of time are all scale transforms of the mean and standard deviation of one single stochastic process. This also implies that the coefficient of variation (CV) of the time estimate is constant. This property is independent of the actual distribution of X, the CS duration. Note also that in this account subjective time is linearly related to physical time.

Gibbon, 1977 evaluated scalar timing in a variety of operant procedures beyond shock avoidance. The scalar property is shown to hold in all cases, and its predictions shown to be more accurate than Poisson timing. This latter theory predicts that estimates of time are made based on a Poisson process, and hence the mean and variance of time estimates should be proportional to the time interval, in contrast to scalar timing which predicts proportionality of mean and standard deviation.

Among the procedures evaluated in Gibbon, 1977 the most relevant in the context of this thesis is fixed-interval (FI) reinforcement. Gibbon's FI model assumes that subjects make two time estimates, a 'global' estimate at the beginning of the trial and another 'local' running estimate. The global is the time estimate x to reinforcement and the local is a real-time estimate $x - t$ of the time remaining to reinforcement. An expectancy of reward is then formed by multiplying the inverse of the time estimate (reinforcement rate) by a constant H that depends on US properties like excitatory strength or value.

Formally, expectancy is represented as a function of time $h(t)$, global and local expectancies are respectively

$$\begin{aligned} h(0) &= \frac{H}{x}, \\ h(t) &= \frac{H}{x - t}. \end{aligned}$$

Timed responding is assumed to be controlled by a comparison between these two expectancies. The comparison is suggested to take the form of a ratio $r(t)$,

$$r(t) = \frac{h(t)}{h(0)} = \frac{x}{x-t} = \frac{1}{1-t/x}. \tag{2.79}$$

Responding starts when this ratio crosses a threshold, say $r(t) > b$. Note that in this account H, the reinforcement value, does not influence response timing. Using data from an FI procedure where the time when subjects switch from a low rate of responding to a high rate (called the break-point) was analysed (Schneider, 1969) Gibbon showed that both mean and standard deviation of break-point increase linearly with FI.

An information processing model of timing in FI was proposed in Gibbon, Church and Meck, 1984 which has remained unaltered to this day (see figure 2.13). A pacemaker marks the passage of time by emitting pulses. These pulses can be gated to an accumulator via a switch which closes at the start of a relevant interval and opens when the interval is finished. The accumulator count is kept in working memory. At the end of the interval the current count is transferred to a long-term reference memory. Behaviour is guided by the action of a comparator which actively compares the count in working memory to the one retrieved from reference memory.

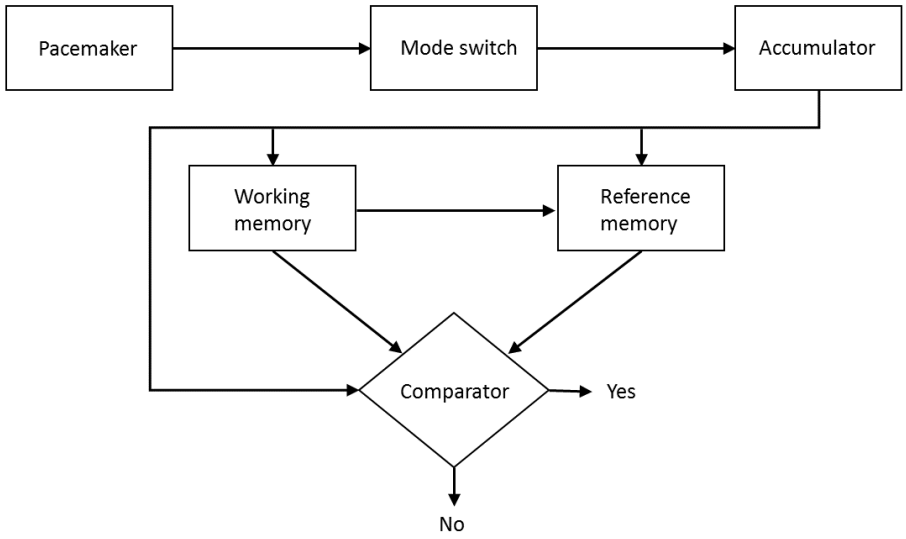


FIGURE 2.13: An information processing flowchart of Scalar Expectancy Theory. Counts from the pacemaker are accumulated in the working memory. A comparator compares the current count with a previously stored target count in reference memory. When the current count reaches the target count it triggers a response ('yes'). Adapted from Allman et al., 2014.

Pacemaker. In the original version (Gibbon, Church and Meck, 1984) assumed a scalar pacemaker, i.e. the time between pulses was considered fixed during the trial

but the pulse rate was thought to vary from trial to trial following a normal distribution. But another possibility, one that has more neurological realism is the Poisson pacemaker, which was analysed in Gibbon, 1991 and Gibbon, 1992. Under this hypothesis the pacemaker emits pulses with rate λ, assumed to be high in comparison to the interval being timed. It follows from the Poisson assumption that the time between pulses x is described by an exponential distribution with pdf $f(x) = \lambda e^{-\lambda x}$ and mean $E(x) = 1/\lambda$.

Mode switch. The switch is assumed to take some time t_1 to close and t_2 to open. The mean time τ during which pulses are counted is then taken to be $\tau = T - T_0$ where $T_0 = T_1 - T_2$, i.e. the mean of the latency differences (lowercase symbols denote random variables and uppercase their expected values).

Accumulator. When the switch is closed the number of pulses N in the accumulator grows according to $N = \lambda\tau$.

Working memory. This is taken to hold the current count in the accumulator $N = \lambda\tau$.

Reference memory. At the end of the FI the count in working memory is $N^* = \lambda\tau^*$. An important assumption is that noise in reading and writing in the reference memory is multiplicative, hence the count stored here is multiplied by a constant k^*, making it k^*N^*.

Comparator. There is more than one option of comparator to use, but they all involve the comparison between the current count n and the count retrieved from reference memory at the start of the trial n^*. Gibbon, 1991 uses the following:

$$\frac{n^* - n}{n^*} < 1 - b$$

where $0 < b < 1$ is a threshold.

Using this model Gibbon, 1992 showed that the accumulator induces a Poisson random walk, and so the time of the peak of responding τ in the peak procedure should follow a gamma distribution. This by itself does not conform to the scalar property as the mean peak time is $E(\tau) = n^*/\lambda$ and its variance $\mathrm{Var}(\tau) = n^*/\lambda^2$, with coefficient of variation γ,

$$\gamma_\tau = \frac{\sqrt{n^*/\lambda^2}}{n^*/\lambda} = \frac{1}{\sqrt{n^*}},$$

which decreases with increasing FIs and not constant. However, if a sample n^* is chosen from the reference memory at each trial, this sample will vary according to a Poisson distribution and this needs to be taken into account. This is equivalent to a Poisson random walk with a Poisson distributed threshold. In this case τ is given by a compound gamma distribution, which again does not conform to the scalar property as its mean and variance are respectively (see Gibbon, 1992) $E(\tau) = N^*/\lambda$ and $\mathrm{Var}(\tau) = 2N^*/\lambda^2$. But so far the assumption of multiplicative noise in reference memory has not been used. When that is taken into account, a biased sample is

then assumed to be retrieved from memory at each trial, bk^*n^*, which the current count n must meet and where b is the threshold and k^* the noise constant. Hence the mean count in reference memory is $\mu^* = bk^*\lambda\tau^*$. Gibbon, 1992 assumes that $s = bk^*$ is a random variable summarizing the contributions of bias in the threshold and encoding/decoding noise, with mean and variance $E(s) = BK^*$ and $\text{Var}(s) = \gamma_s^2$, where γ is the coefficient of varation. Therefore, the compound gamma distribution of τ has mean and variance

$$E(\tau) = BK^*\tau^*, \tag{2.80}$$

$$\text{Var}(\tau) = 2(BK^*/\lambda)\tau^* + (\gamma_s\tau^*)^2, \tag{2.81}$$

and CV

$$\gamma_\tau = \frac{\sqrt{\frac{2BK^*}{\lambda\tau^*} + \gamma_s^2}}{BK^*}. \tag{2.82}$$

Equation 2.82 predicts a CV that is not completely independent of FI time, but that can behave almost as a straight line in the FI range usually analysed.

SET strengths are its information processing architecture, where precise predictions can be made and tested for each separate module. Criticisms have been made on the assumptions required to make the Poisson pacemaker conform to the scalar property (see for example Staddon and Higa, 1999). But as shown above the Poisson hypothesis is not strictly necessary (although it does have neurophysiological realism) and SET's information processing architecture provides a framework for further model development.

2.2.2 Behavioral Theory of Timing

Another model that relies on a counter/accumulator and pacemaker is the Behavioral Theory of Timing (BeT, Killeen and Fetterman, 1988). As the name implies, BeT is mainly a behavioristic theory, and as such does not make assumptions about internal states or processing units (with the exception of the internal pacemaker). It is based on the observation that during timing experiments animals appear to engage on behaviors that transition from one to the other in a serial fashion (e.g. eat then drink then run around the cage, etc). As this sequence is repeated from trial to trial, reinforcement will occur more often during one particular behavior and therefore strengthen its associative link with the response (see figure 2.14). The transition from one behavior to another is probabilistic and can be modeled by a Poisson process

$$p(N(t) = n) = \frac{(t/\tau)^n e^{-t/\tau}}{n!},$$

where $p(N(t) = n)$ is the probability of being in the n-th state at time t with τ the average time between states. Note that the Poisson rate parameter is $\lambda = 1/\tau$.

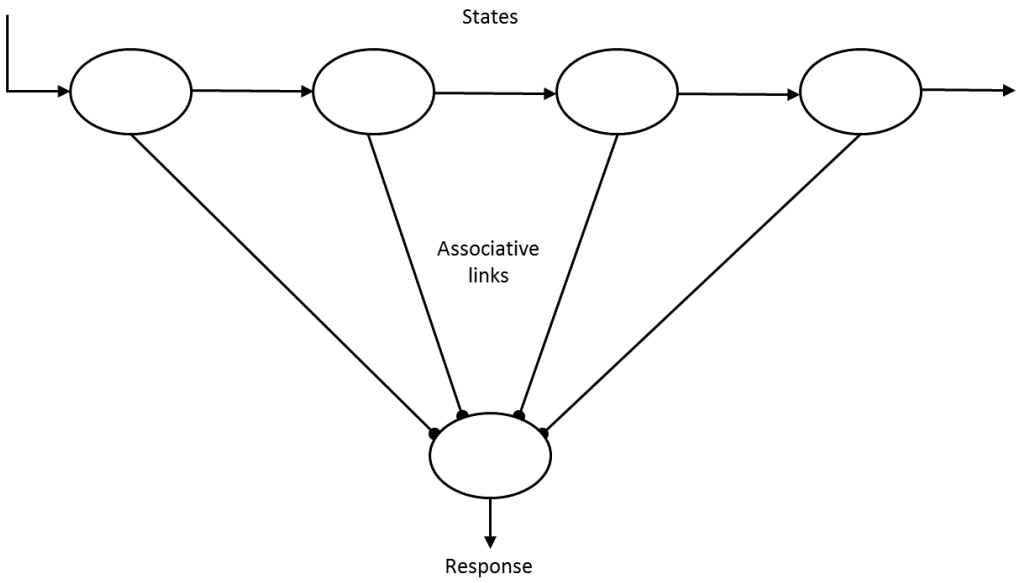


FIGURE 2.14: The basic structure of BeT and LeT. The presence of an external stimulus initiates activity over a series of internal states (top). Each internal state is connected to a response unit (bottom) via modifiable associative links. Adapted from Machado, Malheiro and Erlhagen, 2009.

One important difference between SET and BeT is that, while in SET the pacemaker rate remains constant for all time intervals, in BeT the rate is variable, changing in a manner that is inversely proportional to the time interval. Or, put in another way, the pacemaker rate is proportional to the reinforcement rate. This variable rate is what allows BeT to rely on a Poisson distribution whilst avoiding its inconvenient inability to reproduce the scalar property. To see why, consider again the peak procedure. BeT supposes that the break-point, the time subjects switch from a low to a high rate of responding, is reached after some number of pulses from the pacemaker. This means that the break-point is distributed as a gamma density with mean $\mu = n/\lambda$ and variance $\sigma^2 = n/\lambda^2$, or using the fact that $\lambda = 1/\tau$ we have $\mu = n\tau$ and $\sigma^2 = n\tau^2$. Because of BeT's assumption of proportionality between Poisson rate and reinforcement rate, say $\tau = kT$ where T is the time of reinforcement, the model produces a constant CV $(1/\sqrt{n})$.

Although BeT's main idea of a Poisson pacemaker with rate proportional to reinforcement rate continues to be cited as plausible, the model has not been studied much further. However, it served as a building block for another influential timing model: Learning to Time.

2.2.3 Learning to Time

Machado, 1997 proposed a formalization for BeT that is called Learning to Time (LeT). It goes beyond BeT in proposing associative rules to connect the behavioral states to the stimulus and responses. In its latest formulation (Machado, Malheiro

and Erlhagen, 2009), LeT departs from BeT by adopting a Gaussian instead of a Poison pacemaker. Its main assumptions are the following:

1. Behavioral states are activated serially at a rate λ per second. The rate λ is a normal random variable, sampled at the beginning of every trial, with mean μ and standard deviation σ.

2. Each state n is connected to the response by an associative link. At the end of a trial, the strength W of these links are updated as follows:

   (a) For the active state at reinforcement, n^*, the update rule is

   $$\Delta W(n^*) = \beta(1 - W(n^*)), \tag{2.83}$$

   where β is a constant.

   (b) For inactive states, $n < n^*$, the update rule is

   $$\Delta W(n) = -\frac{\alpha}{n^*} W(n), \tag{2.84}$$

   where α is a constant.

   (c) For states that did not become active during the trial, $n > n^*$, the rule is

   $$\Delta W(n) = 0. \tag{2.85}$$

3. Responses are emitted at a constant rate if the current active state has associative strength $W(n)$ greater than threshold θ.

LeT is able to account for timescale invariance and other properties of timing. Together with BeT, these two models are of particular importance since, through their modifiable associative links, they make a connection with learning theories such as RW.

2.2.4 Timing Drift-Diffusion Model

The Timing Drift Diffusion Model (TDDM, Rivest and Bengio, 2011; Simen et al., 2011) builds on the strengths of BeT and LeT, by having a variable pacemaker rate that is inversely proportional to time interval, and the Drift-Diffusion Model. The latter was originally devised as a theory of memory retrieval (Ratcliff, 1978) and has since then risen to become the standard model in decision making (Voss, Nagler and Lerche, 2013). TDDM makes the following assumptions:

1. A pacemaker Φ in the shape of a drift-diffusion process, incremented at each time step by

   $$\Delta\Phi = A \cdot \Delta t + m \cdot \sqrt{A \cdot \Delta t} \cdot \mathcal{N}(0,1), \tag{2.86}$$

   where A is the pacemaker rate and m is a constant.

2. Upon reinforcement, the rate A is adjusted to make $A = 1/T$ where T is the time of reinforcement. The adaptation rules are:

   (a) If $\Phi < 1$ at the time of reinforcement then the update rule for A is

   $$\Delta A = A \cdot \frac{1 - \Phi}{\Phi}. \tag{2.87}$$

   (b) If Φ reaches 1 before the time of reinforcement, then ΔA (but not A itself) is updated at every time-step from $\Phi = 1$ to the time of reinforcement according to

   $$\Delta A = \Delta A - (A + \Delta A)^2 \Delta t. \tag{2.88}$$

   (c) After reinforcement, A is updated by taking a percentage of the total change ΔA

   $$A_{\text{new}} = A_{\text{old}} + \alpha \Delta A. \tag{2.89}$$

TDDM is able to derive timescale invariance directly from a drift-diffusion process crossing a threshold. Just as LeT, it is capable of learning reinforcement rates. TDDM's main strenghts lie in its neurological plausibility, firm recognition in other areas of psychological research (due to its DDM architecture) and the relative simplicity with which it can explain timescale invariance. Although its link with learning theory is not as immediate as in LeT, TDDM may aid learning models by suggesting a new type of stimulus representation based on its drift-diffusion process.

2.2.5 Multiple Time Scales

The Multiple Time-Scales (MTS) model (Staddon and Higa, 1999) is different than the previous models in that it is not based on the idea of a pacemaker. Instead, it uses a decaying memory trace as its "clock". This memory trace starts at reinforcement and decays continuously until the next reinforcement, when it receives another bump in activation. As reinforcements continue to be delivered at fixed intervals, the memory trace stabilizes into a periodic curve. Thus, time can be "read" directly from this decaying trace.

Formally, MTS can be described as a cascade of integrators V_i,

$$V_i(t) = a_i V_i(t-1) + b X_i(t) \tag{2.90}$$

where $0 < a_i < 1$ and $b > 0$ are constants that determine the rate of decay and stimulus weight respectively. The V_i integrators are connected serially in such a way that the output of one is the input of the next in the cascade. At reward, the input to

the first integrator V_1 is $X_1(t) = 1$ and the input for the ith integrator is

$$X_i(t) = \begin{cases} X_{i-1}(t) - V_{i-1}(t) & \text{if } X_i(t) > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{2.91}$$

The memory trace $v(t)$ is formed by summing the activity of all integrators at each time-step:

$$v(t) = \sum_{i=1} V_i(t). \tag{2.92}$$

In order to generate timed behaviour, a response rule is added which effectively triggers responding whenever the trace v falls below threshold θ defined on every trial n by:

$$\theta(n) = v(n-1) + \xi X + \eta\epsilon(n) \tag{2.93}$$

where $v(n-1)$ is the value of the trace at the time of previous reinforcement, X is reinforcement magnitude, $\epsilon(n)$ is uniformly distributed noise, and ξ and η are constants.

MTS is particularly suited to explain timing behaviour observed in cyclic schedules of reinforcement (Luzardo, Ludvig and Rivest, 2013). It explains timescale invariance due to its periodic memory trace and dynamic noisy threshold. With regards to learning theory, and as is the case with TDDM, its value may lie in the stimulus representation it suggests: a decaying memory trace made up of the sum of activation of individual integrating units.

2.2.6 Spectral Timing Model

The Spectral Timing Model (STM, Grossberg and Schmajuk, 1989) is a neural network type of model. It postulates a neural circuit similar to the one in Figure 2.15. When a CS is presented, it activates neurons $x_1, x_2, ..., x_n$ which in turn release neurotransmitters $y_1, y_2, ..., y_n$. Learning occurs in the z_i synapses due to the co-occurrence of the US, which then activates a CR.

Formally, the activation of x_i units is given by

$$\dot{x}_i = \alpha_i \left[-Ax_i + (1 - Bx_i)CS\right] \tag{2.94}$$

where CS is a step function and α_i, A and B are constants. Neurotransmitter activity is given by

$$\dot{y}_i = C(1 - y_i) - Df(x_i)y_i \tag{2.95}$$

with C, D being constants and $f(x_i)$ a sigmoid function of activation x_i. Note that the

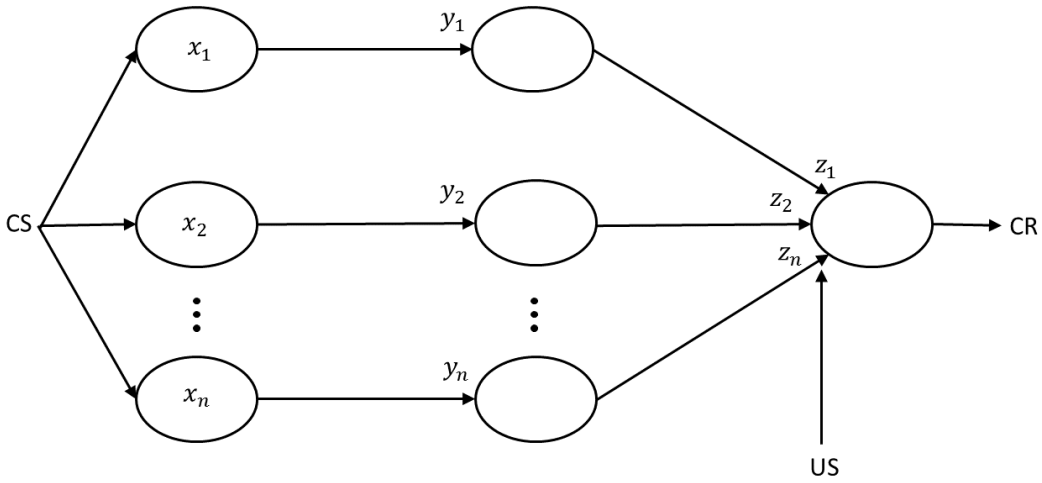


FIGURE 2.15: An example of the neural circuit in Spectral Timing Model. Each separate CS activates a neuronal unit x_i. Each of these units release neurotransmitter y_i which will act on an intermediary neuronal layer. This intermediary layer is connected to the output node via modifiable synapses z_i. Adapted from Grossberg and Schmajuk, 1989.

term $f(x_i)y_i$ in equation (2.95) is the same as equation (2.47). It produces gaussian-looking traces, with higher peaks and smaller widths near CS onset. These timing traces are very similar in shape to the ones postulated by the Microstimuli model (see right panel in figure 2.5).

The equation governing learning is

$$\dot{z}_i = Ef(x_i)y_i\,[-z_i + US] \tag{2.96}$$

where US is a step function and E a constant. Responses are generated by the sum of the product of all activations in the network:

$$R = \sum_i f(x_i)y_i z_i - F \tag{2.97}$$

where F is a threshold and R is corrected to 0 if $R < 0$. Timing arises from the product of the stimulus activation $f(x_i)$ with the neurotransmitter activity y_i, which generates a net stimulus signal $g_i(t) = f(x_i(t))y_i(t)$. This net activation resembles the Microstimuli activation on the right panel of figure 2.5.

STM generates response curves that fit response frequency curves in FI experiments, showing that it can reproduce timescale invariance. It can also handle multiple timing peaks, end effects related to US magnitude and CS intensity. Grossberg and Schmajuk, 1989 also integrate STM into a larger neural network called Gated Dipole Opponent Process (which is a model for associative learning) demonstrating

that STM may be linked with learning theory.

2.3 Existing Hybrid Models

A few attempts have been made at combining learning and timing elements into one single *hybrid* model. However, it can be difficult to establish the criteria for classifying a model as hybrid. How many conditioning and timing phenomena would this model need to explain? With this caveat in mind, two models will be covered here that have been regarded in the literature as hybrids.

2.3.1 Packet Theory

Packet theory (Kirkpatrick, 2002; Kirkpatrick and Church, 2003) is built on two distinct mechanisms, one dedicated to encode time in memory in a way that is similar to SET, and the other dedicated to controlling the shape of behaviour. The timing mechanism is subdivided into three modules:

Perception. This module marks the passage of time by defining the expected time interval e_t as the duration of the previous interval d minus the current interval t:

$$e_t = d - t. \tag{2.98}$$

Memory. Each new expectation is combined with the previous by means of a weighted sum rule:

$$\Delta E_t = \alpha(e_t - E_t), \tag{2.99}$$

and E_t is stored in memory.

Decision. Behaviour is controlled by a decision module which defines the probability of occurrence of a response *packet* by

$$p_t = nE_t^*, \tag{2.100}$$

where E_t^* is a normalised transform expectation and n is a constant.

The mechanism for controlling packet emission is based on empirically gathered statistics on response bouts, namely the distribution of responses per bout and the distribution of the interval between each response. Real-time behaviour in the model is produced by these statistics together with the probability of packet occurrence p_t.

Kirkpatrick, 2002 tested the model on three different schedules of reinforcement and on timing effects. Packet theory provided good fits to real-time behaviour in fixed, random and tandem (fixed plus random) intervals. The results were also good with timing effects, with the model fitting data on ISI, ITI and I/T effects.

Packet theory can accomplish much with relatively little. It uses mainly three free parameters. However, this comes with considerable limitations. First, responding on the model relies on empirical distributions obtained from data. This may be a problem if the model is applied to conditioning protocols that do not produce the same distributions of responses. A second limitation is that there is no mechanism to stop responding once the expected interval is finished. This presents a significant problem when explaining behaviour in the peak procedure, an important timing task. But the most severe limitation is the lack of mechanisms to deal with basic learning phenomena such as acquisition, extinction and blocking. Without these Packet theory is primarily a timing model, albeit with a wider application.

2.3.2 Modular Theory

In order to overcome its initial learning limitations, Packet theory was developed into *Modular Theory* (Guilhardi, Yi and Church, 2007). In this new, more complex, formulation, a new mechanism was included to control the strength of memories based on reinforcement, increasing its strength if reinforcement is present and decreasing it if not. With this addition, the model can handle acquisition and extinction.

Figure 2.16 is a flow diagram version of Modular theory. The three main modules are the same as Packet theory: perception, memory and decision. Here however memory is subdivided in two:

Pattern memory. This controls the expectation of time to reinforcement and is the same exponential moving average described by equation (2.99). This update is applied only if a reinforcement is delivered.

Strength Memory. This module controls the strength of memory w, and is updated by the following rule:

$$\Delta w_t = \begin{cases} \beta_e(0 - w_t) & \text{if US is absent,} \\ \beta_r(1 - w_t) & \text{if US is present,} \end{cases} \tag{2.101}$$

with β a constant that can determine different rates of update for acquisition (β_r) and extinction (β_e). Equation (3.19) is applied in real-time.

Guilhardi, Yi and Church, 2007 obtained good fits with acquisition, extinction and reacquisition in fixed intervals, demonstrating the potential of Modular theory as a learning model. Closed-form model equations were also derived, which can facilitate theoretical analysis.

Modular theory successfully overcame the learning limitations of its predecessor, however, it still does not incorporate one of the main achievements of Rescorla-Wagner theory, namely that temporal contiguity between the CS and US is not enough for learning to occur. Blocking shows that memory strength does not simply increase or decrease with reinforcement; it only does so if there is a discrepancy in the expectation to reinforcement. But given its modular nature, the present theory may easily

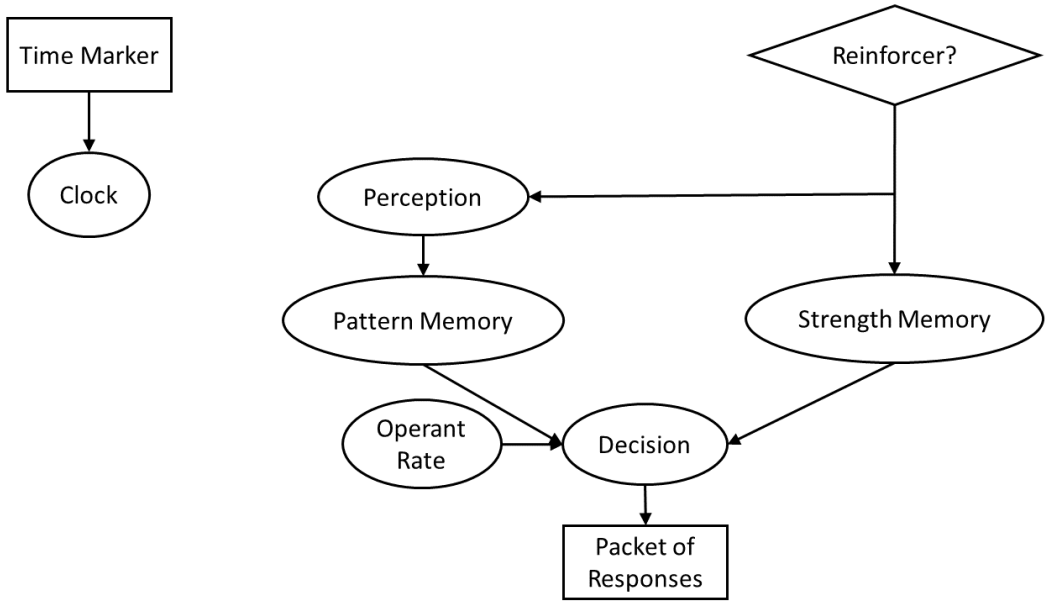


FIGURE 2.16: Flow diagram of Modular Theory. Reproduced from Guilhardi, Yi and Church, 2007.

be reformulated to accommodate a RW-type rule whilst still preserving its perception, pattern memory and decision modules.

In this section I have provided a review of the main timing models, including some hybrid models that can also reproduce a few conditioning phenomena. It may be seen from the exposition above that timing models are mainly concerned with the scalar property or timescale invariance. Very few possess any associative mechanism. Even the hybrid models, only two, are very limited in terms of their associative learning mechanisms. Both Packet Theory and Modular Theory cannot, for example, account for cue-competition phenomena, because they lack a learning rule like RW, which can deal with compound stimuli. A singe computational model that could explain both timescale invariance and cue-competition would thus be a step forward. It is important to note that some conditioning models, notably Schmajuk's (2.1.5) and TD-MS (2.1.5) intend to do just that. However, they only manage to approximate timescale invariance, and the full extent of their timing abilities has not been sufficiently explored. In the next section, I will introduce a new model that can achieve perfect timescale invariance and can explain cue-competition phenomena.

Chapter 3

Model

3.1 The Rescorla-Wagner Drift-Diffusion Model

I follow most classical conditioning theories in conceptualizing the conditioning process as the formation of an association between the internal representations of CS and US. Arguably, one of the most influential rules describing the evolution of this association through training is the Rescorla-Wagner (Rescorla and Wagner, 1972) rule. As mentioned previously, other models exist which have a similar scope to RW, both trial based (Mackintosh, 1975a; Pearce and Hall, 1980) and real-time (Buhusi and Schmajuk, 1999; McLaren and Mackintosh, 2000; McLaren and Mackintosh, 2002). However, my goal was to take advantage of TDDM's time representation, so I sought a theoretical associative framework that could incorporate such a representation. Since trial-based conditioning theories lack any time representation, they are a natural place to start. Out of those theories the RW is perhaps the simplest whilst also retaining the greatest possible explanatory power. Its basic formalism consists of the following rule for updating associative strength:

$$\Delta V_i(n) = \alpha\beta \left(\lambda - \sum_{j=1}^{l} V_j(n) x_j(n) \right) x_i(n) \tag{3.1}$$

where $V_i(n)$ denotes associative strength for CS_i at trial n, λ the asymptote of learning which is set by the US representation, $x_i(n)$ which marks the presence ($x_i = 1$) or absence ($x_i = 0$) of the i-th CS representation at trial n, $0 < \alpha < 1$ a learning rate set by the CS and $0 < \beta < 1$ a learning rate set by the US. The summation term in the equation (3.1) sums over all CSs present in the trial. The top panel of figure 3.1 shows a diagram of a basic *perceptron* for classical conditioning which serves as the architectural framework for both RW and RWDDM. The RW rule is used to update the links $V_1, ..., V_l$ that connect the CS input nodes CS_1,...,CS_l. The summation term in the RW rule is represented in the diagram as a summation unit or junction Σ, that sums the inputs it receives from the CSs $j = 1, ..., l$ present in the trial. This sum allows RW to combine (additively) the reinforcement history of each individual CS present in a compound trial. In the neural network literature, equation (3.1) is also referred to as the Widrow-Hoff rule (Widrow and Hoff, 1960) and the Least-Means-Square (LMS; Sutton, 1992). The relationship to the LMS rule is easier to see if we let

$y(n) = \sum_{j=1}^{l} V_j(n)x_j(n)$ be the output of a learning unit that aims to predict a target λ given inputs x_i by adapting the weights V_i. In classical conditioning, λ represents the maximum learning driven by a given outcome (the US), x_i is the CS and V_i the associative strength. If we let $\delta(n) = \lambda - y(n)$ be the error between output and US, equation (3.1) can be obtained with the method of gradient descent by minimizing the squared error $\delta^2(n)$ with respect to the weight V_i.

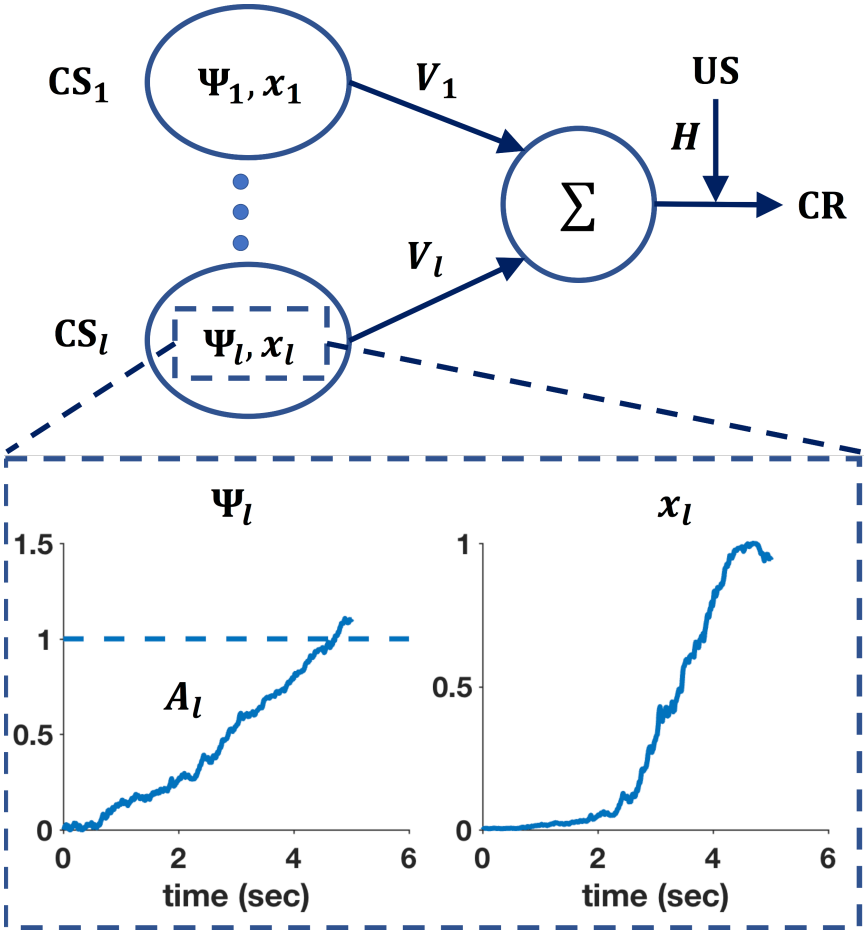


FIGURE 3.1: Connectionist diagram of RWDDM. Each CS unit is connected to a summing junction (labelled Σ) via a modifiable link V. The output of the summing junction is the CR. The US is represented as a teaching signal with a fixed weight H. Each CS unit has its own timer Ψ and representation x. The bottom panel shows a zoomed-in view of the timer Ψ_l and CS representation x_l associated with CS_l. The timer slope A_l is tuned to a 5-second CS duration.

In spite of the relative success in explaining a wide range of conditioning phenomena (for a list of successes, and failures, see Miller, Barnet and Grahame, 1995), the Rescorla-Wagner rule lacks a mechanism to account for the microstructure of real-time responding during conditioning procedures. In terms of the order of CS-US presentation conditioning procedures may be either forward (CS followed by US) or backward (US followed by CS). Two common types of forward conditioning are delay and trace. In delay conditioning the US always occurs a fixed time after CS onset. In trace conditioning the US occurs at a fixed duration after CS offset.

After sufficient training with delay or trace conditioning, responding begins some time after CS onset, increases rapidly in frequency until it reaches a maximum level where it stays until US onset (Gormezano, Kehoe and Marshall, 1983). The RW rule alone does not account for CR level as a function of time. This role is usually fulfilled by the choice of CS representation. I base my choice on a timing model called Timing Drift-Diffusion Model (TDDM, Simen et al., 2011; Rivest and Bengio, 2011; Luzardo, Ludvig and Rivest, 2013; Balcı and Simen, 2016). I chose the TDDM because it possesses a number of interesting features. It is part of a family of pacemaker based models like SET and LeT (Simen et al., 2013) which are arguably two of the most successful timing theories to date. The TDDM is a modified version of the drift-diffusion models that have been extremely successful at modelling reaction time in decision making tasks (Ratcliff, 1978; Voss, Nagler and Lerche, 2013). Evidence of climbing neural activity related to timing that resembles the TDDM has been extensively reported (Komura et al., 2001; Leon and Shadlen, 2003; Brody et al., 2003; Wittmann, 2013; Jazayeri and Shadlen, 2015). The TDDM consists of a drift-diffusion process with an adaptive drift or rate. The drift-diffusion process is defined by a continuous random walk called Wiener diffusion process. The two main components of Wiener diffusion are the drift and the normally distributed noise. The Wiener diffusion process may be visualized by imagining a two-dimensional grid with time in the horizontal axis and displacement on the vertical axis. If we imagine a purely linear and non-random walk that starts at the origin and moves up at a constant rate then the resulting walk would be a straight line and the drift would be equal to the slope of the line. With normally distributed noise, the walk becomes a random walk and it looks like a jagged curve, since at each time step there is now only a probability that the displacement will be up or down. For the purposes of timing, the slope is always positive and the random walk can be interpreted as a noisy accumulator (or timer) $\Psi(t)$, which starts at the beginning of a salient stimulus and stops (and resets) at the end. In a conditioning experiment the CS is usually the most salient stimulus in the uneventful context of the conditioning chamber, so it is well placed to serve as a time marker. When timing starts, accumulator increments are performed at each time-step according to

$$\Delta\Psi_i(t) = A_i(n) \cdot \Delta t + m \cdot \sqrt{A_i(n) \cdot \Delta t} \cdot \mathcal{N}(0,1), \tag{3.2}$$

where $A_i(n)$ is the rate (slope) of accumulation for CS_i in trial n, m is a noise factor, Δt is the time-step size and $\mathcal{N}(0,1)$ denotes a sampling from the standard normal distribution. An interval is timed by the rise in the accumulator to a certain fixed threshold, say $\Psi_i(t) = \theta$. The TDDM adjusts to new intervals by keeping the threshold fixed but adapting the rate of accumulation $A_i(n)$. The bottom left panel of figure 3.1 shows a typical trajectory (or realization) of a CS's TDDM timer after one 5-second trial.

In its original formulation (Rivest and Bengio, 2011; Simen et al., 2011) the accumulation process was not allowed to continue beyond the threshold value θ, a constraint that gave rise to two distinct rules for rate adaptation, one for when the US arrived earlier than expected and another for when it arrived later. The constraint fixing a maximum level of accumulation was driven by the neurophysiological assumption that a linear neural accumulator is not likely to continue to perform effectively beyond a certain level. The neural implementation so far proposed for TDDM's linear accumulator (Simen et al., 2011) is based on a feedback control mechanism that is tuned to balance excitation and inhibition in a neuron population. Tuning of this kind requires great computational precision, which may not be easily kept for very long in a biological system. Neurophysiology notwithstanding, we will drop that requirement here for simplicity and use instead only one update rule. We demonstrate how this single update rule can be derived by the method of gradient descent. The model learns a new interval by adapting its slope A_i so that the accumulator Ψ_i reaches the threshold value θ at the target time t^*, which may be the time of reinforcement for example. The target slope will therefore be θ/t^*. The error $\delta(n)$ between the target slope and the current slope is $\delta(n) = \theta/t^* - A_i(n)$. By minimizing the squared error $\delta^2(n)$ using gradient descent we can derive the slope update rule. The squared error as a function of A_i forms a curve. Moving in the direction opposite the slope of this curve and taking a step of size $\alpha_t/2$ we form the equation:

$$A_i(n+1) = A_i(n) - \frac{\alpha_t}{2}\frac{d\delta^2(n)}{dA_i(n)}. \tag{3.3}$$

Solving the derivative yields

$$\begin{aligned} A_i(n+1) &= A_i(n) - \frac{\alpha_t}{2} 2\delta(n)(-1) \\ &= A_i(n) + \alpha_t \left(\theta/t^* - A_i(n)\right). \end{aligned} \tag{3.4}$$

Since the organism only has access to the psychological time given by its internal timing mechanism, and not the physical time t, we assume that an internal estimate for t is formed by dividing the current pacemaker count by the current slope, $t = \Psi_i(t)/A_i(n)$. Substituting this estimate into equation (3.4) we get:

$$\begin{aligned} A_i(n+1) &= A_i(n) + \alpha_t \left(\frac{\theta A_i(n)}{\Psi_i(t^*)} - A_i(n)\right) \\ &= A_i(n) + \alpha_t A_i(n) \left(\frac{\theta}{\Psi_i(t^*)} - 1\right) \\ &= A_i(n) + \alpha_t A_i(n) \frac{(\theta - \Psi_i(t^*))}{\Psi_i(t^*)}. \end{aligned} \tag{3.5}$$

Hence, the update rule for slope A_i to be applied at target time t^* (the end of the trial or of the interval being timed) is

$$\Delta A_i(n) = \alpha_t A_i(n) \frac{(\theta - \Psi_i(t^*))}{\Psi_i(t^*)}. \tag{3.6}$$

Equation (3.6) is the slope update rule we use. Note that n above is indexing the number of occurrences of a specific interval that the timer is timing. These intervals may be the duration between CS onset and US onset (the usual 'trial' in delay conditioning for example), but they may be any other salient time interval such as CS or intertrial duration. Figure 3.2 shows timer slope adaptation during three timing scenarios: timing a novel stimulus (row 1), timing a long-short change in stimulus duration (row 3), and timing a short-long change in stimulus duration (row 5).

In the top row of figure 3.2 and throughout the paper we assume that the initial value of slope A for a novel stimulus is so low as to overestimate the stimulus duration. This overestimation will only last for a few trials, the number of which can be made arbitrarily small by choosing a high adaptation rate α_t. Alternatively, it would be possible to use a very high initial value for A so as to underestimate the stimulus duration. However this alternative does not seem neurophysiologically plausible as the brain would need to keep a pool of neurons firing very rapidly as its 'standby' timer.

In TDDM, timescale invariance arises from the nature of the noise in the accumulator. After repeated training, say in delay conditioning with a CS of fixed duration, equation (3.6) will converge to a value of A_i which will make the accumulator reach the threshold value θ at the time of stimulus offset, but only on average. In some trials the accumulator will reach the threshold sooner, in which case the organism will underestimate the stimulus duration. In other trials the accumulator will reach the threshold later, causing overestimation. The variability of this time estimate relative to the mean is given by the coefficient of variation (CV). It has been well established experimentally that the CV of time estimates in humans and other animals is approximately constant over a wide timescale (Gibbon, 1977; Gallistel and Gibbon, 2000; Allman et al., 2014). The CV of TDDM's time estimate is (see equation 3 in Luzardo et al., 2017)

$$\mathrm{CV} = \frac{m}{\sqrt{\theta}}, \tag{3.7}$$

which depends only on the choice of threshold θ and noise factor m. As these are constant, the CV of TDDM's time estimate is also constant. Note that because the timer adapts its slope gradually, if the duration of a CS is changed, CV measurements will only match the one given by equation (3.7) after the slope has finished adapting. The number of trials to adaptation will vary depending on the adaptation rate α_t.

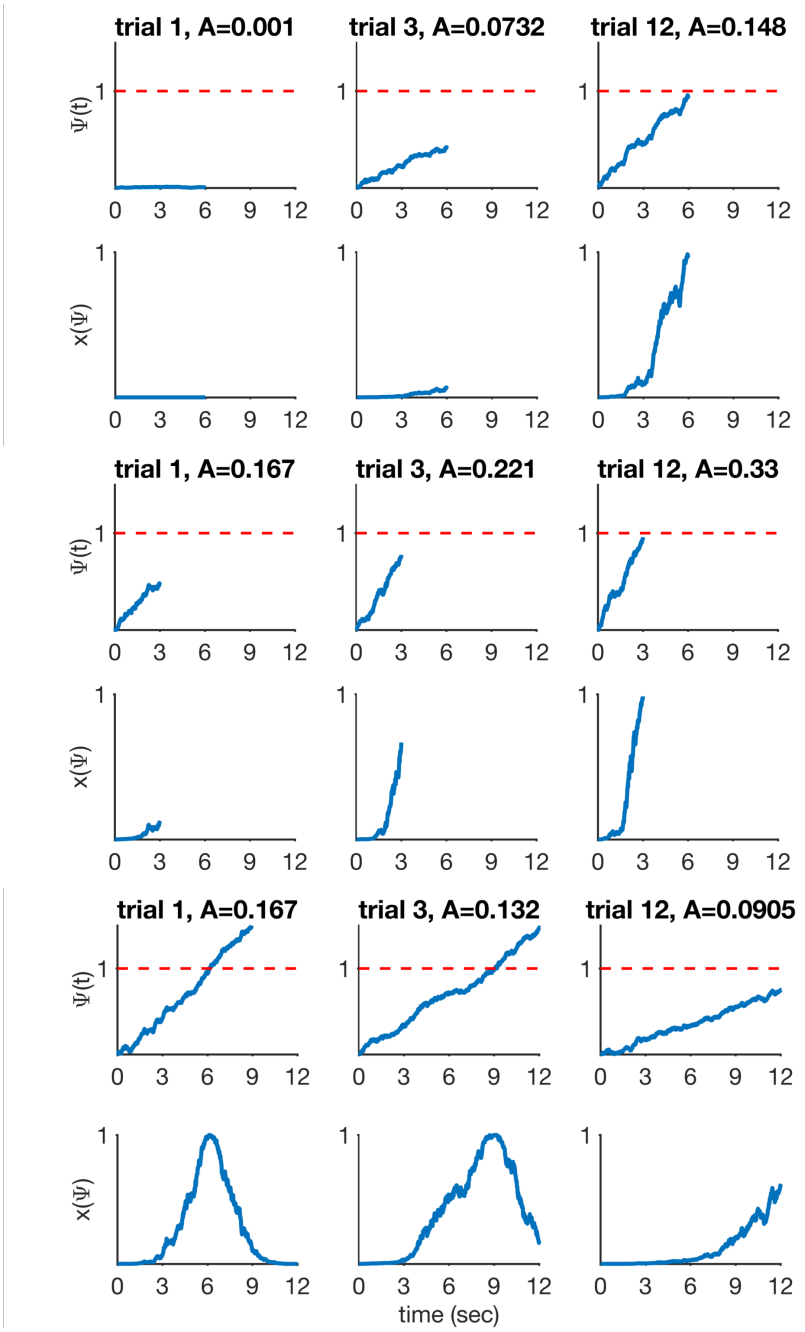


FIGURE 3.2: RWDDM timer and CS representation during three 12-trial timing scenarios. Top two rows: timing a novel 6 second stimulus. Timer starts with a low baseline slope ($A = 0.001$) on trial 1 and gradually adapts over training to reach approximately the required slope. Middle two rows: stimulus duration change from 6 to 3 seconds. Bottom two rows: stimulus duration change from 6 to 12 seconds. Parameters: $\alpha_t = 0.215$, $\theta = 1$, $\sigma = 0.25$, $m = 0.15$.

We substitute the presence representation used in the original RW model by a Gaussian radial basis function. Its input is provided by the TDDM accumulator:

$$x_i(\Psi_i) = \exp\left(-\frac{(\Psi_i(t) - \theta)^2}{2\sigma^2}\right). \tag{3.8}$$

This representation may be interpreted as the receptive field of time-sensitive neurons that read the signal coming from the accumulator neurons. Their receptive fields are tuned to the accumulator threshold value θ. The bottom right panel in figure 3.1 shows the representation for CS_l generated from the input provided by the timer on the left. Note how x_l reaches its maximum value at the same time that Ψ_l crosses the threshold at 1. Figure 3.2 shows $x(\Psi)$ adapting in the three different timing scenarios explained previously. As can be seen, x_i is a dynamic representation of CS_i that adapts to the temporal information conveyed by the stimulus. Other representation shapes could be used, like a sigmoid for example, but a Gaussian is mathematically simple and has been used before by at least one other timing model (MS-TD, Ludvig, Sutton and Kehoe, 2008).

We follow Gibbon, 1977 and Gibbon and Balsam, 1981 in assuming that time sets the asymptote of learning, λ, in equation (3.1). They were led to this hypothesis by investigating CR timing in fixed interval conditioning schedules, a type of delay conditioning. After enough training in this procedure, subjects begin responding some time after CS onset, with a slow rate at first which then increases rapidly until it reaches asymptotic level some time before reinforcement delivery. Gibbon, 1977 proposed that subjects make an estimate of time to reinforcement which is used to generate an expectancy of reinforcement. The expectancy for a particular CS_i with duration t^*, h_i, was hypothesised to be $h_i = H/t^*$, where H was a motivational parameter which was assumed to depend on the reinforcing properties of the US. The reinforcing value of the US is thus spread evenly over the CS length. It was assumed that this expectancy would be updated as time elapsed during the CS, such that $h_i(t) = H/(t^* - t)$. Hence, expectancy would increase hyperbolically until the estimated time to reinforcement $t = t^*$. Responding would reach asymptotic level when the expectancy crossed a threshold value $h_i(t) = b$ (see section 2.2.1 of this thesis for a fuller account).

Here we will not use Gibbon's concept of expectancy update. A similar role is fulfilled by the TDDM accumulator in our formalization. But we hold on to his argument that the reinforcing value of the US is spread over the CS length. Within the Rescorla-Wagner modelling framework, Gibbon's expectancy value may be interpreted as setting the asymptotic level of learning in equation (3.1), namely $\lambda = H/t^*$. Under this interpretation, λ may be said to implement hyperbolic delay discounting of rewards. Similarly to the argument used above in the derivation of the slope update rule, we use the psychological time estimate from TDDM in place of the physical time t^*, such that $t^* = \Psi_i(t^*)/A_i(n)$. The value we use is then $\lambda = \frac{HA_i(n)}{\Psi_i(t^*)}$. Another possibility would be simply $\lambda = HA_i(n)$. Both alternatives yield the same asymptotic value, but $HA_i(n)$ converges gradually (with the rate set by α_t) whilst $\frac{HA_i(n)}{\Psi_i(t^*)}$ immediately. Our version of equation (3.1) for updating associative strength

then becomes:

$$\Delta V_i(n) = \alpha_V \left(\frac{HA_i(n)}{\Psi_i(t^*)} - \sum_{j=1}^{l} V_j(n) x_j(\Psi_j) \right) x_i(\Psi_i). \tag{3.9}$$

In the trial-based RW model, equation (3.1) is applied at the end of a 'trial', which is usually taken to be the event starting at CS onset and ending at US delivery. We follow the same practice here and apply equation (3.9) at the end of a trial, i.e. at US delivery. Note that because $x_i(\Psi_i)$ is a dynamic CS representation, its activation (or strength) level at the end of the trial will vary from trial to trial, as can be seen in figure 3.2. Equation (3.9) is applied using the activation level of $x_i(\Psi_i)$ current at the end of the trial.

We assume that real-time responses to a CS_i are emitted according to the product of its associative strength $V_i(n)$ and representation $x_i(\Psi_i)$, that is, it is the output of the summing junction in figure 3.1:

$$CR_i(t) = V_i(n) x_i(\Psi_i). \tag{3.10}$$

Equations (3.2), (3.6), (3.8), (3.9), (3.10) fully define the basic model. Its six free parameters are: $0 < m < 1$ (accumulator noise), $0 < \alpha_t < 1$ (learning rate for accumulator slope), $0 < \theta < 1$ (accumulator threshold), $0 < \sigma$ (gaussian width), $0 < \alpha_V < 1$ (learning rate for associative strength), $H > 1$ (US reinforcing value).

3.2 Relationship with Other Models

Among the theories capable of providing an account of both timing and conditioning, arguably four stand out for their scope or influence. They are CSC-TD, MS-TD, LeT and MoT.

TD has been developed primarily as a learning model, without the explicit intention of addressing timing. It may be visualized as a real-time rendition of the RW rule. Its basic learning algorithm is given by[1]:

$$V_t(\mathbf{x}_t) = \sum_i w_t(i) x_t(i), \tag{3.11}$$

$$\delta_t = \lambda_t - (V_t(\mathbf{x}_{t-1}) - \gamma V_t(\mathbf{x}_t)), \tag{3.12}$$

$$\mathbf{w}_{t+1} = \mathbf{w}_t + \alpha \delta_t \mathbf{e}_t \tag{3.13}$$

where V_t is the US prediction at time t, formed by a linear combination of the weights $w(i)$ and the CS representation values $x(i)$. This update algorithm is performed at each time step, and not only at the end of a trial like RW and RWDDM. Another important difference is that equation (3.12) computes a difference between the current US value and the temporal difference between predictions. Hence, $\delta_t > 0$ if

[1] See also section 2.1.5 for a description of CSC and MS-TD.

the US is higher than this temporal difference in prediction, and $\delta_t < 0$ if the US is lower. The constant $0 < \gamma < 1$ is termed a discount factor. Equation (3.13) updates the weights for the next time step. The vector $\mathbf{e}_t$ stores *eligibility traces*, which are functions describing the activation and decay of representations $\mathbf{x}_t$. The three most common eligibility traces used are: accumulating traces, bounded accumulating and replacing traces. These three types accumulate activation in the presence of the CS and discharge slowly in its absence, the first accumulates with no upper bound, the second only until the upper bound and the third is always at the upper bound whilst the CS is present (Sutton and Barto, 1998, pp. 162-192).

The richness of TD's timing account relies on the choice of CS representation $\mathbf{x}$. The Complete Serial Compound representation (CSC, Moore, Choi and Brunzell, 1998) postulates one CS element $x(i)$ per time unit of CS duration. Each element is only switched on at its activation time unit, and then decays afterwards following its choice of eligibility trace $e(i)$ (usually an exponential decay function). This componential representation, which increases in size linearly with CS duration, should be contrasted with RWDDM's molar representation (equation (3.8)) which requires only one element. CSC may be called a time-static representation, whilst RWDDM is a time-adaptive representation, with a rule to change its structure based on a change in time (equations (3.6) and (3.8)). CSC-TD also lacks any mechanism to explain timescale invariance of the response curve, which is present in RWDDM. A modification of CSC has recently been developed, the Simultaneous and Serial Configural-Cue Compound (SSCC, Mondragón et al., 2014). SSCC-TD formalizes the idea that when multiple stimuli are presented together in time, a configural cue–a novel stimulus that is unique to the current set of present stimuli–is formed. SSCC follows on the CSC representation, but, unlike any other TD model, it allows for the representation of compounds and configurations of stimuli. Because SSCC-TD is a real-time model, it also allows for the simulation of CR timing during compounds and configurations. However, its approach to timing is still the same as CSC, i.e. it breaks down the stimuli into a series of elemental units which are activated in series. Therefore, with respect to timing only we will consider SSCC to belong to the family of CSC representations.

The Microstimuli representation (Ludvig, Sutton and Kehoe, 2008; Ludvig, Sutton and Kehoe, 2012) introduced a more realistic description of time. Unlike CSC, it uses a fixed number of elements $x(i)$ per stimulus. The ith microstimulus is given by:

$$x_t(i) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(y_t - i/m)^2}{2\sigma^2}\right) \cdot y_t \tag{3.14}$$

where m is the total number of microstimuli, y is an exponentially decaying time trace set at 1 at CS onset. It will be noted that a microstimulus is a Gaussian curve modulated by the decaying trace y_t. The set of microstimuli generated by the CS will then give rise to partially overlapping Gaussians, with decreasing heights and

increasing widths across time. The fact that only a fixed number of microstimuli are required per CS is an improvement to the potentially large numbers of elements in CSC. The MS representation tries to capture the idea that as time elapses, the stimulus leaves a more diffuse and faint impression. However, even though it is more realistic than CSC, it still lacks a mechanism to produce exact timescale invariance.

Learning to Time[2] is primarily a theory of interval timing which can also account for some aspects of conditioning. Here we will deal with its most recent version in Machado, Malheiro and Erlhagen, 2009, which differs somewhat from the earlier version in Machado, 1997. Its CS representation resembles CSC in postulating a long series of elements (or states) that span the whole stimulus duration. Unlike CSC, it transitions from state to state at a rate that varies from trial to trial, and that is normally distributed. Hence, time during a trial is represented as a noiseless linear increase from states $n = 1, 2, 3, ...$ (one per time-step) at a fixed rate. This linear time representation resembles the linear accumulator in RWDDM, except that the latter has noise built into the linear accumulator, whilst LeT assumes noise only at the intertrial level. Each state n is associated with the US via an associative link. At the end of a trial, the strength w of these links are updated as follows:

- For the active state at reinforcement, n^*, the update rule is

$$\Delta w(n^*) = \beta(1 - w(n^*)), \tag{3.15}$$

  where β is a constant.

- For inactive states, $n < n^*$, the update rule is

$$\Delta w(n) = -\frac{\alpha}{n^*} w(n), \tag{3.16}$$

  where α is a constant.

- For states that did not become active during the trial, $n > n^*$, the rule is

$$\Delta w(n) = 0. \tag{3.17}$$

Note that unlike RWDDM's associative update rule, equations (3.15) to (3.17) do not include a summation term. This places a severe limitation on the ability of LeT to deal with compound conditioned stimuli. LeT's strength lies on its being able to explain timescale invariance of the response curve. Machado, Malheiro and Erlhagen, 2009 showed that it is possible to derive timescale invariance using only the assumption of intertrial normality of state transition rate. Finally, LeT assumes that responses are emitted at a constant rate if the current active state has associative strength $w(n)$ greater than a threshold θ. The fact that responding depends on the associative strength of the current state, and that this strength only changes with US

[2] See also section 2.2.3 for a description of LeT.

associations, prevents LeT from accounting for changes in timing that are not related to US occurrence. For example, there is evidence that animals learn the timing of a preexposed CS (Bonardi, Brilot and Jennings, 2016) and are sensitive to changes in timing during extinction (Guilhardi and Church, 2006), two situations that do not involve the occurrence of a US.

Modular Theory[3] is another primarily timing theory that can also deal with some aspects of conditioning. It treats the onset of a stimulus as signalling a time expectation to reinforcement. Its time representation T is, like LeT, an accumulator that increases linearly with time t, $T = ct$, where c is a constant. When reinforcement is delivered the current reading from the accumulator is stored in what is called *pattern memory*. Pattern memory is updated at each trial n according to

$$m(n) = m(n-1) + \alpha(T^* - m(n-1)) \tag{3.18}$$

where α is a learning rate and T^* is reinforcement time. Equation (3.18) may be contrasted to (3.6) from RWDDM. The main difference is that pattern memory in MoT stores a moving exponential average of intervals, whilst the slope in RWDDM stores a moving exponential harmonic average of intervals. However, both models are similar in that they can potentially time the occurrence of any event, not only rewards. MoT's pattern memory and RWDDM's slope can be made, for example, to adapt to mark the end of stimuli that are not necessarily paired with a reward.

A stochastic threshold b is used to mark response initiation. The threshold distribution is set so as to yield timescale invariance of the response curve. Its mean, B, is a fixed proportion of the value in pattern memory, $B = km(n)$, where k is the proportionality constant, and its standard deviation is γB, where γ is the coefficient of variation of B. Hence, the coefficient of variation of the threshold, i.e. of response initiation, is constant for all intervals, which is the timescale invariance of the response curve. RWDDM derives timescale invariance of response curve from noise in the accumulator (equation (3.2)), not from the threshold.

This account of time from MoT is an instantiation of Scalar Expectancy Theory, arguably one of the most successful timing models to date. Being a purely timing theory, SET does not address associative learning directly, so it does not have a rule for changes in association between stimuli. MoT bridges this gap by adding a rule to update what is termed *strength memory*, $w(n)$. Strength memory holds the associative strength between stimulus and reinforcement. The rule consists of a linear operator:

$$\Delta w(n) = \begin{cases} \beta_e(0 - w(n-1)) & \text{if US is absent,} \\ \beta_r(1 - w(n-1)) & \text{if US is present,} \end{cases} \tag{3.19}$$

with β a constant that can determine different rates of update for acquisition (β_r)

[3] See also section 2.3.2 for a description of MoT.

and extinction (β_e). Equation (3.19) may be compared with (3.9). Note that, unlike RWDDM, equation (3.19) does not contain the summation term from RW based rules.

MoT also includes a rule for response rate that is more realistic than RWDDM's given by (3.10). It is partly derived from an empirical analysis of real-time responding in animals. We refer the interested reader to Guilhardi, Yi and Church, 2007 for a fuller description. We will only mention here that MoT generates a two-state response pattern, low and high. The transition between states is determined by the crossing of threshold B, and the high state is proportional to strength memory $w(n)$.

Other theories exist which are similar in scope to CSC-TD, MS-TD, LeT and MoT. Two notable examples are the Componential version of the Sometimes Opponent Process model (C-SOP, Brandon, Vogel and Wagner, 2003) and the Adaptive Resonance Theory - Spectral Timing Model (ART-STM Grossberg and Schmajuk, 1989). C-SOP builds a CS representation based on two sets of elements, or components, one that includes elements activated as a function of time and another whose elements are randomly activated. Associative strength for each element is updated using the standard trial-based RW rule. Simulations in Brandon, Vogel and Wagner, 2003 have demonstrated that C-SOP can produce some degree of timescale invariance. ART-STM is a neural net with an input layer and one hidden layer, which allows it to explain nonlinear conditioning phenomena (such as negative pattern) that a single-layer RW neural net cannot. It employs a CS representation that is very similar to the microstimuli used in MS-TD, so it also shows a degree of timescale invariance. Other theories could be mentioned (for two influential examples see Buhusi and Schmajuk, 1999; McLaren and Mackintosh, 2000; McLaren and Mackintosh, 2002) but we will limit the analysis to CSC-TD, MS-TD, LeT and MoT for two reasons: a) these four models collectively embody most of the conditioning and timing mechanisms used in modelling these areas, and b) our goal here is not to provide a comprehensive review, but rather focus on the mechanisms that are shared by our proposed model and the others.

Table 3.1 summarizes the main mechanisms/features of the models described above. In terms of the type of time representation, it may be observed that the models fall roughly into two categories: (a) those that employ a chain of units or states activated sequentially (CSC-TD, MS-TD, LeT), and (b) those that employ an accumulator (MoT and RWDDM). Those in category (b) may be considered more economical both computationally and biologically, as they don't require a number of units that increase with time. In terms of what the representations can time, two categories may be discerned: (a) those that time only rewards (CSC-TD, MS-TD and LeT), and (b) those that can time any stimuli (MoT and RWDDM). Models in category (b) have more flexibility to create a temporal map involving all stimuli present, including those not signalling reward. In terms of timescale invariance, the models are basically divided between those that can account for it (MS-TD, LeT, MoT and RWDDM) and the one that cannot (CSC-TD). Finally, in terms of the type of associative learning

TABLE 3.1: Summary of the main features of the models.

model	type of time representation	what it can time	timescale invariant	associative learning rule
CSC-TD	units/states, one per time step	only rewards	no	TD/RW, cue competition
MS-TD	units/states, fewer than one per time step	only rewards	approximately	TD/RW, cue competition
LeT	units/states, one per time step	only rewards	yes	linear operator, no cue competition
MoT	linear accumulator	any stimuli, not only rewards	yes	linear operator, no cue competition
RWDDM	noisy linear accumulator	any stimuli, not only rewards	yes	RW, cue competition

rule used, models are divided between those that use a RW-type rule (CSC-TD, MS-TD, RWDDM) and those that use the linear operator (LeT and MoT). The ones that use RW are wider in scope, being able to account for cue-competition phenomena, which form the core of classical conditioning.

The main innovation of RWDDM over its predecessors is the combination of a noisy linear accumulator for timing with the RW rule for associative learning. As table 3.1 shows, linear accumulator theories are the only ones in our sample of the models that can fully account for timescale invariance. But because they rely on the linear operator rule, they cannot account for cue-competition and other compound stimuli phenomena in conditioning. Therefore RWDDM extends the application of the linear accumulator to compound stimuli, covering a wider range of conditioning phenomena.

In summary, the model I propose is, to the best of my knowledge, the only one that unites the flexibility, computational economy and timescale invariance of the linear accumulator as a time representation, to the RW associative learning rule, which accounts for many more conditioning phenomena than the linear operator. In the next section I evaluate the models against a number of phenomena in conditioning and timing.

Chapter 4

Results

The long history of experimental work in classical conditioning has allowed the discovery of a rich variety of phenomena–a recent review (Alonso and Schmajuk, 2012) has catalogued approximately 87. This forces theorists to be selective when deciding which phenomena to simulate when presenting a new model. I searched the literature for phenomena that could test each feature of the model. Table 4.1 lists the main RWDDM features, together with the corresponding phenomena found in the literature that can test each.

Table 4.2 contains the design for each simulation performed with the model. The model parameters used in all simulations were kept almost constant but in some cases a few adjustments were found necessary to obtain a better agreement between model and data. I report their values in each simulation below. The time-step was the same for all simulations: $\Delta t = 10$ msec. Simulations were performed using MATLAB version R2016b. The code to generate the figures in each result section is available on Github.

4.1 Faster reacquisition

A conditioned response emerges gradually over the course of several trials where the CS signals the arrival of a US. If a measure of CR strength (such as rate or magnitude) is plotted against the number of trials, the shape and rate of this acquisition curve will depend largely on the CR and organism, but it usually follows a negatively accelerated curve (Pavlov, 1927; Kehoe and Macrae, 2002). Pavlov, 1927 believed timing of the CR would emerge only later in acquisition, through a process he described as *inhibition of delay* whereby the initial part of the CS would become inhibitory. Recent and more detailed analyses suggest that an estimate for the time to reinforcement is acquired very early in training, possibly even after one or two trials, although the expression of such estimation may not be observable until later in training (Holland, 2000; Ohyama and Mauk, 2001; Balsam, Drew and Yang, 2002; Drew et al., 2005).

If the CS no longer signals reinforcement, CR strength gradually decreases over the course of these extinction trials, until it finally disappears. If the CS is made to signal the US again, the CR returns, a process that is called reacquisition. It is a

TABLE 4.1: Model features and the experimental findings they can explain.

RWDDM feature	phenomenon for which it can account
independent update rules for time and associative strength	faster reacquisition, time change in extinction, latent inhibition and timing
RW rule for associative strength	blocking with different durations, time specificity of conditioned inhibition, inhibition in trace conditioning
intertrial variability in time estimation	compound peak procedure
asymptote of associative strength set by time	ISI effect, mixed FI
a memory that learns the rate of reinforcement	VI and FI, temporal averaging

consistent finding that reacquisition is faster than acquisition (Ricker and Bouton, 1996; Guilhardi, Yi and Church, 2007; Kehoe and Macrae, 2002, p. 185).

Learning is loosely defined as an enduring change in behaviour as a result of experience. Acquisition of a CR is the most basic demonstration that classical conditioning is a form of learning. As such, all classical conditioning models provide an account of it.

Simulations

Figure 4.1 (top left panel) shows a plot of RWDDM's associative strength as given by equation (3.9), in a simulation of acquisition and extinction. Acquisition consisted of 80 presentations of a 5-sec CS followed by reinforcement, after which there were 100 extinction trials where H was set to zero. The simulations match with experimental data from acquisition and extinction (bottom left panel of figure 4.1). The simulated acquisition curve asymptotes around the theoretical value given by setting $\Delta V(n) = 0$ in equation (3.9) and solving for V, yielding

$$V_\infty = \frac{HA_\infty}{x(\Psi_{t^*})\Psi(t^*)}, \tag{4.1}$$

which in this particular case is $V_\infty \approx 1$, since $H = 5$, $A_\infty \approx 1/5$, $\Psi_{t^*} = \Psi(t^*) \approx 1$, $x(\Psi_{t^*}) \approx 1$, where t^* is the time of reinforcement. Because $\Psi(t^*)$ is a random variable, $x(\Psi_{t^*})$ and V_∞ are also random variables and their values are reported as approximations to their expected values (but not the actual expected values).

Figure 4.1 (top middle panel) shows the adaptation of timer slope A given by equation (3.6). This equation precludes the initial value of A from being zero, so I set it to the very low value of $A(1) = 10^{-6}$. I also set the threshold $\theta = 1$, which by equation (3.6) means that $A_i(n)$ encodes the exponential moving average of the rate of reinforcement signalled by CS_i. Or, equivalently, $1/A_i(n)$ encodes the moving harmonic average of the intervals since last reinforcement during CS_i. In this

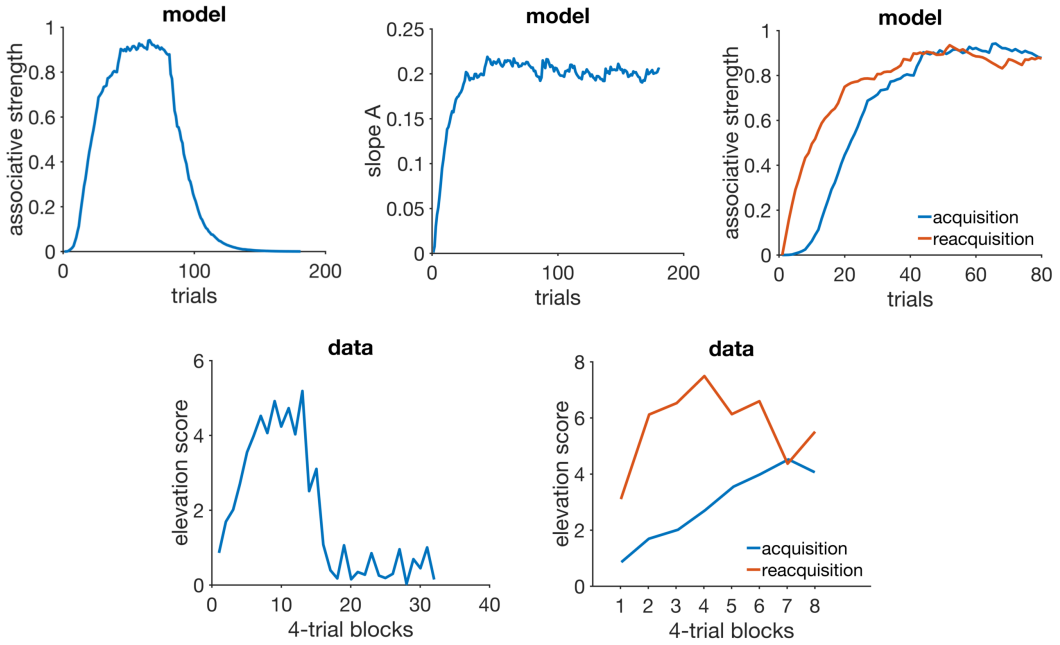


FIGURE 4.1: Acquisition and reacquisition. Top left: simulated associative strength V in acquisition and extinction. Top middle: adaptation of RWDDM slope A. CR extinction began at trial 80 but has no effect on the RWDDM slope. Top right panel: simulated V curves in acquisition and reacquisition. Bottom left panel: response strength data from an experiment in acquisition and extinction, redrawn from figure 1 in Ricker and Bouton, 1996. Bottom right panel: data from an experiment in acquisition and reacquisition, redrawn from the top panel of figure 3 in Ricker and Bouton, 1996. Model parameters: $m = 0.15$, $\theta = 1$, $\sigma = 0.3$, $\alpha_t = 0.1$, $\alpha_V = 0.1$, $H = 4$ in acquisition and $H = 0$ in extinction.

simulation, since there is only one US which is delivered always at the same time at CS offset (5000 msec), A converges to $A_\infty = 1/5000$. Note that the value of A does not decline after extinction begins at trial 80. It continues to be updated since the stimulus is still present, even if its presence no longer signals reinforcement.

The top right panel of figure 4.1 shows the acquisition and reacquisition curves using RWDDM. Reacquisition produced by the model is evidently faster than the simulated acquisition, but not as fast as the reacquisition seen in the data on the bottom left of figure 4.1.

Discussion

In RWDDM acquisition and extinction of associative strength follow from the same mechanism as RW. The only difference is the noisy stimulus representation $x(\Psi_{t^*})$, which induces noise into the acquisition curve. Changes in associative strength and timing are treated independently. In particular, the memory for time encoded by the slope A is not affected by extinction. This leads to a faster reacquisition following extinction. This is because RWDDM's time-adaptive CS representation $x(\Psi_{t^*})$ reaches

its maximum activation value right from the beginning of reacquisition, since the timer slope A is already tuned to the current CS duration (see equation (3.8)).

Modular theory (Guilhardi, Yi and Church, 2007) is another model that treats timing and associative strength separately. It postulates two memories, one for the pattern of reinforcement and another for the strength of the association between CS and US. The pattern memory stores an exponential moving average of the intervals to reinforcement which, like RWDDM, does not change with extinction. However, its strength memory $w(n)$ is updated according to the linear operator rule,

$$w(n+1) = w(n) + \beta(\lambda - w(n)) \tag{4.2}$$

which, unlike RWDDM, does not include a term for a time-adaptive CS representation. Thus, the way MoT accounts for rapid reacquisition is by using different learning rates β for acquisition and reacquisition. The same strategy may be employed with the TD and LeT models.

In summary, RWDDM explains reacquisition as the persistence of a memory for time, whilst TD, LeT and MoT explain it as a permanent change in the learning rate for associative strength.

4.2 Time change in extinction

When a previously conditioned stimulus is no longer followed by reinforcement, the conditioned response gradually decreases. An important theoretical question for hybrid timing/conditioning models concerns what happens to the timing of responses in extinction. Using the peak procedure Ohyama et al., 1999 found that although the maximum (peak) response rate decreased in extinction, peak time and sensitivity (measured by the coefficient of variation) remained virtually unchanged. Drew et al., 2004 investigated the behaviour on extinction by changing CS duration between acquisition and extinction. Groups where the CS changed to a shorter or longer duration were compared to another where the duration did not change. They found that CS duration had little effect on the rate of extinction, with all groups taking about the same number of trials to achieve CR extinction. However, when the CS used in extinction was considerably longer (4 times) than the one acquired, extinction was facilitated. Guilhardi and Church, 2006 performed a similar experiment (experiment 2) and observed that when stimulus duration is changed from acquisition to extinction, the pattern of responding during extinction gradually shifts to the new duration over extinction trials. Following the same procedure, Drew, Walsh and Balsam, 2017 also used partial reinforcement to slow down the rate of acquisition, and thus observe if response patterns really do shift gradually to the new duration. They confirmed that when CS duration was increased from acquisition to extinction, the within-trial response peak shifted gradually to the right over the course of extinction. When the CS was shortened, the results were not conclusive. Also, when CS

duration was changed from training to extinction, the speed of extinction increased, but this appeared to be explained at least in part by the shifting of response patterns.

In summary: a) peak timing and CV are not altered in extinction when using a peak procedure, b) changing the CS duration from training to extinction causes the within-trial response peak to shift to the new duration, and c) changing the CS duration in extinction can speed up extinction, but this may be due to the shifting of the response peak and not to changes in associative strength. These results pose a challenge to the models analysed here. Out of CSC-TD, MS-TD, LeT and MoT, only MoT has a mechanism that would allow it to account for time change in extinction.

Simulations

RWDDM provides an account for these findings as follows. In the case of the peak procedure, the occurrence of the longer peak trials may be considered too infrequent to cause a shift to the longer time. In this case, equation (3.6) is not applied in peak trials so RWDDM predicts that both slope A and CV will remain unaltered in extinction. In the case of a permanent change in CS duration from acquisition to extinction, the slope update rule is applied and the response peak will shift gradually to the new duration.

I have simulated RWDDM in two extinction conditions, one where the CS presented in extinction was longer than the one acquired (20 sec to 40 sec, short-long) and another where the extinction CS was shorter than the acquired CS (20 sec to 10 sec, long-short). Figure 4.2 summarizes the main results.

The panels on the left column show response strength during a trial in conditions short-long (top) and long-short (bottom). In the early stages of extinction (early) the response curves peak around the time of US arrival in acquisition (20 sec). This is more evident in the condition short-long (top left) because in the other condition (bottom left) the trial ends 10 seconds before the peak at 20 seconds occurs. Had the stimulus remained on for a full 20 seconds, the response curve in the early stages of long-short would have continued to increase until the 20 second mark. In middle and late extinction the response peak slowly shifts to the new duration in both conditions, and their heights decrease. Compare the simulated curves in the left column of figure 4.2 to the actual experimental data in the right column. The panels on the middle row of figure 4.2 show the adaptation of time estimate $1/A$ in conditions short-long (top) and long-short (bottom). They demonstrate that RWDDM adapts exactly to time change in extinction.

To investigate if the rate of acquisition changes with CS duration, I have plotted the extinction curves for each CS duration in the left panel of figure 4.3. Decreasing CS duration from acquisition to extinction slightly facilitates extinction, but increasing CS duration markedly delays extinction. However, these are only the V values, a theoretical construct that accounts for the associative strength of the stimulus as a whole. Actual behaviour measurements of extinction are based on how much response frequency changes from trial to trial. But response frequency also changes

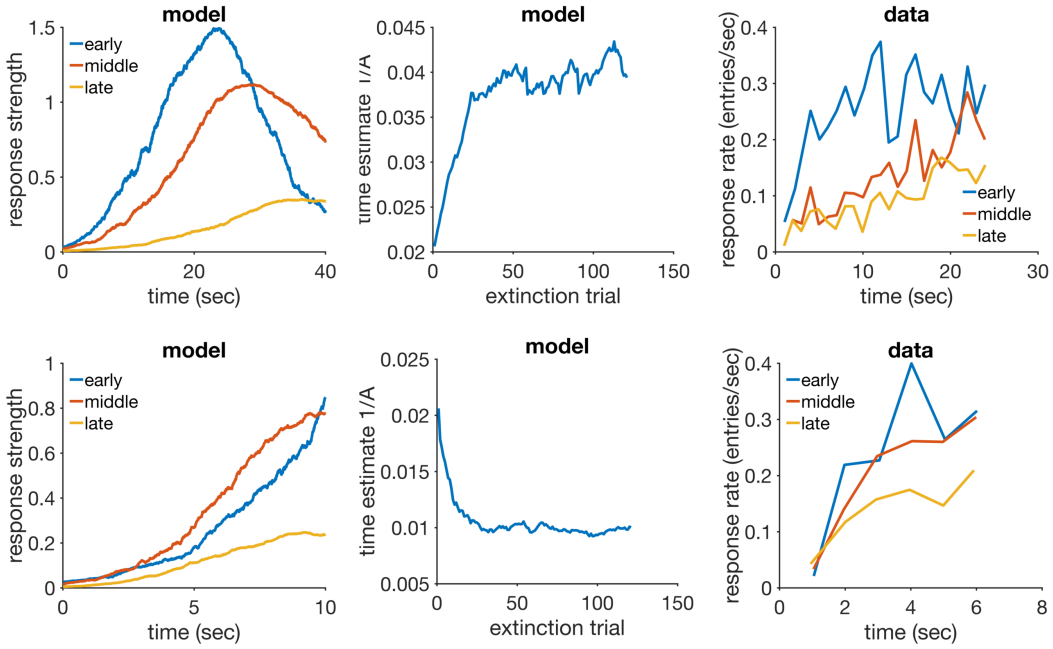


FIGURE 4.2: Time change in extinction. Left column: simulated response strength averaged over trials in extinction short-long (top) and long-short (bottom). Middle column: time estimate adaptation of the model during extinction short-long (top) and long-short (bottom). Right column: experimental data from an experiment where the CS duration changed from 12-sec in acquisition to either 24-sec (top) or 6-sec (bottom) in extinction. Data plots redrawn from figure 10 in Drew, Walsh and Balsam, 2017. Model parameters: $m = 0.25$, $\theta = 1, \sigma = 0.35, \alpha_t = 0.08, \alpha_V = 0.09, H = 30$.

within the trial. As pointed out by Drew, Walsh and Balsam, 2017, the value obtained for the rate of extinction may be affected by which portion of the CS was measured. To analyse this, Drew, Walsh and Balsam, 2017 measured response frequency only during the first 6-sec (half the duration of the CS in acquisition) of each CS duration in extinction. I have followed the same procedure and the results can be seen on the middle panel of figure 4.3. They show a marked delay on extinction when the CS duration was shortened, but not when it was lengthened. Compare these curves with the actual data analysed by Drew, Walsh and Balsam, 2017 and displayed in the rightmost panel of figure 4.3. The simulations conflict in part with the same analysis in Drew, Walsh and Balsam, 2017, which showed no delay on extinction, only facilitation in the case of extending CS duration.

Discussion

RWDDM predicts that a change in CS duration from acquisition to extinction will always cause a rescaling of the response curves in extinction. This is largely in agreement with the data. However, RWDDM seems to predict a degree of delay on extinction, whilst the data seems to point to a facilitation of extinction when the CS changes duration. When only the first half of the CS response curves are analysed,

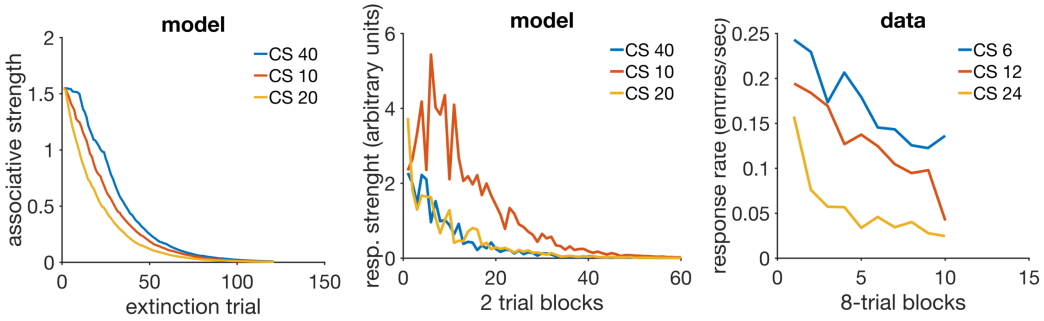


FIGURE 4.3: Extinction curves. Left panel: model V values for each CS duration in extinction. Middle panel: simulated CR values calculated only for the first 10 seconds of the CS. Each data point is calculated by summing the output of equation (3.10) over the first 10 sec of each trial, then averaging these trial values two by two, and dividing by 100 to rescale. Right panel: actual CR data for the first 6 sec of the CS in extinction, redrawn from figure 8 (C) in Drew, Walsh and Balsam, 2017

the data suggests that extending CS duration in extinction can speed up extinction, whilst RWDDM predicts that shortening CS duration will delay extinction.

RWDDM's prediction for a delay in extinction following a change in CS duration is due to the shifting of the response curve. At the beginning of extinction, a trial ends either before the CS representation has reached its peak (CS shortening) or after its peak (CS lengthening). This makes equation (3.9) update with a small value for $x(\Psi)$, resulting in a smaller update than with the higher $x(\Psi)$ value of the unchanged CS.

As mentioned above, time change in extinction is a difficult phenomenon for the current models to explain. CSC-TD does not have a mechanism to change the peak of responding when a US is not present. Neither does MS-TD or LeT. These models assume that extinction can only weaken existing links between CS and US representations. Because in these models timing usually depends on the sequential activation of these links, changing the CS duration in extinction would not alter the timing but only the magnitude of responding. RWDDM explains time change in extinction because its rule for time adaptation is independent of a change in associative strength. Thus, when the duration changes in extinction, RWDDM's accumulator slope tracks this change, whilst associative strength decays as a function of US absence. Regarding the extinction facilitation caused by a change in CS duration, none of the models analysed here currently have a mechanism to explain this either.

It would be possible to allow the average rate of state transition in LeT to vary as a function of CS duration, which would cause timing to adapt to the new time in extinction. However, in its latest formulation (Machado, Malheiro and Erlhagen, 2009) LeT relies on a fixed average rate of state transition to explain timescale invariance. Thus, if the rate is made to change as a function of CS duration, this would break timescale invariance.

As for MS-TD, one interesting modification that would likely allow it to explain

time change in extinction is to make the microstimuli themselves time-adaptive. Like RWDDM's time-adaptive CS representation, the microstimuli could be made to 'stretch' or 'compress' when stimulus duration shortens or lengthens.

Modular Theory is likely to account for time change in extinction, since its pattern memory for time could be made to update even in extinction. That would shift the response pattern to the new time whilst strength memory, which depends only on US presentation, would decay.

4.3 Latent inhibition and timing

When a subject is exposed to repeated and non-reinforced presentations of a stimulus it has never encountered before, this procedure is called preexposure. If reinforcement is subsequently paired with the preexposed CS, the initial rate of CR acquisition is usually lower compared to acquisition to a nonpreexposed stimulus, a phenomenon called latent inhibition (Lubow and Moore, 1959). The asymptotic level of conditioning, however, is not normally affected by preexposure (Lubow, 1989). Latent inhibition is an important representative of a class of phenomena involving latent effects. Collectively, these phenomena demonstrate that something is learned about the stimulus even when it does not signal reinforcement. Therefore, latent inhibition cannot be accounted by the Rescorla-Wagner model, since the theory only applies when there are changes in associative strength.

A question relevant for real-time conditioning models is what happens to timing when a preexposed stimulus is conditioned. To answer this question, Bonardi, Brilot and Jennings, 2016 used CSs of variable and fixed durations (the variable duration CS had the same mean as the duration of the fixed CS) to vary the temporal conditions between preexposure and conditioning phases. Latent inhibition was observed even when the temporal information from the two phases was different. Crucially, timing, as measured by the response gradient within a trial, appeared to improve in the preexposed CS even when the temporal information was different between the two phases.

As alluded to above, latent inhibition cannot be accounted by the associative learning update rule used in RWDDM, the Rescorla-Wagner. However, I show here that RWDDM is compatible with the Pearce-Hall rule (Pearce and Hall, 1980; Pearce, Kaye and Hall, 1982), one of the most widely used models for explaining latent inhibition and other latent learning effects. I demonstrate that this modification maintains the basic framework of the RWDDM, and that it can account for latent inhibition and improved timing with preexposure. None of the other models analysed here can account for latent inhibition without modifications. Improved timing with preexposure could be accounted by Modular Theory, but not by the the current version of the other models.

Simulations

The Pearce-Hall model is basically a rule for adapting the learning rate α_V based on the error δ between the predicted US outcome and the actual US outcome. It was originally formulated by Pearce and Hall, 1980 and updated by Pearce, Kaye and Hall, 1982. I have maintained equation (3.9) for associative strength, but changed α_V on every trial n according to

$$\alpha_V(n+1) = \alpha_V(n) + \gamma(|\delta| - \alpha_V(n)), \tag{4.3}$$

$$\delta = \left(\frac{HA(n)}{\Psi(t^*)} - V(n)x(\Psi) \right) \tag{4.4}$$

where $0 < \gamma < 1$ is a parameter that sets the rate of learning rate adaptation. Equation (4.3) is basically the Pearce-Hall rule, except that instead of using 1 as the asymptote of learning I use $\frac{HA(n)}{\Psi(t^*)}$.

I simulated latent inhibition with a 5-sec CS. Preexposure consisted of 80 trials of the CS without reinforcement ($H = 0$). The preexposed CS was then reinforced for 250 trials. Figure 4.4 (top left panel) compares the acquisition curves for the preexposed CS and a control CS in the reinforced trials. The preexposed CS acquisition curve increases at a lower rate than the control CS, the latent inhibition effect (see data from a corresponding experiment at the bottom left panel of figure 4.4).

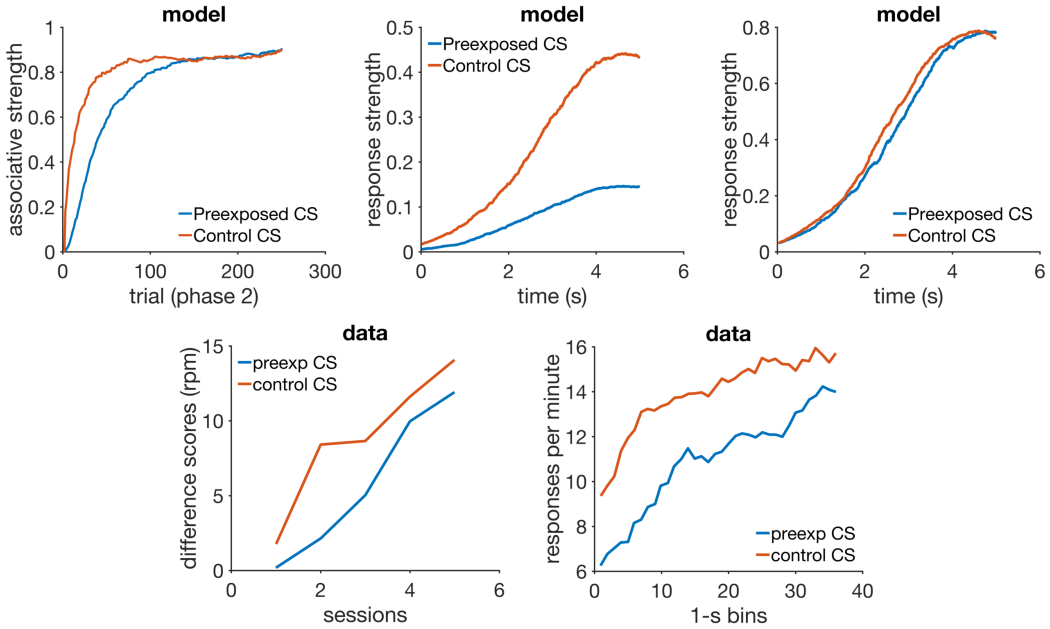


FIGURE 4.4: Latent Inhibition. Top row: simulated associative strength in latent inhibition (left), simulated CR averaged over the first 30 trials of conditioning phase (middle), and simulated CR averaged over the last 30 trials of conditioning phase (right). Bottom row: acquisition curves from an actual experiment in latent inhibition (left), and response rate data during the CS (right). Data plots redrawn from figures 1 and 2 respectively in Bonardi, Brilot and Jennings, 2016. Model parameters: $\alpha_t = 0.1$, $\alpha_V = 0.08$, $\mu = 1$, $\sigma = [0.6 - 0.35]$, $m = 0.2$, $H = 4$, $\alpha_{PH} = 0.4$, $\gamma = 0.03$.

Improved timing with preexposure follows directly from the fact that RWDDM adapts its accumulator slope A to the CS duration during preexposure. However, our choice of a Gaussian for stimulus representation does not allow for this change to become visible. Bonardi, Brilot and Jennings, 2016 demonstrated improved timing by showing that the slope of the response curve from the preexposed CS was higher in the first few trials of acquisition than the one from the control CS (see bottom right panel of figure 4.4). In general, animal response curves tend to be quite flat during the beginning of acquisition. There is evidence that the response curves appear to change from negatively accelerated to a sigmoidal shape over the course of training (see figure 1 in Meck and Church, 1984, for an example). This means that in the early stages of acquisition, within-trial response frequency increases very early in the trial and then stays at a constant level until the end. As training progresses, the increase in frequency moves slowly to the right, giving rise to the sigmoidal shape that peaks just before the end of the trial. In these cases a higher slope of the response curve would indicate improved timing. But in our model the curves are sigmoidal from start of acquisition, so they will always peak at the end of the trial, even if the timer slope has not adapted to the interval yet, as is the case with a novel stimulus. Therefore, during the acquisition phase of latent inhibition, RWDDM predicts that only the peaks of the response curves will gradually increase over the trials. Because of the learning decrement caused by preexposure, the peak of the control CS will increase faster than the preexposed CS, as the top middle panel of figure 4.4 demonstrates. The response curve of the control CS will have a higher slope than the preexposed CS, even though the preexposed CS's timer rate has been adapted to its duration. Hence, the improved timing found in the data is explained by adaptation of RWDDM's timer slope, but RWDDM's CS representation cannot make this visible.

I have tried adding an adaptable σ in equation (3.8) so as to decrease the width of the gaussian curve gradually over trials. I chose a simple linear operator rule to adapt the Gaussian width:

$$\sigma(n+1) = \sigma(n) + \alpha_\sigma(0.35 - \sigma(n)), \tag{4.5}$$

and set $\sigma(1) = 0.6$ and $\alpha_\sigma = 0.025$.

Figure 4.4 (top middle panel) shows response strength of control and preexposed CSs averaged over the first 30 trials of the conditioning phase. The preexposed CS already shows a clear sigmoidal shape, whilst the control is slightly wider and linear. But the effect is too small to be able to account for the one seen in the data from Bonardi, Brilot and Jennings, 2016. Towards the end of the conditioning phase the two curves converge (figure 4.4, top right panel).

Discussion

The simulations show that the model can account for latent inhibition adequately if the Pearce-Hall rule is used (in which case the model would be more appropriately named PHDDM). The PH rule adapts the learning rate α_V based on the level of associative learning between stimulus and reward. When the subject encounters a novel stimulus, it is assumed that α_V has some non-zero starting value α_V^{novel}, which allows learning in equation (3.9) to take place. If this novel stimulus does not signal reward, as is the case in the preexposure phase of latent inhibition, $\sigma = 0$ and equation (4.3) will simply decay the value of the learning rate across trials until it reaches zero. If at this point the stimulus begins to be followed by reward, $\sigma > 0$ and equation (4.3) will begin to raise the value of the learning rate, which in turn will allow equation (3.9) to begin increasing the value of V. Since the increase in the value of the learning rate is gradual, determined by the rate γ, there will be a number of trials in the beginning of the conditioning phase where $\alpha_V < \alpha_V^{\text{novel}}$, which leads to the initial impairment in the learning curve when compared to the learning curve of a non-preexposed CS, as seen in the top left panel of figure 4.4.

The separate rule for time adaptation allows the model to account for improved timing after preexposure, but the model cannot make this effect visible even if we allow for Gaussian width adaptation. In view of this it seems more likely that a two-state CS representation may be a better solution. As mentioned above, figure 1 in Meck and Church, 1984 suggests that during the initial stages of training a CS representation may be modelled by the following leaky integrator

$$x_i(t+1) = x_i(t) + \frac{1}{\tau}(I_i - x_i(t)) \tag{4.6}$$

where I_i is the indicator function marking the presence of CS_i, and τ a time constant. In the latter stages of training, when timing is expressed, the organism switches to the Gaussian representation given by equation (3.8). When the switch between representations is made and how abruptly remains to be investigated.

Latent inhibition cannot be accounted by any of the other models analysed here without modifications. Also, models that rely on the US for time adaptation, like CSC-TD, MS-TD and LeT, cannot account for improved timing by preexposure. Modular Theory is the only one that can time any stimulus like RWDDM, so it could account for the improved timing. But it would also need a modification like (4.3) to adapt its learning rate to account for latent inhibition.

4.4 Blocking with different durations

Arguably, the most important compound conditioning phenomenon is blocking. It is part of a class of cue competition and compound phenomena discovered in the late 1960s which challenged the view that conditioning was driven by the pairing, or contiguity, of CS-US. These results suggested that conditioning with compound

stimuli was influenced by the reinforcement histories of the elements forming the compound (Rescorla, 1988; Gallistel and Gibbon, 2001). This led to the development of a new generation of models that could account for those findings (Rescorla and Wagner, 1972; Mackintosh, 1975a; Pearce and Hall, 1980). The rule I use, the Rescorla-Wagner, provides an explanation for blocking that is based on the summation term in equation (3.1).

In a blocking procedure a CS is first paired with a US in phase 1 of training. During phase 2 a novel CS is presented in compound with phase 1 CS and paired with the US for just a few trials. Subsequently, when tested alone the novel CS elicits less responding than if it had been trained in compound with another novel stimulus (Kamin, 1968). The previously reinforced CS is said to block the novel CS. The temporal information encoded by each CS has an effect on the amount of blocking observed. Schreurs and Westbrook, 1982 varied the ISI in the pre-training and compound phases, and observed less blocking when the durations were different in both phases than when they were the same. Barnet, Grahame and Miller, 1993 performed a similar experiment but with forward and simultaneous conditioning varying between phases, and also found that blocking was stronger when blocked and blocking CSs had the same temporal history. Jennings and Kirkpatrick, 2006 used compounds where the elements had different durations. They observed that a long blocking CS could block a co-terminating short Cs, but a short blocking CS failed to block a co-terminating long CS (see rows 1 and 3 in figure 4.5). Amundson and Miller, 2008 performed four blocking experiments using trace conditioning. In two of them the blocking CS trace duration changed between phases, and blocking was not observed. In the other two experiments the trace duration was held fixed between phases, and the blocking and blocked CSs were presented serially and not in a compound (see rows 2 and 4 of figure 4.5). Blocking was observed when the blocking CS followed the blocked CS, but not in the reverse condition.

The studies reviewed above appear to show that changing the ISI of the blocking CS between phases may attenuate blocking. Another finding is the apparent asymmetry of blocking when the ISI of the blocking CS is kept constant between phases. Rows 1 and 2 of figure 4.5 suggest that a long blocking ISI can block a short blocked ISI. Rows 3 and 4 suggest that a short blocking ISI does not block a long blocked ISI.

As mentioned above, RWDDM can account for blocking because it uses the RW rule. The summation term in equation (3.1) formalizes the widely held view that a given US can only confer a limited amount of associative strength which CSs must compete for. Different theories exist that take other approaches to blocking (see for example Mackintosh, 1975a; Harris, 2006; Stout and Miller, 2007) but among the ones analysed here (for their ability to handle timing also) only CSC-TD and MS-TD are equipped to deal with it. I show next that RWDDM can account for the blocking of a short CS by a long CS, and that by making the reasonable assumption of second-order conditioning it can also account for the lack of blocking of a long CS by a short CS. CSC-TD and MS-TD are also capable of providing an account of both blocking

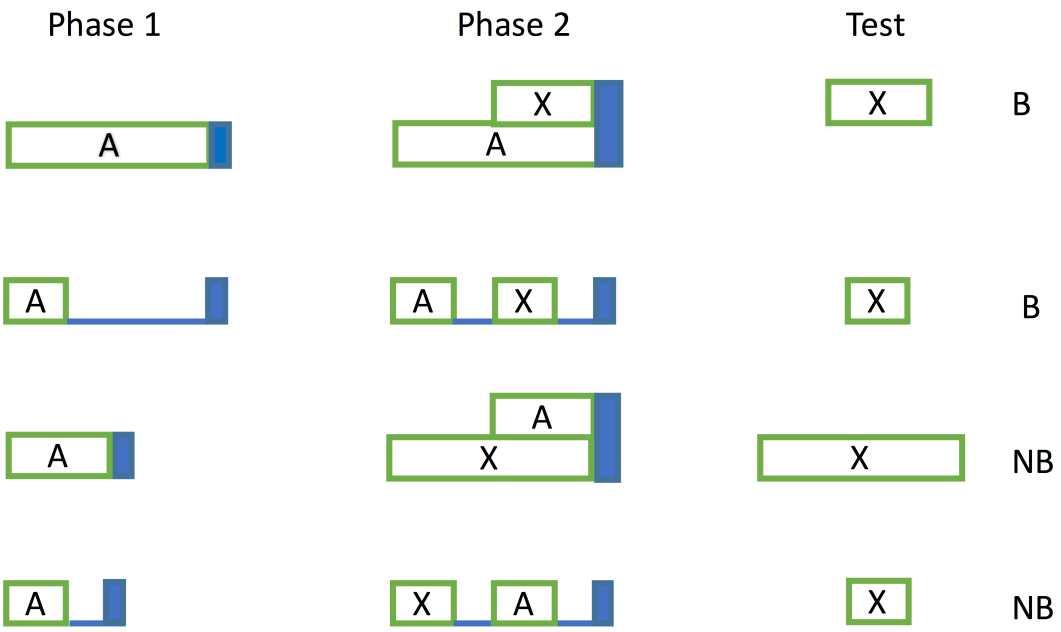


FIGURE 4.5: Experimental designs from two blocking experiments. CS X was blocked (B) in rows 1 and 2, and not blocked (NB) in rows 3 and 4. Blue bar indicates US presence.

conditions.

Simulations

Because RWDDM is based on the RW rule, it produces virtually the same results as the latter when the CSs have the same duration. Our interest here is to test whether it can reproduce the finding that a long CS can block a shorter CS but a shorter CS does not block a longer one. I performed a simulation following the design in rows 1 and 3 of figure 4.5. In the first phase a CSA (blocking CS) of duration either 10 or 15 seconds was followed by reinforcement until its associative strength V reached asymptote. In phase 2 CSA was joined with a CSX (blocked CS), of either 15 or 10 seconds, in a coterminating compound and followed by US. The top left panel of figure 4.6 shows the acquisition of associative strength for CSX and its control during phase 2 for the condition CSA-15sec and CSX-10sec. A considerable amount of blocking is observed, matching with the data (bottom left panel).

The top right panel of figure 4.6 shows the results for condition CSA-10sec and CSX-15sec. In this condition the model diverges considerably from the data (bottom right panel) and predicts that CSX should actually become inhibitory.

Discussion

The blocking and inhibition seen in figure 4.6 is a result of a discrepancy in the asymptote of learning between the CSs. After phase 1, CSA has associative strength

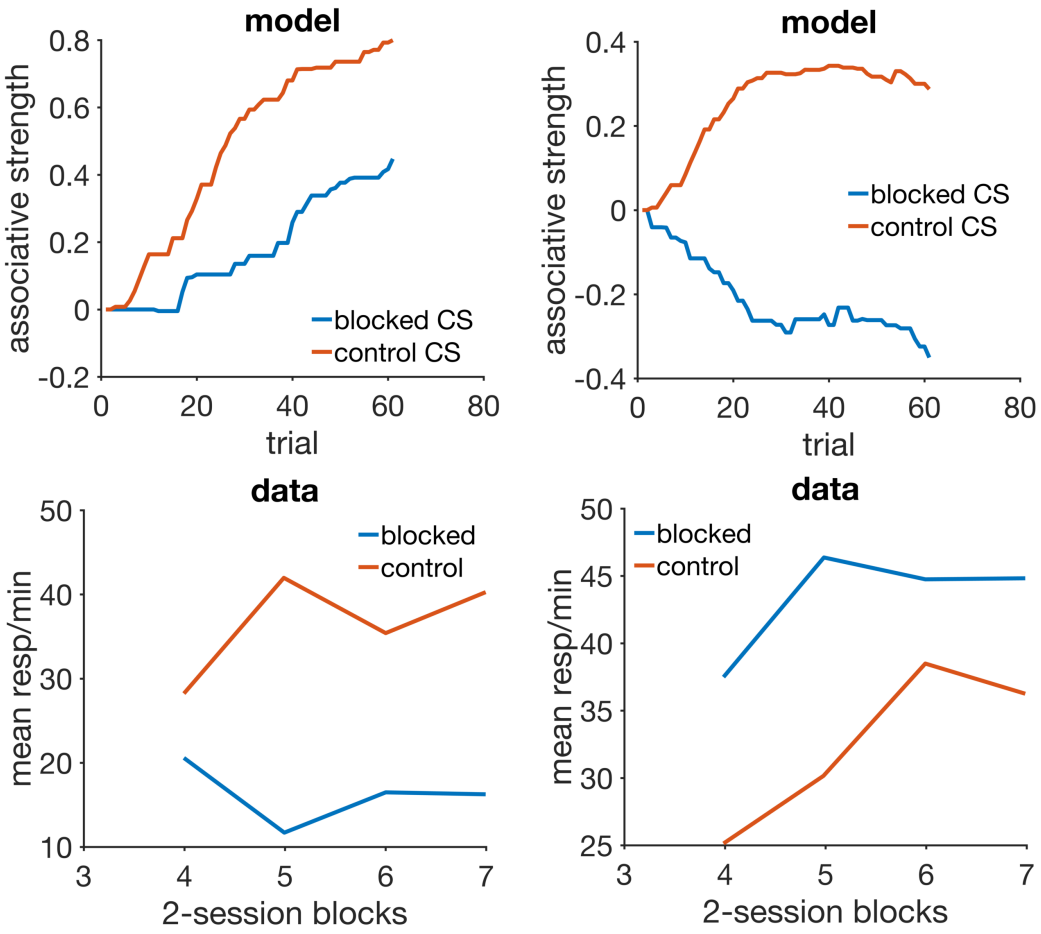


FIGURE 4.6: Blocking with different durations. Left column: simulation (top) with a 15 sec blocking CS and 10 sec blocked CS, and animal data (bottom) from an experiment with the same design. Right column: simulation (top) with a 10 sec blocking CS and 15 sec blocked CS, and animal data (bottom) from an experiment with the same design. Data panels redrawn from the top right panel in figure 5 in Jennings and Kirkpatrick, 2006. Model parameters: $\alpha_t = 0.2$, $\alpha_V = 0.1$, $\mu = 1$, $\sigma = 0.35$, $m = 0.2$, $H = 10$.

$V_A \approx HA_A$. During phase 2, CSX's associative strength changes according to:

$$\begin{aligned}\Delta V_X &\approx \alpha(HA_X - (V_A + V_X)) \\ &= \alpha(HA_X - (HA_A + V_X)) \\ &= \alpha(H(A_X - A_A) - V_X)\end{aligned}$$

and since $(A_X - A_A) < 0$, V_X becomes negative.

However, it could be argued that the short CSA becomes a secondary reinforcer which is signalled by the onset of the long CSX. In this case, the onset of CSX would serve as the time marker for the onset of CSA, and not for the onset of US. Hence, during the first 5 seconds of CSX responding would be under the control of this 5-sec stimulus representation which would not overlap, thus not compete, with CSA's later representation. It would follow from this account that no blocking would be

observed, and that responding during test phase with CSX would peak at the 5-sec mark. This is a testable prediction that, if shown to be the case, could validate RWDDM's account.

Also note that the time-dependent associative strength asymptote assumed by RWDDM implies that learning during a compound where the elements are of different durations is not stable. In particular, if CSA and CSX are the two elements of the compound phase of blocking, their associative strengths are updated by RWDDM as

$$\Delta V_A = \alpha_V(HA_A - (V_A + V_B))$$
$$\Delta V_B = \alpha_V(HA_B - (V_A + V_B)),$$

which in the steady state form an inconsistent system of linear equations,

$$V_A + V_B = HA_A$$
$$V_A + V_B = HA_B.$$

Since the compound phase of blocking only lasts for a few trials, RWDDM could produce the blocking seen on the left panel of figure 4.6. But if training with the compound was carried out for longer, the V values would grow without bound. However, there is evidence that in compounds formed by elements with asynchronous onsets, like in the compound phase of the blocking experiments here, the shorter stimulus comes to control CR timing and there is no summation of associative strengths (Fairhurst, Gallistel and Gibbon, 2003). Hence, it appears that with compounded asynchronous CSs, the shorter CS, more proximal relative to the US, comes to dominate and a summation rule like RW would not be applicable beyond the first few trials of training.

A model that is well placed to explain these results is CSC-TD. A long blocking CS will completely overlap a short blocked CS, blocking all units in the blocked CS. But in the case of a short blocking CS, there will be free units in the beginning of the blocked CS which will acquire associative strength, attenuating blocking. Given its similarity, MS-TD would likely produce comparable results. MoT and Let would not be able to account for any type of blocking given their current choice of rule for associative strength. Unlike RWDDM and the TD models, they both rely on the linear operator rule, which antedates the transition to the rules that sum associative strengths in the compounds as mentioned previously. MoT and LeT would need, at the very least, to replace the linear operator by the RW or other equivalent rule to be able to account for blocking and other compound phenomena.

4.5 Time specificity of conditioned inhibition

Learning occurs not only when a CS signals the occurrence of a US, but also when a CS signals the omission of a US. It is commonly assumed that the excitation caused by the former is counteracted by an inhibition produced by the latter. This is again formalized by the summation term in the RW rule. Conditioned inhibition is thus one of the phenomena that, together with blocking and other compound phenomena, challenged the contiguity interpretation of classical conditioning.

A conditioned inhibition procedure involves reinforced trials with a CS, say A+, intermixed with non-reinforced trials with a compound AB-. Conditioned responding develops during A+ trials but not during AB-. Hence, conditioned inhibition is a key conditioning phenomenon since it is also a form of discrimination learning.

Conditioned inhibition poses higher technical challenges for a model of learning and timing as responses cannot be directly observed. To assess conditioned inhibition two types of measures are used (Denniston and Miller, 2007): summation and retardation tests. There are different procedures that can generate inhibition, so I refer here specifically to the inhibition produced by alternating A+ with AB- trials. CSA is called a training excitor, and CSB an inhibitor. In summation tests, this inhibitor is then presented together with a different excitor, and the inhibitor is said to pass the test if there is a decrement in responding compared to the excitor alone. In retardation tests, the inhibitor by itself is now paired with the US, and it is said to pass the test if acquisition is slower than with a neutral stimulus. Denniston and Miller, 2007 reviewed a series of studies that varied the durations of the training excitor and that between the inhibitor and the training excitor. The studies showed that conditioned inhibition is observed when the temporal relations between training and testing are preserved, and not otherwise.

However, the studies reviewed by Denniston and Miller, 2007 used as measure of conditioned inhibition the time to resume drinking (licking suppression) when presented with the inhibitor. Williams, Johns and Brindas, 2008 investigated inhibition caused by reinforcement omission in excitatory conditioning, a more direct measure than licking suppression. In their experiments the inhibitor stimulus signalled the omission of one of two USs (at 10 or 30 seconds) that had been associated with the excitor stimulus. Using summation tests they found that the inhibitor would suppress responding only at the specific time of predicted US omission. Retardation tests confirmed that the time of US omission is encoded by the inhibitor.

I show here that RWDDM can account for inhibition and its time specificity. CSC-TD and MS-TD are also equipped to deal with these results. MoT and LeT do not currently have the necessary mechanisms to explain inhibition.

Simulations

I demonstrate time specificity of inhibition with simulations of Williams, Johns and Brindas, 2008 experiment. Excitors E1 and E2 signalled reinforcement after 10 and 30

seconds respectively, and inhibitors I1 and I2 signalled US omission after 10 and 30 seconds respectively. During phase 1, E1 and E2 were always reinforced, whilst the compounds E1I1 and E2I2 were never reinforced (see table 4.2). In phase 2 a transfer excitor E3 was trained on a mixed FI schedule, where in half the trials E3 lasted 10 seconds and in the other half 30 seconds. Phase 3 consisted of nonreinforced peak trials that lasted 90 seconds, a third with E3 compounded with I1, a third with E3I2, and a third with E3 alone. Figure 4.7 summarizes the results. Responding during E3 alone shows the two peaks characteristic of mixed FIs. As figure 4.7 shows, the compound excitor and inhibitor inhibits responding only at the time encoded by the inhibitor.

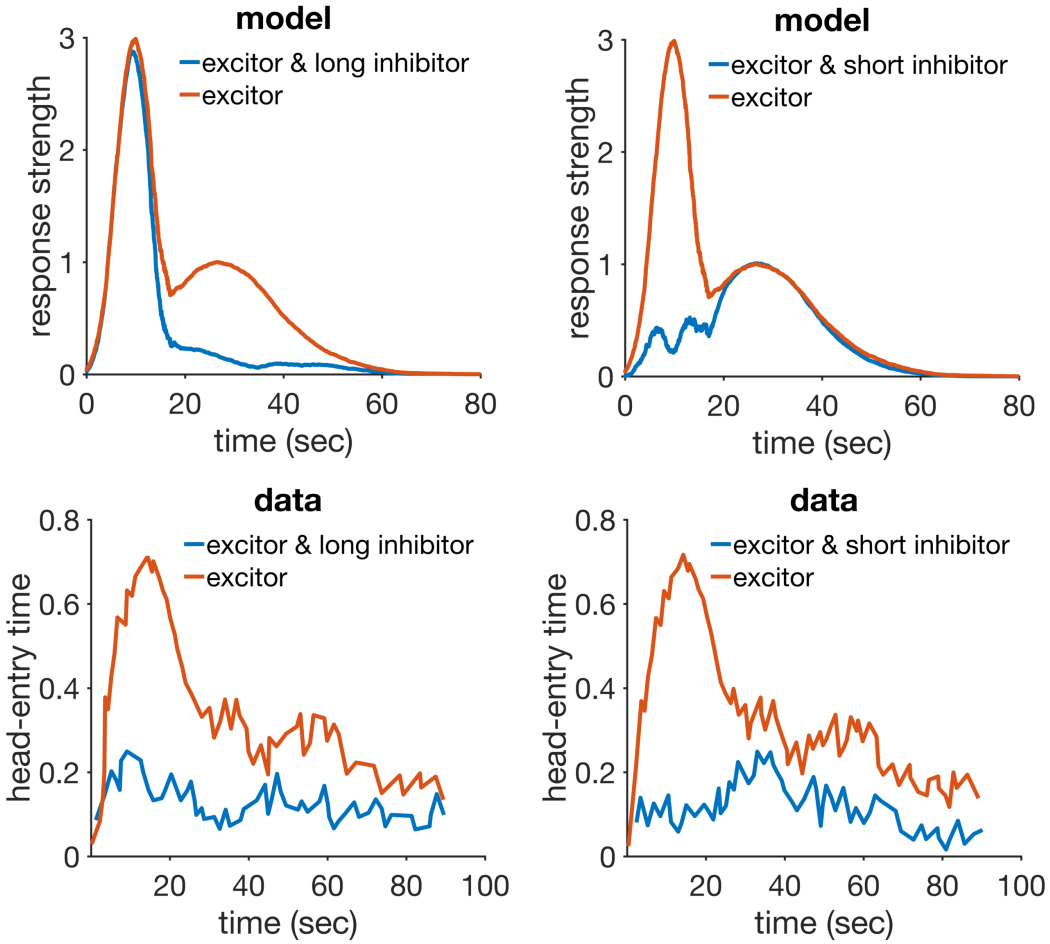


FIGURE 4.7: Conditioned inhibition. Left column: simulation (top) and data (bottom) from conditioned inhibition with a long inhibitor. Right column: simulation (top) and data (bottom) from conditioned inhibition with a short inhibitor. Data plots redrawn from figure 4 Williams, Johns and Brindas, 2008. Model parameters: $\alpha_t = 0.09$, $\alpha_V = 0.06$, $\mu = 1$, $\sigma = 0.35$, $m = 0.16$, $H = 30$.

Discussion

The account provided of inhibition by RWDDM relies on the traditional summation term inherited from the RW rule. Time specificity comes from the inhibitor CS timer

being treated just like any other CS timer, except that instead of timing the arrival of the US it times the arrival of US omission.

RWDDM predicts that the representation of an inhibitor CS has the same shape as of an excitor CS. This implies that inhibition is the exact opposite of excitation. This is a testable prediction which the empirical results above provide some validation.

The TD models provide a similar account of these data. Both CSC and MS TD have CS representations that allow for time specificity of US omission. Because the TD relies on the RW summation term, they can account for inhibition. LeT and MoT can also represent such time specificity, but because they rely on the older linear operator rule, they do not have a mechanism to account for inhibition.

4.6 Disinhibition of delay and compound peak procedure

The two related phenomena described here are important in that they appear to challenge the summation effect. A common observation is that a compound of two previously conditioned CSs usually produces more responding than its individual components (Rescorla, 1997; Kehoe and Macrae, 2002, p. 204). However, failure to obtain summation is also common (Rescorla and Coldwell, 1995; Pearce, George and Aydin, 2002), and the precise conditions when it is observed or not is still a current topic of debate (see Harris and Livesey, 2010, for a discussion). Here we consider two cases in which summation was not observed and that RWDDM can offer a possible explanation.

Aydin and Pearce, 1995 used an autoshaping procedure to condition pigeons to stimuli of 30 second duration. They observed little or no summation in compound trials, but a response curve with a consistent shift to the left. This earlier start of responding was observed even when one of the components was a neutral preexposed CS. The shift of the response curve to the left was termed disinhibition of delay.

Meck and Church, 1984 performed an analogue experiment using the peak procedure. They trained rats to associate a light and a sound (both of 50 second duration) individually to a reinforcement, and then used a peak procedure to investigate what happens to timing in their compound. Like Aydin and Pearce, 1995 they also found no summation and a shift to the left in the compound. Furthermore, rats also stopped responding earlier in the compound peak trials.

Taken together, these results appear to show that in some cases summation is not observed, and responding in the compound starts earlier than in the component CSs. One possible explanation for this effect is that the subject fails to recognize the two individual components of the compound, what is known as generalisation decrement. If this is the case then it would be a performance effect, and not a learning phenomenon. I cannot rule this out, but I show that RWDDM's trial variability in time estimation provides a plausible mechanism to explain this effect. The only other models in my analysis set that can account for this are MoT and LeT.

Simulations

RWDDM is capable of accounting for the earlier responding in compounds by noise in the timer. When a compound formed by CSA and CSB is presented, its two timers $\Psi_A(t)$ and $\Psi_B(t)$ will run in parallel. However, their rates A_A and A_B will have slightly different values due to noise. This implies that on every compound trial, one timer will be running slightly faster than the other. In contrast, on trials where only one CS is present, the timer will run faster in some trials and slower in others. Therefore, if on compound trials responding is guided by the faster timer, the average response curve for compounds will be shifted to the left when compared to the averaged response curve for a single CS.

Figure 4.8 shows simulations of disinhibition of delay and compound peak procedure. The figures were constructed by averaging the responses produced by equation (3.10) over 50 trials. The simulations reproduce in part the anticipation in responding during the compound that is observed in the data in both experiments (see top right and bottom left panels of figure 4.8). Meck and Church, 1984 reported a median peak time of 40±4 seconds for the response curves in compound trials, and 50±3.5 seconds in the individual trials. I ran 15 simulations as the one shown at the bottom row of figure 4.8, and analysed the peak times produced by each. I found an average peak time of 42±3 seconds in the compound trials, and 47±4 in the individual trials. Both results are within the error bounds in Meck and Church, 1984. Aydin and Pearce, 1995 did not analyse peak times or shift in the response curves, so I cannot make a quantitative comparison with our simulations.

Discussion

RWDDM can offer a good account for the lack of summation and earlier responding in compound trials in the two cases analysed here. It does so by having trial to trial variability in time estimation. However, the model shows a slightly higher maximum response frequency in compounds than in their components (top and bottom left of figure 4.8) something not observed in the data. This is not the product of summation, but of the slightly different asymptotes of learning in the faster and slower timers in the reinforced trial immediately preceding the peak trial. Our assumption was that in compound trials the timer running faster, with a higher slope A, would be the one guiding responding. When timing adaptation has reached asymptotic levels, the updates on slope A are due to noise in the value of the timer at reinforcement time, $\Psi(t^*)$. The two slopes, A_A and A_B, will have very similar values. In the reinforced trial preceding the compound peak trial, whichever timer produces a value of $\Psi(t^*)$ lower than the threshold will have its slope A adjusted up by the the slope update rule, likely causing it to overtake the other slope. This slightly higher slope will then be chosen in the peak trial that follows. But the corresponding V associated with that timer will have been updated on the previous reinforced trial based on the lower $\Psi(t^*) < \theta$ value. Because that is the denominator in $HA/\Psi(t^*)$,

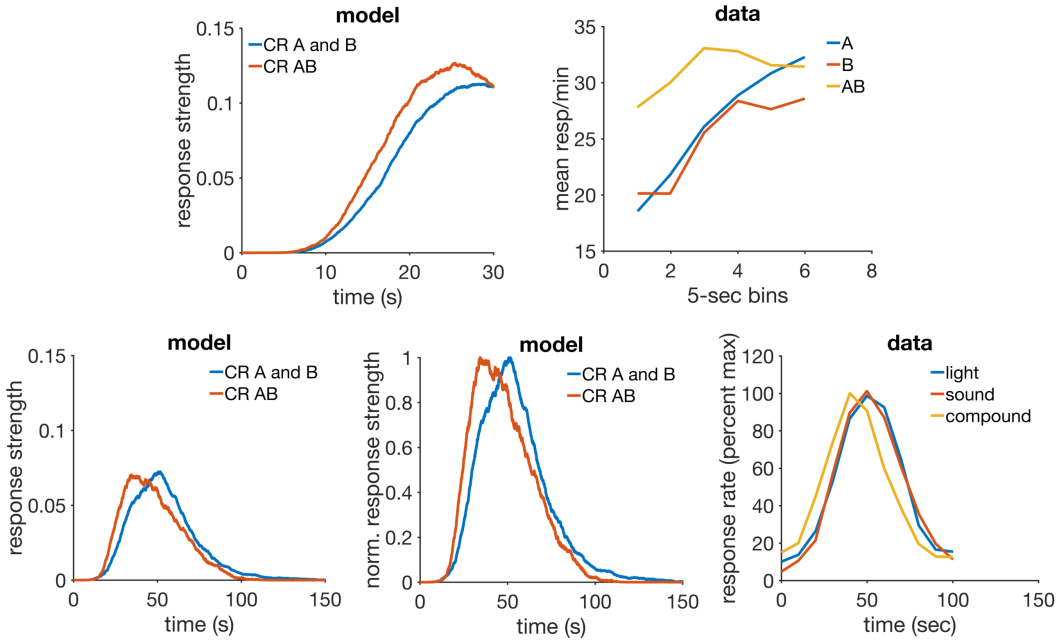


FIGURE 4.8: Disinhibition of delay and compound peak procedure. Top row: simulation (left) and data (right) of disinhibition of delay. Bottom row: simulation (left and middle) and data (right) of a compound peak procedure. The middle panel is a normalized (proportion of maximum response strength) version of the left panel. Data plot redrawn from figure 13 in Meck and Church, 1984. Model parameters: $m = 0.25$, $\theta = 1$, $\sigma = 0.18$, $\alpha_t = 0.75$, $\alpha_V = 0.1$, $H = 5$.

the V value of the chosen timer will be consistently slightly higher on the compound peak trials.

Other theories that might account for the data in this phenomenon are LeT and MoT. Both theories postulate intertrial variability in timer rate, the same mechanism used by RWDDM to explain this data. TD in any of its current versions lacks a mechanism to explain these data.

4.7 ISI effect

The interval between CS onset and US onset is called *Inter Stimulus Interval* (ISI). In general, measures of CR strength such as response frequency and amplitude decrease with longer ISIs (Smith, 1968; Gormezano, Kehoe and Marshall, 1983; Kehoe and Macrae, 2002). Response timing is commonly analysed by using fixed interval (FI) schedules of reinforcement, which rely on a fixed ISI. It is a well established result that the peak in the response curve decreases with longer FIs (Catania and Reynolds, 1968; Gibbon et al., 1997). However, the entire response curve approximately scales with FI. This is obtained by plotting different FI response curves as the proportion of maximum response strength versus the proportion to FI, a normalization procedure. The resultant normalized curves roughly superimpose (Rakitin et al., 1998; Matell and Meck, 2000; Matell and Meck, 2004; Allman et al., 2014). This

is sometimes called scalar timing, and it is one of the manifestations of the more general property of timescale invariance.

CSC-TD does not have a mechanism to explain either timescale invariance or the ISI effect. Its more recent development, MS-TD, can approximately reproduce both timescale invariance and the ISI effect. LeT is also a timescale invariant model, but does not appear to show the decrease in response peak as a function of FI. MoT, at least in its earlier version (Kirkpatrick, 2002), can reproduce both the ISI effect and timescale invariance.

Simulations

To demonstrate how RWDDM can reproduce the ISI effect I have simulated a delay conditioning procedure using three fixed interval stimuli. Figure 4.9 shows RWDDM simulations with FIs 5, 10 and 20 seconds. The top left panel shows within-trial response rate (given by equation (3.10)) averaged over 50 trials for each FI. The response curves show the same pattern as the data (bottom panel) from the ISI effect: a sigmoidal shape with a maximum that decreases as a function of FI duration. Note that because the curves are averages of 50 trials, the noise is averaged out.

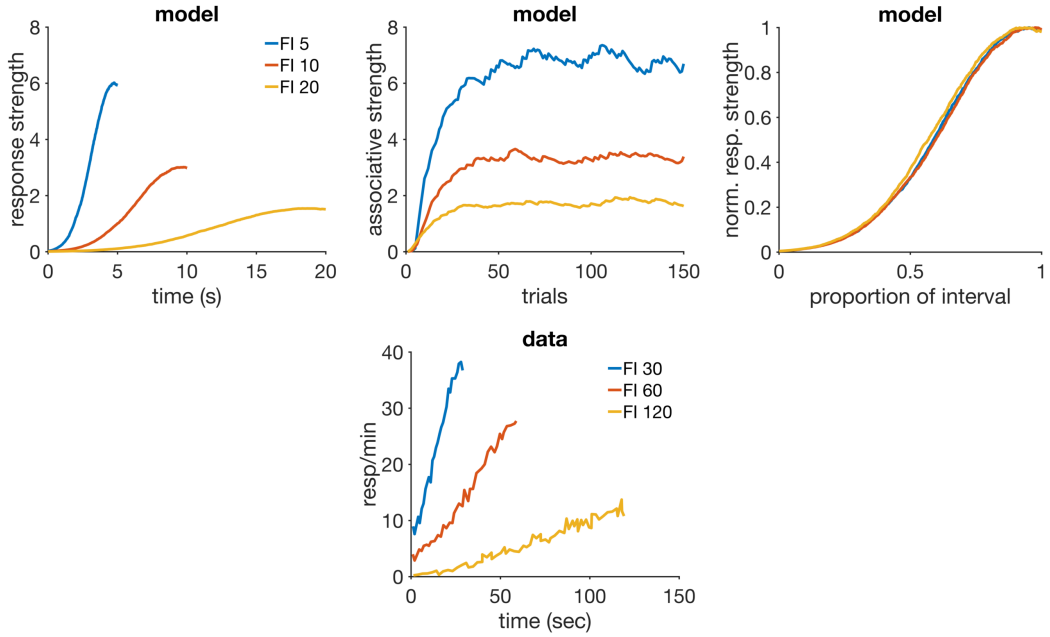


FIGURE 4.9: ISI effect. Top row: simulated average response rate during CSs (left), associative strength over trials (middle), and superimposition of response curves (right). Bottom row: average response rate data from an FI experiment, redrawn from bottom right panel of figure 4 in Kirkpatrick and Church, 2000. Model parameters: $m = 0.15$, $\theta = 1$, $\sigma = 0.3$, $\alpha_t = 0.2$, $\alpha_V = 0.1$, $H = 5$.

The top middle panel of figure 4.9 shows the associative strength acquisition curves for each FI. Their asymptotic levels are given by equation (4.1). V_∞ is approximately a linear function of A_∞, the TDDM slope. The different asymptotic levels of

associative strength are responsible for the different response peaks in the left panel of figure 4.9.

RWDDM also reproduces the superposition observed when FI response curves are normalized by maximum response rate and time to reinforcement (top right panel of figure 4.9).

Discussion

Gibbon and Balsam, 1981 attributed the ISI effect to the expectancy to reinforcement. A specific reinforcer carries, according to their view, an amount of expectancy H. This expectancy is spread back in time over the stimulus that signals US occurrence. Hence, for a CS of fixed duration T and US with expectancy amount H, the total expectancy during the CS is $h_T = H/T$. Our RWDDM account follows the same principles. The time to reinforcement T is computed by the ratio between the accumulation height at time of reinforcement $\Psi(t^*)$ and the timer slope at the current trial $A(n)$. This leads to the asymptote of learning in equation 3.9 being set to $HA_i(n)/\Psi_i(t^*)$. Superimposition of the response curves follows directly in RWDDM from the nature of noise in the linear accumulator. This noise guarantees that the time estimate produced by the model is timescale invariant (Simen et al., 2013).

The ISI effect can also be explained by the TD model with the Presence representation (Sutton and Barto, 1990) and with the more recently developed Microstimuli representation (Ludvig, Sutton and Kehoe, 2012). The Presence representation consists of a single element x which has the value 1 when the CS is present, and 0 otherwise. Its associative strength V is updated by the TD rule at every time step within a trial. In longer trials (longer FIs) the strength V will decay more, since it is updated more times in the absence of the US. This will lead to a lower asymptotic value for V. However, Presence TD cannot account for the superimposition of intratrial response curves. The CSC-TD fares even worse, unable to account for either ISI effect or superimposition (see Ludvig, Sutton and Kehoe, 2012, for a comparison between MS, CSC and Presence TD). The Microstimuli representation treats the stimulus as if it were composed of many units activated in sequence. Their activations follow a Gaussian shape which partially overlap. Later units have lower peaks and are wider than earlier ones. Because the number of Microstimuli are fixed, in longer FIs there is less temporal resolution which causes the US prediction to be lower than in shorter FIs, so it can explain the ISI effect. MS-TD's account of superimposition is only partial, although clearly better than CSC and Presence-TD.

LeT in its current version lacks a mechanism to produce decreasing response peaks with increasing FIs. But it can account very well for superimposition, since its time representation is timescale invariant. The earlier version of Modular Theory, called Packet Theory, has been shown to produce the ISI effect (see top row of figure 3 in Kirkpatrick, 2002). This prediction comes from longer interval durations

decreasing the probability of response packet generation in the model. MoT is also timescale invariant, so it generates superimposition quite easily.

To summarise, the ISI effect is explained either by time setting the asymptote of learning (RWDDM) or by a time representation that gets more diffuse with time, lowering the US prediction (MS-TD). Superimposition is explained either by the type of noise in the linear accumulator (RWDDM, LeT) or by stimulus units which have an approximately timescale invariant activation profile (MS-TD).

4.8 Mixed FI

Procedures where a stimulus signals reinforcement at more than one location in time are called mixed FI or two-valued interval schedules. A mixed FI involves only one CS which could be of short or long duration, and the subject has no way of knowing which duration it is currently experiencing until the US is delivered. Catania and Reynolds, 1968 conditioned pigeons in a mixed FI and reported a pattern of responding during the long CS that resembles a combination of two distinct FIs (with two peaks) when the separation between the intervals was in the ratio 8:1 but not at smaller proportions. Cheng, Westwood and Crystal, 1993 found a similar result (experiment 2) when the intervals were in 5:1 proportion and Leak and Gibbon, 1995 showed that with intervals in the 8:1 proportion the scalar property (measured by the CV) holds approximately even for three-valued interval schedules. Whitaker, Lowe and Wearden, 2003 ran three experiments with Mixed FIs in rats and found two peaks with the same CV when the proportion between the durations was greater than 4:1, but not for smaller proportions. They also found that the peak height at the short duration was higher than at the long duration in most cases. Whitaker, Lowe and Wearden, 2008 used intervals in the very small proportion 2:1 and still found two peaks that became more distinct when the short interval was presented more often than the long.

These results are interesting because they challenge in particular models of timing. They have served to provide evidence in favour of SET, and against BeT and the first version of LeT (Leak and Gibbon, 1995). Subsequently, they provided motivation for the development of the current version of LeT Machado, Malheiro and Erlhagen, 2009. LeT can now account for the multiple response peaks in Mixed FIs, and their superimposition, but it cannot produce peaks with decreasing heights. Modular Theory has the necessary mechanisms to account for all the features of the data above. The TD models, MS and CSC, could both account for multiple peaks, but their account of superimposition would vary, with MS being superior than CSC. I show next that RWDDM can account for all features of the data in Mixed FIs.

Simulations

In this simulation one CS was used which was followed by reinforcement either after 15 or 75 seconds randomly chosen, a proportion of 5:1. My assumption was that in

Mixed FI experiments subjects form two independent stimulus representations, one for the short interval x_S, and another for the long interval x_L, each with its respective associative strength (V_S, V_L) and timer (Ψ_S, Ψ_L). At CS onset, both timers begin timing, generating the two representations x_S and x_L, and at each point in time behaviour is guided by the representation with the highest activation value. When a reinforcement occurs, the CS representation with the highest activation value is the one to which credit is assigned.

The left panel of figure 4.10 shows the simulated responses averaged over 50 trials of the long 75-second duration. Two peaks, centred roughly at 15 and 75 seconds, of decreasing heights and increasing widths are clearly seen, matching roughly with the data (right panel).

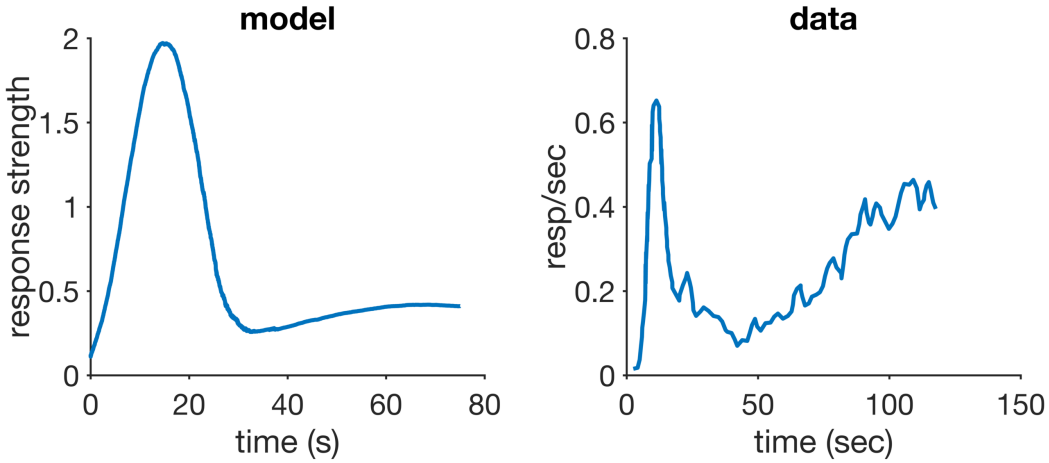


FIGURE 4.10: Mixed FI. Left: simulated response strength during long trials. Right: response strength data from a mixed FI experiment, redrawn from figure 3 in Leak and Gibbon, 1995. Model parameters: $\alpha_t = 0.2$, $\alpha_V = 0.1$, $\mu = 1$, $\sigma = 0.425$, $m = 0.2$, $H = 30$.

Discussion

RWDDM's mechanism for dealing with mixed FIs is in essence the same as for single FIs. The only difference is that instead of only one timer (and CS representation) in Mixed FIs RWDDM uses as many timers (and CS representations) as rewards. I have not however addressed explicitly how one CS can give rise to two distinct representations. One possible explanation is that the slope adaptation rule (equation (3.6)) is only applied when the difference between the two intervals is below a certain amount. If the difference is above this amount, then the model would create a new representation. In fact, the data reviewed here suggests that animals may not be able to distinguish two intervals if they are in proportion below 2:1.

To the best of our knowledge, the only other model from our analysis set that has tried to address the behaviour in mixed FIs is LeT. Machado, Malheiro and Erlhagen, 2009 have succeeded in obtaining the two peaks with the same CV using LeT. Their account relies on a single accumulator in the form of a series of states activated at a

fixed rate. This rate is fixed within a trial, but varies from trial to trial. After repeated training with a mixed FI, the states around the reinforced times receive on average more associative strength than the ones away from them. This activation pattern generates the response peaks seen in the data. However, as the authors note, 'in mixed-FI schedules, the response rate [produced by LeT] at the first peak is equal to or lower than the response rate at the second peak, but never higher,' which is the opposite of what the data shows. The authors suggest that a decaying arousal function might need to be added to the model so as to allow response rate to decay with interval duration.

Modular Theory is capable of accounting for the behaviour in Mixed FIs since its pattern memory for time is based on SET, which has been shown to account for these data (Leak and Gibbon, 1995). MoT's account is similar to RWDDM's in that both rely on a separate accumulator (and memory) for each time of reinforcement. CSC-TD would likely produce two peaks, since it relies on a perfect discretization of time into as many units as time-steps. But the curves would not superimpose when scaled as there is no mechanism to account for timescale invariance. MS-TD would also account for the two peaks but superimposition would likely not be fully obtained as its simulations of the ISI effect have only partially reproduced it (see section 4.7 and Ludvig, Sutton and Kehoe, 2012).

4.9 VI and FI

Schedules of reinforcement specify the conditions of reinforcement delivery. There are a number of different types of schedules, some are based on the time elapsed between reinforcements, some on the number of responses emitted between reinforcements, but there can be other possibilities. Of particular interest for a timing and conditioning model are the two most commonly used time-based schedules: variable and fixed interval. Variable Interval schedules of reinforcement (VI) consist in the delivery of a US following a CS that varies in duration from trial to trial. The CS durations are usually derived from an arithmetic or geometric sequence. In contrast, Fixed Interval schedules of reinforcement (FI) use a CS of fixed duration in all trials. Skinner and Ferster, 2015 reported that VIs tend to produce behaviour with a constant rate throughout the trial, whilst FIs produce scalloped curves with a pause following each reinforcement and a rapid increase in rate until the next reinforcement.

Catania and Reynolds, 1968 performed a detailed analysis of behaviour under VIs and found that response rate declined with the average reinforcement rate. Within a trial response frequency increased with time, following approximately a negatively accelerated curve. When normalized by maximum response rate and time to reinforcement, these curves showed a considerable degree of superimposition.

Matell, Kim and Hartshorne, 2014 trained rats on a VI in which intervals were sampled from an uniform distribution $\mathcal{U}(15, 45)$, and then tested using a peak procedure. They compared the VI response peak curve to the peak curve from a control group trained on an FI 30 (the mean of the VI distribution). Although the two curves were not significantly different statistically, the VI response peak curve peaked slightly earlier and was slightly higher than the control group.

Jennings et al., 2013 compared timing performance between VI and FI in three experiments, but found VI timing only in a VI where the average interval was 30 seconds. The other experiments from the same paper produced results more in agreement with the earlier work by Skinner and Ferster, 2015 showing a constant rate of responding during VI trials.

Taken together, these studies appear to show that timing may sometimes be present during VI schedules. In this case, animals appear to be learning the average of the interval distribution. Here I demonstrate with simulations that RWDDM can account for such findings. The only other model in our analysis set that can account for this result is Modular Theory.

Simulations

In this simulation a random VI was produced by sampling intervals from a discrete uniform distribution $\mathcal{U}(15, 45)$. Non-reinforced peak trials of duration 135 seconds were interspersed during the VI, with a probability of 0.25. Our assumption here is that subjects will keep adapting the timer rate A over trials. In this case, equation (3.6) calculates the exponential moving harmonic average of the CS durations. Since it is a moving average, the predicted peak time will depend on the actual intervals used and their presentation order, but the non-moving harmonic average of all intervals is 27.1 seconds. This is earlier than the arithmetic average (30 seconds), which is in line with the trend observed in the data by Matell, Kim and Hartshorne, 2014.

Figure 4.11 (top left panel) compares the response strength averaged over peak trials in the VI and in a regular peak procedure with FI 30. The VI peak is higher and slightly earlier (at roughly 29.68 sec) than the FI peak, matching roughly with the data (bottom row). When normalized both by peak height and time the curves show the superimposition (top right panel) also seen in the data.

Discussion

The model predicts a harmonic mean value for the position of the response peak, which is always less than the arithmetic mean, but because it is a weighted moving average the actual value may vary. As I saw in the simulations, the VI response curve peaked at a value (29.68 sec) very near the arithmetic mean of the intervals (30 sec). This may explain the trend observed in the data by Matell, Kim and Hartshorne,

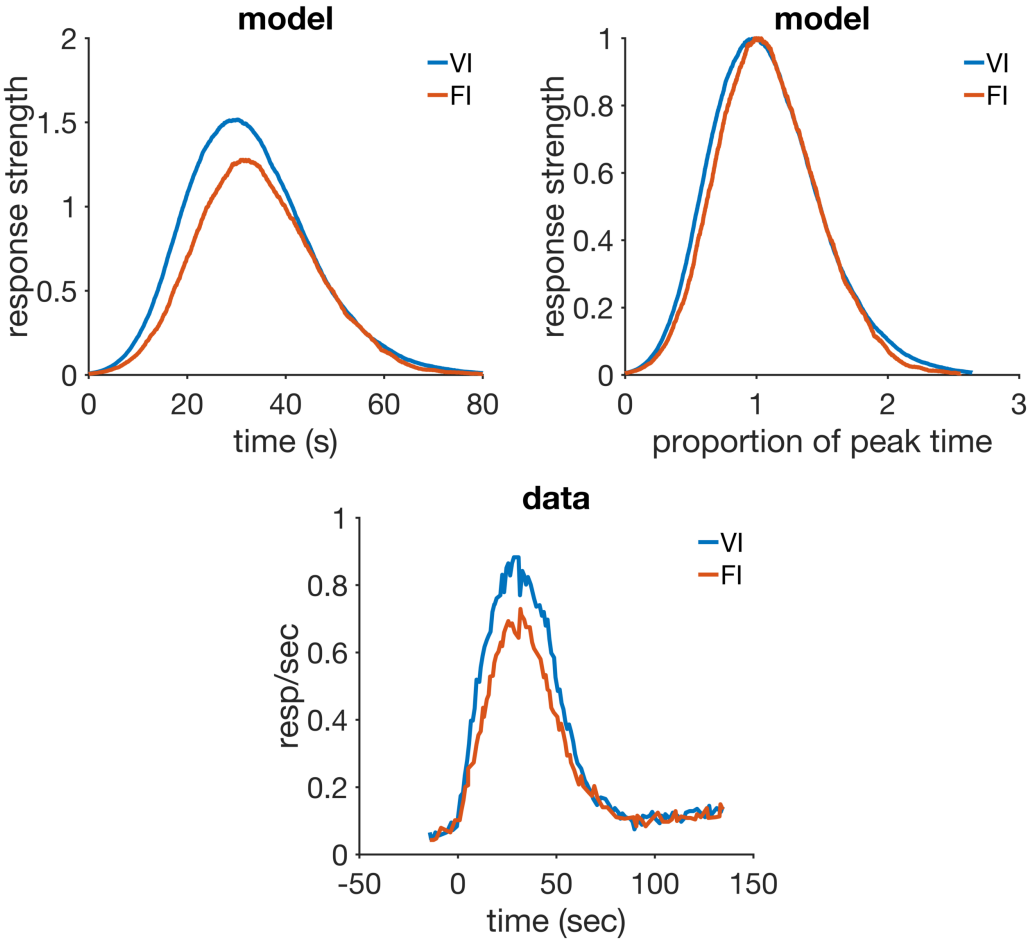


FIGURE 4.11: VI and FI. Top row: simulated average response strength during peak trials (left), and the same data plotted after both axes are normalized (right). Bottom row: average response strength data from an experiment in VI and FI, redrawn from figure 1 in Matell, Kim and Hartshorne, 2014. Model parameters: $\alpha_t = 0.1$, $\alpha_V = 0.1$, $\mu = 1$, $\sigma = 0.3$, $m = 0.2$, $H = 40$.

2014. However, because that trend was not statistically significant, further experiments would be needed to establish if the response peak during VIs is nearer to the harmonic or the arithmetic mean.

Taken together, these results are more easily accommodated by theories that can store an average of CS durations like RWDDM. Modular Theory is such an example, since it also stores an average of intervals in its pattern memory. Other models such as LeT and MS or CSC-TD would struggle with this result. The CS representation in these models break down the CS into a sequence of units activated serially in time. With a uniform distribution of CS durations associative strength would likely be spread broadly over the weights that cover the interval, generating a broader pattern of responses that would not be centred on the mean.

4.10 Temporal Averaging

Although animals are able to time different durations simultaneously, as seen in mixed FIs, paradoxically under certain circumstances a type of temporal averaging can be observed. This is a relatively new and important phenomenon, which challenges in particular theories of timing to propose a mechanism that can explain such averaging.

When rats are trained using two distinct stimulus modalities, a visual stimulus (a light) and an auditory (a tone), each signalling reinforcement at a different time, responding during compound presentations of both stimuli peaks roughly in the middle of both durations (Swanton, Gooch and Matell, 2009). This intermediate response curve to the compound superimposes with the two other single stimulus curves when normalized, suggesting that the animal is timing only one average duration. The type of average being computed appears to be modulated by the reinforcement probabilities associated with each stimulus duration, with the weighted geometric average fitting the data better than a weighted arithmetic average or a non-weighted average (Swanton and Matell, 2011; Matell and Henning, 2013; Matell and Kurti, 2014). Significantly, temporal averaging in rats is only consistently observed when the auditory stimulus signals the short interval and the visual stimulus signals the long interval (Swanton and Matell, 2011; Delamater and Nicolas, 2015). Even when each stimulus is associated with a different response option (light reinforced with a left nosepoke, tone with a right) rats still tend to mix the temporal information during compound trials (De Corte and Matell, 2016).

I do not make a strong claim about RWDDM's ability to explain this data. Rather, I show that it has the necessary elements from which an account can begin to be formulated. MoT also has similar elements from which an account can be built. CSC-TD, MS-TD and LeT do not appear to be equipped to deal with this phenomenon.

Simulations

In RWDDM the accumulator is the mechanism that marks the passage of time. The temporal proximity to an event is determined by how close the level of accumulation is to a fixed threshold value. A CS that signals reward later than another CS, will have a lower rate (A_{low}) of accumulation than the shorter CS (A_{high}). Because in RWDDM associative strength is set by time to reward, the two CSs will also have different associative strengths, V_{low} and V_{high} respectively. We may assume that under temporal averaging circumstances the stimuli are of such nature that they cause the subject to integrate their information. At the start of the compound trials, the ambiguity presented by the compound stimulus may cause the representations of the two component stimuli to be only partially retrieved. If the subject fails to represent the two stimuli separately, the result may be the formation of a single representation composed by only a fraction of the timing rate A and associative strength V of each individual stimulus. The fractions are then added into one single rate and one

single associative strength, and processed as if they were the components of a single stimulus representation. For the simulation below, I assume that the fractions added are exactly half of their individual values: $A_{\text{compound}} = A_{\text{low}}/2 + A_{\text{high}}/2$, and $V_{\text{compound}} = V_{\text{low}}/2 + V_{\text{high}}/2$.

I used a long CS of duration 20 seconds and a short CS of duration 10. I simulated a peak procedure with each CS and with the compound. A plot of the response strength averaged over peak trials is shown in the top left panel of figure 4.12. The three peaks scale when normalized (top right panel).

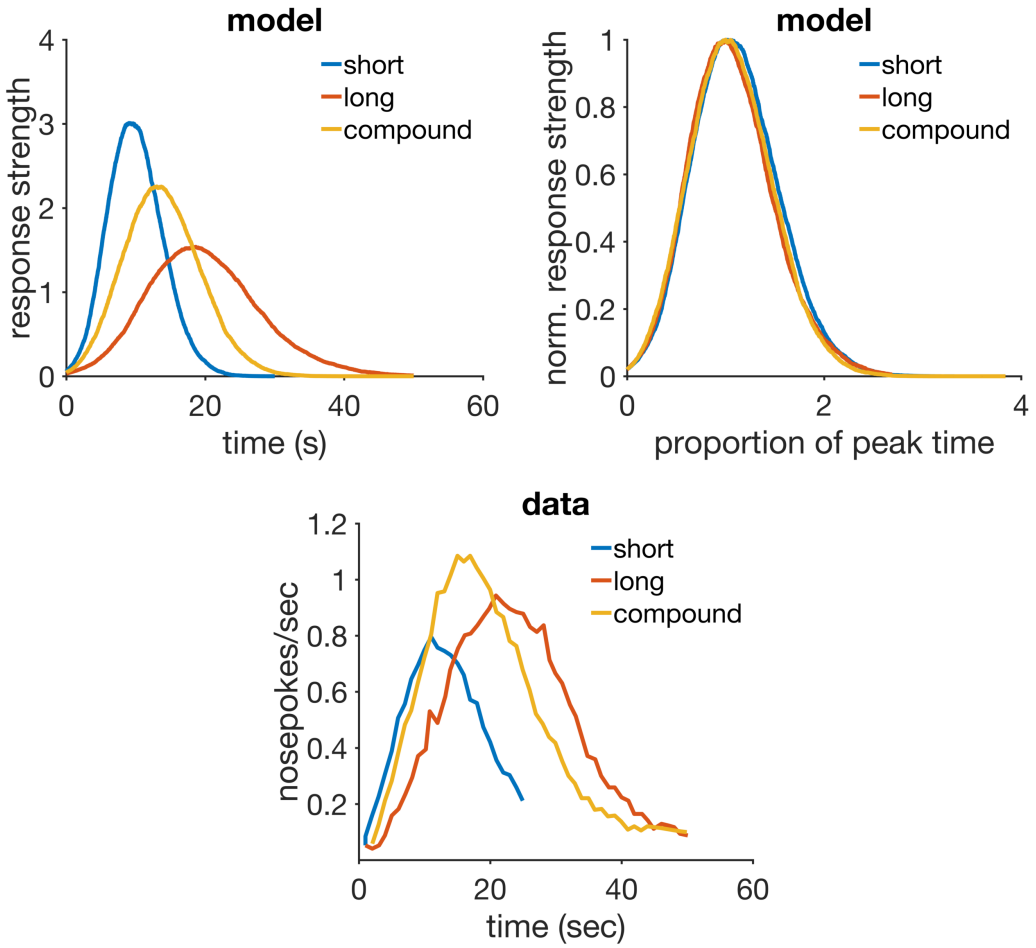


FIGURE 4.12: Temporal averaging. Top row: simulated response strength averaged over peak trials in temporal averaging (left), and the same data normalized by maximum response strength and peak time (right). Bottom row: peak trial response strength data from an experiment in temporal averaging, redrawn from figure 1 in Swanton, Gooch and Matell, 2009. Model parameters: $\alpha_t = 0.2$, $\alpha_V = 0.1$, $\mu = 1, \sigma = 0.35, m = 0.2, H = 30$.

The peak of the compound is roughly at 13.33 sec, which would be the expected value for an averaged rate $A = (1/10 + 1/20)/2$, the harmonic average of the intervals. The height of the compound peak is also at an intermediate level between the two end peaks. The simulations match roughly with the data (bottom row of figure 4.12)

Discussion

The assumption I made here, that temporal averaging is the result of only one accumulator being active during the compounds and fed with half the rate for each of the stimuli, is plausible and can accommodate the main features of the data. However, given the evidence from mixed FIs it seems animals are capable of keeping multiple timers running in parallel, without averaging their rates. Also, if averaging of rates always happened during compounds, then the explanation provided by RWDDM for the left shift in the response curve in the compound peak procedure would not hold. I suggest one possible way of interpreting these three phenomena based on a failure of representation selection caused by the ambiguity of the signal. In mixed FIs there is one single CS that signals two rewards at very different times. There is not much ambiguity in how to interpret the signal, so the subject keeps two timers running in parallel. In the case of compounds formed by individual CSs that signal reward at the same time, as in the compound peak procedure, there is also not much ambiguity. There's very little difference between the time memories evoked by the CSs, so choosing only one, the faster one, leaves no ambiguity as to which CS is signalling reward. In the case of compounds formed by individual CSs of different modalities that signal reward at different times, the ambiguity might be such that cannot be resolved easily. The information from each CS may then be only partially retrieved and added into one representation, resulting in temporal averaging.

As mentioned previously, this is not a strong account of the conditions that generate temporal averaging. But whatever the final word on this may be, RWDDM has components that allow it to generate averaging and timescale invariance. However, RWDDM predicts this average to be the harmonic mean, and not the geometric mean weighted by reinforcement probabilities that has been frequently found (Swanton and Matell, 2011; Matell and Henning, 2013; Matell and Kurti, 2014). Also, Matell and Henning, 2013 reported evidence of summation of response rates during the compound trials. In my simulations here I assumed that equal fractions were taken of the rates of each CS, resulting in a combined non-weighted harmonic average of rates, but different fractions (or weights) may be taken. In particular, the data indicates that the weights are set by the reinforcement probabilities of each individual stimulus. Since this information is stored in the associative strength V, we could assume the subject integrates the two timer rates as follows:

$$A_{\text{compound}} = \left(\frac{V_{\text{low}}}{V_{\text{low}} + V_{\text{high}}}\right) A_{\text{low}} + \left(\frac{V_{\text{high}}}{V_{\text{low}} + V_{\text{high}}}\right) A_{\text{high}}.$$

Although this would produce a weighted average, it is still a weighted harmonic average of the intervals and not a weighted geometric average found in the data, so the account given by RWDDM would still be partial. As for the summation of response rates observed in the compound trials, this could be explained by RWDDM if instead of taking a fraction of the V values for each stimulus to form the Vcompound,

the subject simply summed, or partially summed, both V values.

Another model that is equipped to deal with averaging is Modular Theory. If we allow for one single accumulator fed by one half of each time memory, then MoT would predict a peak of responding at the arithmetic mean of the two intervals. A weighted average could also be obtained following the procedure I sketched above for RWDDM. However, this would yield a weighted arithmetic mean, and not the weighted geometric mean obtained in the data. As for timescale invariance, MoT relies on a noisy timer threshold whose mean is always a fixed proportion of the time memory, with a standard deviation proportional to this mean. Therefore, timescale invariance is guaranteed for all time memories, averaged or not.

LeT would not be able to explain temporal averaging without modifications. It cannot change its average transition rate between states without compromising timescale invariance. Without changing the transition rate it is difficult to see how else LeT could account for a different timing in the presence of the compound. CSC-TD and MS-TD also lack any mechanism that could be used to account for temporal averaging.

4.11 Trace Conditioning

In a trace conditioning experiment the CS terminates before the occurrence of the US. Pavlov, 1927 observed that in spite of the gap separating CS and US, conditioning still occurred. This observation led him to formulate the theoretical construct of a *stimulus trace*, a neural activity initiated by the stimulus and that persists for a while after the stimulus has ended. This stimulus trace would bridge the gap between CS and US, allowing the two stimuli to become associated.

CR frequency during the CS is usually lower in trace conditioning than in delay conditioning (Cole, Barnet and Miller, 1995), but CR timing in trace conditioning is less well understood. Buhusi and Meck, 2000 investigated timing by first training animals on a trace conditioning procedure where the gap between CS and US was 30 seconds. Little or no CR was observed during the CS, but responding increased from CS offset and reached a peak at the time the US was delivered. They then tested the subjects by varying the duration of the CS. When the CS duration was shortened or lengthened by 15 seconds the CR peak shifted to the left or to the right respectively by roughly the same 15 seconds. This suggests that animals use the CS offset as a signal to start timing the US occurrence.

Williams et al., 2016 used trace conditioning to evaluate whether the CS would actually become inhibitory. They trained two groups of subjects, one group had USs presented randomly during the intertrial period (but not during the trace gap), and the other had a US-free intertrial period. Each group was subdivided into three different conditioning procedures: trace, delay and embedded (the US appeared a fixed time after CS onset but before CS offset). In all groups and conditioning procedures

the CR peaked at the expected time of US. When tested for inhibition, using summation and retardation tests, the ITI US trace CS was found to be strongly inhibitory in comparison to the no-ITI US trace CS. Furthermore, the whole CS appeared to have become inhibitory, and not only its beginning.

The experiment by Williams et al., 2016 presents a challenge to learning models that rely on a post-CS trace to explain trace conditioning. In TD models for example this trace is called an eligibility trace (Sutton and Barto, 1998, p. 163). CSC and MS-TD both predict that, in the case of an excitatory context (ITI USs) only the initial part of the CS would become inhibitory. This is because the eligibility trace would cause the later part of the CS, the most proximal to the US, to become excitatory, and this excitation would propagate back with decreasing strength in such a way that the early part of the CS would have zero or negative associative strength. The other models evaluated here, LeT and MoT, because they rely on the linear operator rule do not have a mechanism to explain inhibition.

RWDDM takes a different approach which does not require the use of an eligibility trace. It treats the gap between CS offset and US onset as a separate stimulus, with its own timer and associative strength. Under the ITI US condition, the context would become excitatory and the CS (with its own timer and associative strength) would become inhibitory for its entire duration. Below I present simulations of the experiment in Williams et al., 2016 and show that RWDDM can reproduce the inhibition in the CS obtained in the data.

Simulations

I simulated RWDDM with the same 2×3 factorial design as in Williams et al., 2016, namely the two groups (no-ITI USs and ITI USs) and the three conditioning procedures (embedded, delay and trace). For the purposes of timing, I split the trial into three distinct parts: the ITI, the CS, and in the case of trace conditioning only, the gap. In terms of stimulus representations, I split the stimuli as: the context, the CS, and also only for trace conditioning, the gap. Therefore, the following arrangement of timers and representations were used:

ITI The US at the end of the gap marked the beginning of the ITI and the CS onset marked the end. The timer $A_{\mathrm{CX,ITI}}$ was tuned to either the CS onset (no-ITI USs) or the next US (ITI USs). V_{CX} was updated at every US occurrence (ITI USs) or at the end of the ITI (no-ITI USs).

CS At CS onset, two timers were started, one belonging to the CS, A_{CS}, and another belonging to the context $A_{\mathrm{CX,CS}}$. Both timers were tuned to either the US (embedded and delay) or the CS offset (trace), at which time V_{CS} and V_{CX} were updated.

Gap In the case of trace conditioning only, CS offset marked the beginning of timers $A_{CX,Gap}$ and A_{Gap}, both tuned to the US onset. At US onset, V_{CX} and V_{Gap} were updated.

The ITI lasted for 340 seconds, and the CS for 120 seconds. In embedded conditioning the US appeared at 110 seconds, and in delay conditioning at 120 seconds. In trace conditioning the gap lasted for 10 seconds. Figure 4.13 compares the response strength obtained in the simulations (top panels) to those from the experimental data (bottom panels).

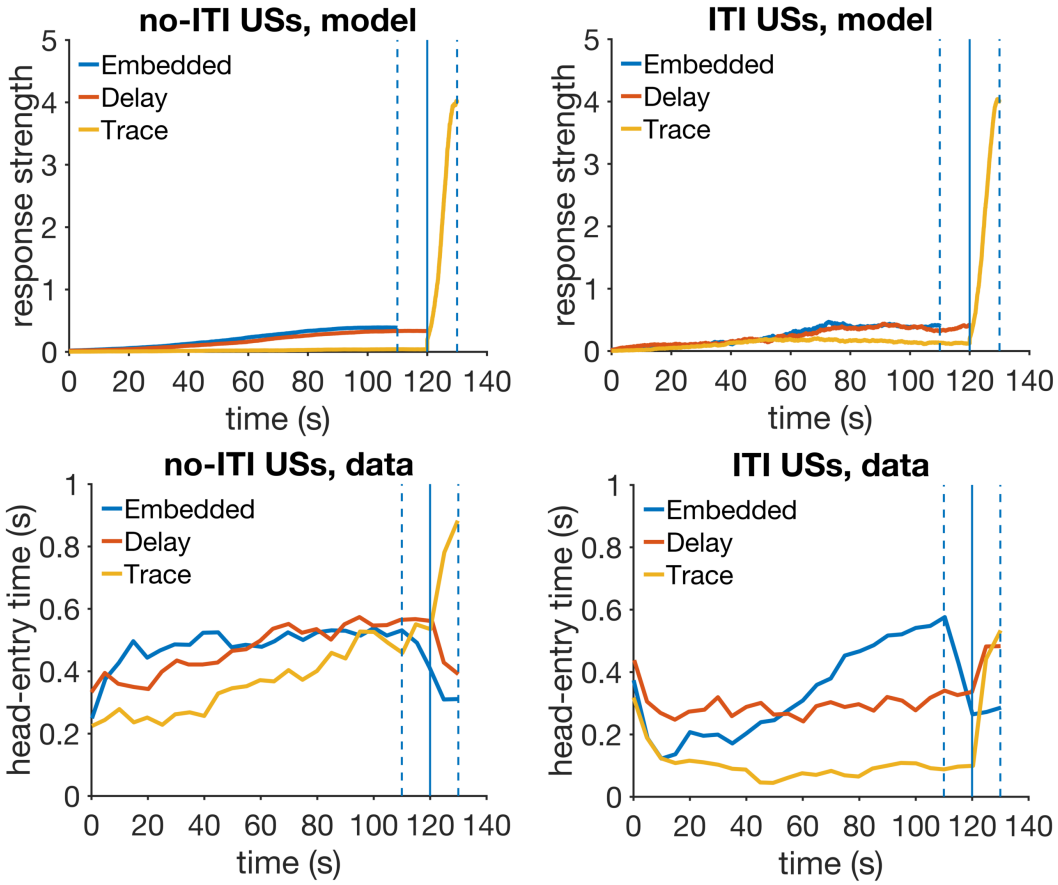


FIGURE 4.13: Trace, embedded and delay conditioning. Top row: simulated response strength averaged over 30 trials for the no-ITI USs (left) and the ITI USs groups. Bottom row: experimental data redrawn from figure 2 in Williams et al., 2016. Model parameters: $\alpha_t = 0.1$, $\alpha_V = 0.07$, $\mu = 1$, $\sigma = 0.4$, $m = 0.15$, $H = 40$.

Although all simulated response curves peak at the time of US onset, the difference in peak height between CS and gap is considerably larger in RWDDM than in the data. This is due to the hyperbolic reward discounting in RWDDM (equation (3.9)). The asymptotic value of associative strength is $H/120$ for the CS in delay conditioning, $H/110$ for the CS in embedded conditioning and $H/10$ for the gap. Hence, there is a 12-fold net associative strength difference between the gap and the delay CS, and a 11-fold difference between gap and embedded CS.

Figure 4.14 compares the associative strength of the trace conditioning CS in the no-ITI US and ITI US groups. The CS in the ITI US group became considerably more inhibitory (15 times more) than the CS in the No-ITI US group.

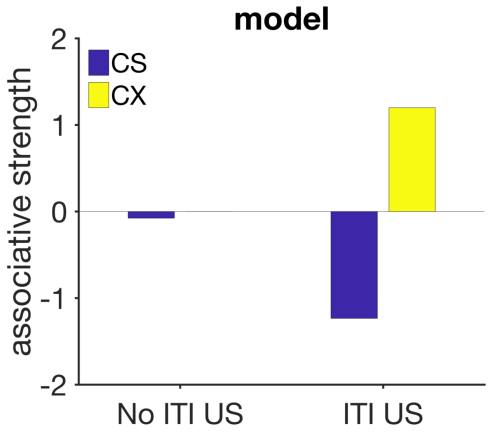


FIGURE 4.14: Simulated associative strength in trace conditioning. The values for the CSs are: $V_{\text{CS}} = -1.23$ for the ITI US group, and $V_{\text{CS}} = -0.08$ for the no-ITI US group.

Discussion

RWDDM produced similar response curves to the data in the embedded, delay and trace conditioning experiments. It also produced a strong inhibitory CS in the trace conditioning ITI USs group, and not in the no-ITI USs group, just as was obtained in the actual experiments. This strong inhibition was due to the context being reinforced at an average rate of once every 30 secs during the ITI period. This caused the context to acquire an associative strength of $V_{CX} \approx H/30$, which then competed with the CS via the summation term inherited from the RW rule yielding a value of $V_{CS} \approx -H/30$. In the case of the no-ITI USs group, the context acquired excitatory associative strength only at the end of the gap, and even then only in competition with another stimulus, the 'gap CS'. It then underwent two decreases in strength, one at the end of the ITI and another at the end of the CS. This very low value for V_{CX} made the CS only very slightly inhibitory in the no-ITI Uss group.

Here I have used a timer for each stimulus, but the simulations indicate that this may have been superfluous. During CS presentation and the gap interval two timers were run in parallel, but only one was actually needed. It would be more parsimonious, and more plausible, that in those cases only one timer was responsible for supplying temporal information to the two stimuli currently present.

Other models cannot easily explain the inhibition observed in the ITI USs trace conditioning. Williams et al., 2016 simulated MS-TD on the same experiment and found that it predicted a temporal pattern of increasing associative strength from CS onset until offset, unlike the data. If however the MS-TD discount factor γ in equation (3.13) is allowed to vary between the ISI USs and no-ITI USs groups then a better

match to the data is obtained. The same strategy could in principle be employed by CSC-TD with similar results. As mentioned previously, LeT and MoT cannot explain inhibition because of their choice of learning rule.

4.12 Summary of Results and Analysis

Table 4.3 summarizes the results from the simulations. RWDDM was able to reproduce the main features of the data in 9 out of the 11 experiments. In the other 2 the model was able to partially account for the data.

To allow for comparison I have offered qualitative predictions for the other 4 models in table 4.3. It is important to note that for most of the 11 phenomena analysed here simulations using these models are not available in the literature. Although I have tried my best to provide predictions based on our understanding of these models, I have not actually simulated them. Therefore it is possible that in some cases a model may produce results that I did not foresee if the right set of parameters is found or some of the assumptions are relaxed. It is also possible that some simple modifications might allow the models to explain the data. I endeavoured to point out some such modifications that seem likely to work when discussing the simulation results above, but I do not make predictions based on them because the purpose here is only to provide a comparison of the current mechanisms of each model and therefore encourage future work on model improvement. With that in mind, Modular Theory has fared best after RWDDM, being able to account for 7 out of the 11 experiments. MS-TD and CSC-TD shared the second place with 4 out of 11. LeT came in last, able to account for 2 experiments. The last column of table 4.3 identifies the main mechanisms responsible for successfully accounting for each phenomenon.

Simulation	Group	Phase 1	Phase 2	Phase 3
Acquisition, extinction and reacquisition	FI 5	80 CS+	100 CS-	80 CS+
Extinction with diff. duration	FI 20-40	150 CS(20)+	150 CS(40)-	—
	FI 20-10	150 CS(20)+	150 CS(10)-	—
ISI effect	FI 5	150 CS+	—	—
	FI 10		—	—
	FI 20		—	—
VI vs FI	VI 30	mixed 1500 CS+, 375 peak	—	—
	FI 30	mixed 500 CS+, 125 peak	—	—
Mixed FI	MFI 15-75	mixed 200 A(15)+, 200 A(75)+	—	—
Latent inhibition	Preexposed A	80 A-	120 A+	—
	Control C	—	120 C+	—
Blocking diff. durations	Blocking A	120 A(10 or 15)+	60 A(10)B1(15)+ or 60 A(15)B1(10)+	—
	Blocked B1, B2	—		—
	Control C	—	60 C(10)B2(15)+ or 60 C(15)B2(10)+	—
Disinhibition of delay	FI 30	100 A+, 100 B-	mixed 300 A+, 300 B+, 100 AB+	—
Compound peak	FI 50	100 A+, 100 B+	mixed 300 A+, 300 B+, 100 AB peak	—
Conditioned inhibition	1 group	mixed 300 each E1(10)+, E2(30)+, E1(10)I1(10)-, E2(30)I2(30)-	mixed 300 each E3(10)+, E3(30)+	100 each peak E3I1, E3I2, E3
Temporal averaging	FI 10-20	700 L(20)+, 700 S(10)+, 154 L peak, 154 S peak, 154 SL peak	—	—
Trace conditioning	embedded	110 CS+	—	—
	delay	120 CS+	—	—
	trace	120 CS-,10 CX+	—	—

TABLE 4.2: Simulation designs.

TABLE 4.3: Summary of main simulation results and comparison with other models. Notes: (1) if learning rate is allowed to vary; (2) if discount factor is allowed to vary.

phenomenon	RWDDM	CSC-TD	MS-TD	LeT	MoT	explaining mechanism
Faster reacquisition	yes	yes[1]	yes[1]	yes[1]	yes[1]	Time-adaptive stimulus representation or changes in learning rate.
Time change in extinction	yes	no	no	no	yes	Separate rules for time adaptation and associative strength.
Latent inhibition and timing	part.	no	no	no	no	PH rule and separate rules for time adaptation and associative strength.
Blocking with diff. durations	part.	yes	yes	no	no	RW rule and ability to time any stimulus or distributed time representation.
Time spec. of conditioned inhibition	yes	yes	yes	no	no	RW rule and concentrated memory for time or distributed time representation.
Compound peak procedure	yes	no	no	yes	yes	Intertrial variability in time estimation.
ISI effect and superimposition	yes	no	part.	part.	yes	Asymptote of assoc. strength set by time and accumulator noise or time representation that gets diffuse with longer time.
Mixed FI	yes	part.	part.	part.	yes	Ability to generate multiple time representations or a single distributed time representation.
VI and FI	yes	no	no	no	yes	Memory that stores average of intervals.
Temporal averaging	yes	no	no	no	yes	Memory that stores average of intervals and the accumulator.
Trace conditioning	yes	yes[2]	yes[2]	no	no	RW rule and TD discount factor.

Chapter 5

Discussion

5.1 RWDDM Mechanisms

RWDDM was able to reproduce faster reacquisition due to its memory for time being conserved during extinction. This memory is used to activate the stimulus representation. Learning is slower in acquisition because RWDDM increases the activation in the stimulus representation gradually over the trials. The stimulus representation needs to be 'built up' first, and this process depends on learning the timing of the US. Extinction eliminates associative strength but leaves the time memory, hence the stimulus representation, intact. Reacquisition proceeds faster because the stimulus representation does not need to be built up again. Other models explain this by allowing the associative strength learning rate to be faster in reacquisition.

Time change in extinction was accounted for because of RWDDM's ability to time CS duration independently from US associations. Time is learned entirely by time markers. The TD models and LeT do not make this separation. These models do not have a mechanism to time stimuli without the US stamping in the changes.

Improved timing in latent inhibition was also accounted by RWDDM's ability to learn timing independently of associations. Preexposure allows the model to build its time representation, which is later expressed by behaviour during the acquisition phase. The only other model that learns to time independently of associations is MoT, but it does not have a mechanism to explain the latent inhibition effect. The latent inhibition effect alone, i.e. the initial decrement in the acquisition curve of a preexposed stimulus, was made possible in RWDDM by using the P-H rule to change the learning rate for associative strength. The use of the P-H rule instead of the RW would certainly have other theoretical implications for the general theory I am introducing in this paper, but I have used it only in this case. I will make further comments in the conclusion. Blocking with different durations was easily accounted in one condition, the short blocked and long blocking CS. The blocking effect in this condition followed from the summation term in the RW rule. For the other condition, long blocked and short blocking CS, a straight application of the model did not yield the results expected. But the experimental results leave open the possibility that this might be a case of second-order conditioning, where the summation term in RW does not play a role. In this case, RWDDM is well placed to

explain the results, since it can time the whole sequence of stimuli. The only other models capable of explaining these results were the TD models.

The time specificity in conditioned inhibition was very well accounted for by the combination of the summation term in the RW rule, which allowed for inhibition to develop, and the independent timing mechanism in RWDDM that allowed it to time US omission. However, the alternative account provided by the different time representation in the TD models was also successful. The other theories failed here for the same reason as in blocking, they lack a rule like RW that can deal with compound stimuli effects.

The response curves centred at the mean of intervals in the VI procedure was well accounted by the ability of RWDDM to learn the average of intervals. This ability is only present in Modular Theory, making it the only other model able to account for the results here.

In the case of temporal averaging, RWDDM was able to account for the general features of the phenomenon, namely a response curve that peaks at the average of the intervals signalled by the compound stimulus. However, RWDDM predicts the peak to be at the harmonic mean, whilst some experimental results suggest it happens at the geometric mean. RWDDM's account of temporal averaging was hypothesised as the result of ambiguity in the signal. In trying to resolve whether the compound should be treated as a single stimulus or as two separate stimuli, the subject settles on using one accumulator that is fed partial timing information from both stimuli. Other hypothesis might turn out to be more adequate, but this is one possibility that fits well with the RWDDM framework. The only other model that would produce averaging under the same hypothesis is MoT.

The classic ISI effect followed from two mechanisms in RWDDM. The lower response curves during longer stimuli were explained by time setting the asymptote of associative learning by hyperbolic delay discounting. The larger spread of response curves during longer stimuli and the superimposition of normalised curves follows from RWDDM's timescale invariant time representation. The noise in RWDDM's accumulator decreases with the interval being timed in such a way that it results in timescale invariance of the response curves. Modular Theory can also reproduce all features in the data. This is because it relies on a timescale invariant response rule function that generates less responding in longer intervals. LeT can account for superimposition, but it does not have a mechanism to account for the lower curves in longer stimuli. MS-TD can account for both elements because of the form of its microstimuli representation.

The double peaks observed in the response curves during mixed FIs is explained by RWDDM using simultaneous timing. It generates two different representations, one for each reward. Thus, it can account for mixed FIs by the same principles used to account for the ISI effect and simple FI schedules. Modular Theory takes the same approach of simultaneous timing and is also successful. The TD models and LeT can provide a partial account due to their distributed time representation. But timescale

invariance of the peaks is not observed in CSC-TD and only approximately in MS-TD. LeT produces the timescale invariance but not the decrease in peak height with time.

The left shift of response curves seen in compound peak procedure and disinhibition of delay was well accounted for by RWDDM. It did so because of intertrial variability in noise estimation. By choosing in every compound trial the time memory that predicts reward sooner, RWDDM produces the left shift in response. The only other models that can appeal to the same principle to explain it are LeT and MoT.

Inhibition in trace conditioning with an excitatory context was well accounted for by RWDDM due to its RW rule and its ability to time any stimulus. The RW rule accounted for the inhibition, due to competition between the CS and the CX, the excitatory context. A well timed response during the gap was due to RWDDM's timer using the CS offset as a cue to start timing. The TD models can explain the increased inhibition only if the γ discount factor is allowed to vary between the excitatory and non-excitatory contexts. This is a plausible mechanism but it adds an extra degree of freedom to these models, decreasing their parsimony. LeT and MoT cannot explain any inhibition because they lack the summation term from the RW rule.

5.2 Comparison with CSC-TD, MS-TD, LeT and MoT

The superiority of RWDDM and MoT in explaining the majority of the phenomena analysed highlights the importance of some of their shared mechanisms. Both models have separate rules for updating time and associative strength. This makes them capable of timing any stimuli, independent of changes in associative strength. Both models represent psychological time as linearly related to physical time through the theoretical construct of the accumulator. Their memory for time stores a moving average of the experienced intervals. They both allow for intertrial variability in time estimation. Among their differences, only one proved crucial in discriminating the two models in the experiments analysed here: the lack of a mechanism in MoT to account for stimulus compounds. RWDDM uses the RW rule, which was developed to deal with phenomena such as blocking and inhibition, whilst MoT uses the linear operator, a historically earlier association rule that cannot handle compounds. This was the single difference that caused the difference between MoT and RWDDM in number of phenomena explained.

MS-TD came in third place in number of phenomena successfully explained, but the gap between it and MoT was comparatively high, with MoT being almost twice more successful than MS-TD. CSC-TD came just half a point below MS-TD. This is certainly a result of their similarities. The only difference between these two TD models is in their time representation. However, this different representation allowed MS-TD to explain only one more phenomenon than CSC-TD, the ISI effect.

Therefore, in the set of experiments analysed here MS-TD did not show a significant improvement on CSC-TD. This does not mean that MS-TD is not a significant improvement on CSC-TD overall. Its superior account of timing is significant. But the set of experiments chosen here are particularly challenging even for a dedicated timing theory, so they raise the bar even higher. The strength of the TD models was in accounting for compound phenomena of blocking and inhibition, due to their RW rule for association. Their weaknesses was that they rely on changes in associative strength to express changes in timing. This prevented them from explaining time change in extinction and improved timing in latent inhibition. They both lack a memory to store the average of intervals, so they could not explain behaviour in VI schedules. Finally, their lack of trial to trial variability in time estimation prevented them from accounting for the left-shift in the compound peak procedure.

With respect to the number of successes only, LeT came in last. The results allowed us to identify at least four limitations in LeT's current formulation. The first is that it ties its time representation to changes in associative strength. This prevented it to explain time change in extinction and improved timing in latent inhibition. The second limitation is that it relies on the linear operator rule for associative strength, which prevented it from accounting for blocking and time specificity in conditioned inhibition. Thirdly, its distributed memory for time does not store the average of the intervals seen. This prevented it from accounting for the behaviour in VI. Lastly, it doesn't have a mechanism to explain the decrease in peak height of the response curves with longer ISIs. However, as a timing model, LeT's strength is in explaining timescale invariance. If it can be made to overcome at least the weakness of its associative learning rule, for example by also adopting the RW to update associative strength, LeT could be on a par with the TD models.

5.3 Limitations and Future Work

RWDDM faced a few problems in explaining the set of phenomena analysed here. In latent inhibition the model was able to learn the timing for the preexposed CS, but our choice of CS representation translates this into a response curve that does not fully match the data. A better solution might involve a two-state CS representation, one state for the early stages of training and the other for the latter stages. RWDDM could not account for the lack of blocking with a long blocked CS and a short blocking CS. One possible solution that does not require changing the model is to treat the blocking CS as a secondary reinforcer. A more difficult problem related to asynchronous co-terminating CSs such as the ones used in the blocking experiment analysed here, is that in its current formulation RWDDM cannot produce a stable solution. Because RWDDM assigns a different learning asymptote for each CS in the compound, it generates an inconsistent system of equations for V. How to fix this remains an open problem. Finally, in temporal averaging RWDDM predicts a peak in CR at the harmonic mean of the intervals, not at the geometric mean as has been

observed in the data. More experiments might help to determine if the harmonic average should indeed be ruled out as an explanation.

One relevant phenomenon that we did not explore here is the peak procedure. In particular, Balci et al., 2009 have produced evidence that in the long peak trials animals don't stop responding immediately after the expected reward time, but instead take a number of peak trials to learn to stop. The Gaussian function $x_i(\Psi_i)$ used as the CS representation in RWDDM ensures that CR levels will begin to decrease after $\Psi_i(t)$ crosses threshold θ without any learning. To address the findings in Balci et al., 2009 the RWDDM CS representation could be changed to a sigmoid, saturating after the timer $\Psi(t)$ crosses a first threshold. A second threshold could then be introduced to mark the time to stop responding. When the timer crosses this stop threshold the saturation process in the CS representation would stop and a decay process would begin. This however would still be an incomplete account, as a mechanism would be needed to explain the learning of the second threshold. But if such a CS representation was used, the model would also fit a larger body of data coming from studies that analyse responding during individual trials of the peak procedure. Schneider, 1969 and subsequently Gibbon and Church, 1990 and others (Cheng and Westwood, 1993; Matell, Bateson and Meck, 2006) have argued that the pattern of responding is better characterized not by a Gaussian but instead by an approximate square-wave function, with a low-high-low response frequency pattern. It can be shown that by introducing a stop threshold to the timer $\Psi_i(t)$, the TDDM timer (used in RWDDM) can fit the data on times of start and stop responding (Luzardo et al., 2017). Alternatively, the accumulator $\Psi_i(t)$ itself could be used as the CS representation, replacing x_i in equations (3.9) and (3.10). In this case, an upper absorbing boundary would need to be set on the accumulator to prevent response strength increasing considerably in the first few trials following a CS duration increase for example. Also, such a choice of CS representation would cause within-trial responding to become linear, rather than the more commonly observed sigmoidal pattern. If a sigmoidal response curve is to be preserved, a different choice of response function would be required.

Another phenomenon that I did not address but deserves mention is the timescale invariance of the acquisition process (Gallistel and Gibbon, 2000). It refers to the general finding that the number of trials required until an acquisition criterion is met depends on the ratio of intertrial (or context) and trial durations, the I/T ratio (Gibbon, 1977; Lattal, 1999; Holland, 2000). Gibbon and Balsam, 1981 provided an account for this that postulates a decision process based on the reward expectancy signalled by the stimulus versus the one signalled by the context. A ratio between the two expectancies is calculated, and once the ratio exceeds a certain value, acquisition starts. If the same postulate of a decision ratio of reward expectancies is made, RWDDM may account for the I/T ratio in a similar manner. If we assume that animals time the interval between USs (the context or I duration) with rate $A_I(n)$ and also the CS duration as usual with rate $A_T(n)$, then we can form the ratio $r(n) = A_T(n)/A_I(n)$. As the number of trials n increases,

the A rates converge to their asymptotic values, and the ratio r will converge to $A_T/A_I = (1/T)/(1/I) = I/T$. This is essentially the same account given by Gibbon and Balsam, 1981, with the timer rates A_T and A_I substituting Gibbon and Balsam's expectancies H/T and H/C.

At least three testable RWDDM predictions came out from the simulations reported here. The first concerns blocking with different durations. A long blocked CS will not be blocked by a short co-terminating blocking CS, and two peaks in responding will be observed during test trials with the blocked CS: one at the time the short blocking CS would normally start, and another at the end of the blocked CS. The second prediction is that conditioned inhibition is the exact opposite of excitation. This means that the behaviour produced by inhibition is timed in the same manner as in excitation. Finally, in temporal averaging the response peak in the compound stimulus should be at the harmonic average, or weighted harmonic average. One prediction that did not come out of the simulations but that is worth mentioning concerns time estimation during very early trials. Our assumption of a low initial value for the accumulator rate A implies that in the initial trials durations will be overestimated. A new experiment testing this prediction could help validate, or invalidate, the model.

RWDDM is, to the best of our knowledge, the first time the RW associative learning rule is coupled with a accumulator-based timing theory. An important implication of this effort for associative learning is that it allows for a richer analysis of the effects of timing in compound stimuli experiments. Here I have analysed blocking and conditioned inhibition, but there is evidence suggesting time may have important effects in other cue-competition phenomena such as overshadowing (Kehoe and James, 1983; Jennings, Bonardi and Kirkpatrick, 2007). Timing effects in compounds has until now received somewhat little attention, with many published experimental studies reporting only aggregate response measures. This is perhaps to be expected, since most associative learning models that can handle compounds do not have any, or a rich enough, time representation. RWDDM is an attempt at filling this theoretical gap.

Another limitation of associative learning models is that they tend to simply postulate the timing features of the stimulus representation, without a detailed account of how these can mechanistically arise and evolve. This is the case with the CS representations of CSC-TD, MS-TD and others like C-SOP (Brandon, Vogel and Wagner, 2003). RWDDM's adaptive timer and time-adaptive CS representation provide a fuller account of the timing mechanism and its dynamics. Another recent model that provides this level of detail is the Timing from Inverse Laplace Transform (TILT, Shankar and Howard, 2012; Howard et al., 2015). It can dynamically develop a timescale invariant representation of stimulus history using a two-layer neural network. It can also reproduce the important I/T ratio conditioning phenomenon, but so far it has only been implemented with the linear operator rule for associative learning, which precludes it from accounting for cue competition phenomena.

The RWDDM architecture suggests that timing is largely independent of the process of association formation and maintenance. Associations however, according to RWDDM, depend on timing both to set the asymptote of associative strength and to build the CS representation so that it can enter into association with the US. Thus, RWDDM implies that interactions between timing and associative learning are mainly one-directional. This appears to match roughly with experimental findings. In a review Kirkpatrick, 2013 found that prediction error influenced measures of time estimation only through changes in reward magnitude and devaluation, whilst effects in the other direction included the appropriate timing of CRs from start of conditioning, trial and intertrial durations affecting strength and probability of CR occurrence, and cues with different temporal information affecting cue competition.

5.4 RWDDM and Machine Learning

Although not the focus of this work, it is interesting to note some implications of RWDDM for the field of machine learning.

RWDDM's learning rule is essentially the RW with hyperbolic delay reward discounting. As such, it is an alternative to the TD reinforcement learning model. It is a real-time model but, unlike TD, it does not require a large number of associative units to represent time. This represents a significant economy in terms of computational resources. However, even though RWDDM only requires one associative unit per stimulus, these units need to be specified in advance, i.e. it still requires a hard-coded temporal representation. This was particularly evident in our simulations of trace conditioning. This procedure required a relatively complex representation of stimuli and timers. The context CX had three distinct timers: one active during the ITI, another active during the CS and the third active during the gap. But these three timers formed part of the same context associative unit. How can the model decide on its own how to temporally 'split' the context in such a way?

In section 2.1.6 I reviewed the LSTM, the only machine learning architecture currently able to learn a temporal representation directly from the data. RWDDM has some similarities with the LSTM. RWDDM's timer is analogous to the recurrent activity of LSTM's memory cell c. The memory cell's activity range between [-1,1] whilst RWDDM's temporal representation x only between [0,1]. To bring RWDDM's activity within the same range one could use the associative strength V multiplicatively, yielding $c_{\text{RWDDM}} = x\dot{V}$. To see how RWDDM could be used as a type of LSTM, consider a time series prediction task where the goal is to predict the next event in the series. Each event serves as the input for the 'RWDDM memory cell', whose output is between [0,1], $y_{\text{RWDDM},t}$, analogous to the LSTM output. The error between $y_{\text{RWDDM},t}$ and the event we are trying to predict is used to adjust the weight of RWDDM (the slope of accumulation), in a similar manner as the forget gate of an LSTM. Furthermore, the weight update rule used by RWDDM (equation (3.4)) provides a more natural estimate of the timing error than LSTM's original cost

function (equation (2.78)). This is because RWDDM's slope update rule is trying to minimize the timing error, i.e. the discrepancy in time between its prediction and the time of the actual event. Its error magnitude is proportional to the size of the timing error, meaning that the number of trials needed to learn the correct time is proportional to the size of the error. This is something that the cost function of a LSTM does not capture. In LSTMs the magnitude of the error is the same whether the prediction was off by one time step or, for example, 100. This slows down LSTM learning considerably.

Although the parallels drawn above between RWDDM and LSTM are interesting to consider in the context of LSTM improvement, they are far from being a recipe to a new model. In particular, they leave open the role played by the input and output gates, if any. They are also not sufficient to predict whether this LSTM/RWDDM hybrid would be able to learn a temporal representation directly from the data. But they may be used to help bring some of the insights gained from computational modelling of animal interval timing and conditioning into LSTM research.

It is also interesting to consider what parallels can be drawn between RWDDM and the current deep learning architectures. As we saw in section 2.1.6, by using backpropagation it is possible to train a perceptron network composed of multiple hidden layers. These deep networks have recently become highly successful in learning data representations. Their powers of representation reside precisely on their deep architecture, with each layer capable of representing separate features of the data. This type of architecture does not appear to lend itself easily to the type of timing data that RWDDM was built to explain. One way to begin to approach this problem is to build a network with a single hidden unit, similar to the one postulated by the S-D model of conditioning (see section 2.1.5 and figure 2.8). A hidden-unit network like S-D can solve negative patterning (and the XOR problem) providing a way to test the model experimentally. Adjusting the weights of this hidden-unit RWDDM can be done by backpropagation so this would not be a challenge. The problem is: how does the hidden unit process the time information coming from the CSs? If the hidden unit does not interfere with timing, then hidden-unit RWDDM would certainly account for negative patterning. But if it does, then there a principle is needed to create the model. But even if a theoretic answer to this problem was found, there does not seem to be any experimental studies on timing in negative patterning to confirm or disprove the theory.

However it is at least conceivable that RWDDM timer units could be connected in series, as a multiple layer perceptron network. For example, an event in the time series would trigger the start of the input timer unit, which would increase and reach its threshold value θ. This would trigger the start of the next (hidden) timer unit, and so on. When the next event in the time series happens, this would generate the error between the time this event happened and the time predicted by the currently running timer. The error would be used to adjust the slope of all timers so that the output of the last timer matches with the data. But it is not clear if this 'deep timer'

network would provide a superior prediction than a single timer.

Chapter 6

Conclusion

In this thesis I introduced a new real-time model for classical conditioning and timing. The model combines elements from two theories, the Rescorla-Wagner conditioning model and the TDDM interval timing theory.

I have simulated the model on 11 conditioning phenomena selected from the literature, which collectively represent a particular challenge for any single model to explain. The model was successful in accounting for 10, and can be made to account for the rest if simple modifications are made. The mechanisms used by other models of similar scope were evaluated to see if they could also account for the data. The model that got closer to this level of success in this set of phenomena was Modular Theory. This was due to MoT and RWDDM having a significant overlap in terms of mechanisms. Both models use an accumulator to mark the passage of time. Both models require only a single associative unit per stimulus that adapts to the temporal information conveyed by the stimulus. Their main difference is that MoT still uses the linear operator rule which precludes it from explaining blocking and other compound phenomena, whilst RWDDM uses the RW rule which can account for those phenomena. The same limitation is faced by TILT, a recent model that I did not analyse but that shows promising results and has desirable timing properties.

RWDDM may be improved in several ways. It is quite likely that the asymptote of learning may not be described by the simple inverse relationship to reinforcement time that I assumed. In some of the experiments modelled here, response peak seemed to decrease slower with ISI than our inverse relationship predicted. Functions other than Gaussians might be used to represent the CS, which could better fit the data in the case of latent inhibition for example. These and other theoretical issues may be better elucidated by new experiments involving compound stimuli and a manipulation of their durations, such as the experiments with blocking, compound peak procedure and temporal averaging analysed here.

I have also adopted the P-H rule in one experiment, but have not explored its application in the others. Making the P-H rule an integral part of RWDDM would add one more parameter but it would also allow RWDDM to account for other preexposure and attentional effects that the rule is designed to account. This is not a difficult modification, and I have already shown it to be feasible.

RWDDM may be regarded, like TD, as a real-time extension of RW. Unlike TD

and LeT, it does not require a number of associative units that grows linearly with time. It adds to RW the powerful timing mechanism of TDDM. But also, by making a link with a version of DDM, it shows that it may be possible to arrive at a unified account of timing, conditioning and decision making.

Bibliography

Allman, Melissa J et al. (2014). 'Properties of the internal clock: first- and second-order principles of subjective time.' en. In: *Annual review of psychology* 65, pp. 743–71. ISSN: 1545-2085. DOI: 10.1146/annurev-psych-010213-115117. URL: http://0-www.annualreviews.org.wam.city.ac.uk/doi/abs/10.1146/annurev-psych-010213-115117.

Alonso, Eduardo and Nestor Schmajuk (2012). 'Special issue on computational models of classical conditioning guest editors' introduction.' In: *Learning & behavior* 40.3, pp. 231–40. ISSN: 1543-4508. DOI: 10.3758/s13420-012-0081-7.

Amundson, J. C. and R. R. Miller (2008). 'CS–US temporal relations in blocking'. In: *Learning & Behavior* 36.2, pp. 92–103. ISSN: 1543-4494. DOI: 10.3758/LB.36.2.92. URL: http://www.springerlink.com/index/10.3758/LB.36.2.92.

Aydin, A and J M Pearce (1995). 'Summation in Autoshaping with Short-Duration and Long-Duration Stimuli'. In: *Quarterly Journal of Experimental Psychology Section B-Comparative and Physiological Psychology* 48.3, pp. 215–234. ISSN: 0272-4995. DOI: 10.1080/14640749508401449.

Balci, Fuat et al. (2009). 'Acquisition of peak responding: What is learned?' In: *Behavioural Processes* 80.1, pp. 67–75. URL: http://linkinghub.elsevier.com/retrieve/pii/S0376635708002222papers2://publication/doi/10.1016/j.beproc.2008.09.010.

Balcı, Fuat and Patrick Simen (2016). 'A decision model of timing'. In: *Current Opinion in Behavioral Sciences* 8, pp. 94–101. ISSN: 23521546. DOI: 10.1016/j.cobeha.2016.02.002. URL: http://www.sciencedirect.com/science/article/pii/S2352154616300249.

Balsam, Peter D, Michael R Drew and C R Gallistel (2010). 'Time and associative learning'. In: *Comparative cognition & behavior reviews* 5, p. 1. ISSN: 1911-4745. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3045055/papers2://publication/uuid/F4402BD3-008F-4A7C-90DF-C03050211FAFhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3045055&tool=pmcentrez&rendertype=abstract.

Balsam, Peter D, Michael R Drew and Cynthia Yang (2002). 'Timing at the Start of Associative Learning'. In: *Learning and Motivation* 33.1, pp. 141–155.

Balsam, Peter D., Stephen Fairhurst and Charles R. Gallistel (2006). 'Pavlovian contingencies and temporal information.' In: *Journal of experimental psychology. Animal behavior processes* 32.3, pp. 284–294. ISSN: 0097-7403. DOI: 10.1037/0097-7403.32.3.284.

Balsam, Peter D and C Randy Gallistel (2009). 'Temporal maps and informativeness in associative learning.' In: *Trends in neurosciences* 32.2, pp. 73–8. ISSN: 0166-2236. DOI: 10.1016/j.tins.2008.10.004. URL: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=19136158&retmode=ref&cmd=prlinkspapers2://publication/doi/10.1016/j.tins.2008.10.004http://www.sciencedirect.com/science/article/pii/S0166223608002798.

Barnet, Robert C., Nicholas J. Grahame and Ralph R. Miller (1993). 'Temporal encoding as a determinant of blocking.' In: *Journal of Experimental Psychology: Animal Behavior Processes* 19.4, pp. 327–341. ISSN: 1939-2184. DOI: 10.1037/0097-7403.19.4.327. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/0097-7403.19.4.327.

Barto, Andrew G. and Richard S. Sutton (1982). 'Simulation of anticipatory responses in classical conditioning by a neuron-like adaptive element'. In: *Behavioural Brain Research* 4.3, pp. 221–235. ISSN: 01664328. DOI: 10.1016/0166-4328(82)90001-8. URL: http://www.sciencedirect.com/science/article/pii/0166432882900018.

Bonardi, Charlotte, Ben Brilot and Dómhnall J. Jennings (2016). 'Learning about the CS during latent inhibition: Preexposure enhances temporal control.' In: *Journal of Experimental Psychology: Animal Learning and Cognition* 42.2, pp. 187–199. ISSN: 2329-8464. DOI: 10.1037/xan0000096.

Brandon, Susan E, Edgar H Vogel and Allan R Wagner (2000). 'A componential view of configural cues in generalization and discrimination in Pavlovian conditioning'. In: *Behavioural Brain Research* 110.1-2, pp. 67–72. ISSN: 01664328. DOI: 10.1016/S0166-4328(99)00185-0. URL: http://www.sciencedirect.com/science/article/pii/S0166432899001850.

Brandon, Susan E., Edgar H. Vogel and Allan R. Wagner (2002). 'Computational Theories of Classical Conditioning'. English. In: *A Neuroscientist's Guide to Classical Conditioning*. Ed. by JohnW. Moore. Springer New York. Chap. 7, pp. 232–310. ISBN: 978-0-387-98805-4. DOI: 10.1007/978-1-4419-8558-3{_}7.

Brandon, Susan E, Edgar H Vogel and Allan R Wagner (2003). 'Stimulus representation in SOP: I. Theoretical rationalization and some implications'. In: *Behavioural Processes* 62.1-3, pp. 5–25. ISSN: 03766357. DOI: 10.1016/S0376-6357(03)00016-0.

Brody, Carlos D. et al. (2003). 'Timing and Neural Encoding of Somatosensory Parametric Working Memory in Macaque Prefrontal Cortex'. In: *Cerebral Cortex* 13.11, pp. 1196–1207. ISSN: 10473211. DOI: 10.1093/cercor/bhg100.

Buhusi, Catalin V. and Warren H. Meck (2000). 'Timing for the absence of a stimulus: The gap paradigm reversed.' In: *Journal of Experimental Psychology: Animal Behavior Processes* 26.3, pp. 305–322. ISSN: 1939-2184. DOI: 10.1037/0097-7403.26.3.305. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/0097-7403.26.3.305.

Buhusi, Catalin V and Nestor A Schmajuk (1999). 'Timing in simple conditioning and occasion setting: a neural network approach'. In: *Behavioural Processes* 45.1-3,

pp. 33–57. ISSN: 03766357. DOI: 10.1016/S0376-6357(99)00008-X. URL: http://www.sciencedirect.com/science/article/pii/S037663579900008X.

Carandini, Matteo and David J Heeger (2012). 'Normalization as a canonical neural computation.' In: *Nature reviews. Neuroscience* 13.1, pp. 51–62. ISSN: 1471-0048. DOI: 10.1038/nrn3136. URL: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=22108672&retmode=ref&cmd=prlinkspapers2://publication/doi/10.1038/nrn3136http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3273486&tool=pmcentrez&rendertype=abstract.

Catania, A C and G S Reynolds (1968). 'A quantitative analysis of the responding maintained by interval schedules of reinforcement.' In: *Journal of the experimental analysis of behavior* 11, pp. 327–383. ISSN: 0022-5002. DOI: 10.1901/jeab.1968.11-s327.

Cheng, Ken and Richard Westwood (1993). 'Analysis of single trials in pigeons' timing performance.' In: *Journal of Experimental Psychology: Animal Behavior Processes* 19.1, pp. 56–67. ISSN: 1939-2184. DOI: 10.1037/0097-7403.19.1.56. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/0097-7403.19.1.56.

Cheng, Ken, Richard Westwood and Jonathon D. Crystal (1993). 'Memory variance in the peak procedure of timing in pigeons.' In: *Journal of Experimental Psychology: Animal Behavior Processes* 19, pp. 68–76. ISSN: 0097-7403. DOI: 10.1037/0097-7403.19.1.68.

Cole, Robert P., Robert C. Barnet and Ralph R. Miller (1995). 'Temporal encoding in trace conditioning'. In: *Animal Learning & Behavior* 23.2, pp. 144–153. ISSN: 0090-4996. DOI: 10.3758/BF03199929.

Dayan, Peter and Yael Niv (2008). 'Reinforcement learning: The Good, The Bad and The Ugly'. In: *Current Opinion in Neurobiology* 18.2, pp. 185–196. ISSN: 09594388. DOI: 10.1016/j.conb.2008.08.003.

De Corte, Benjamin J and Matthew S Matell (2016). 'Temporal averaging across multiple response options: insight into the mechanisms underlying integration'. In: *Animal Cognition* 19.2, pp. 329–342. ISSN: 14359448. DOI: 10.1007/s10071-015-0935-4.

Delamater, Andrew R and Dorie-Mae Nicolas (2015). 'Temporal Averaging Across Stimuli Signaling the Same or Different Reinforcing Outcomes in the Peak Procedure'. In: *International Journal of Comparative Psychology* 28.1. ISSN: 2168-3344.

Denniston, James C. and Ralph R. Miller (2007). 'Timing of omitted events: An analysis of temporal control of inhibitory behavior'. In: *Behavioural Processes* 74.2, pp. 274–285. ISSN: 03766357. DOI: 10.1016/j.beproc.2006.11.003.

Desmond, J. E. and J. W. Moore (1988). 'Adaptive timing in neural networks: The conditioned response'. In: *Biological Cybernetics* 58.6, pp. 405–415. ISSN: 0340-1200. DOI: 10.1007/BF00361347.

Dickinson, Anthony, Geoffrey Hall and N. J. Mackintosh (1976). 'Surprise and the attenuation of blocking.' In: *Journal of Experimental Psychology: Animal Behavior Processes* 2.4, pp. 313–322. ISSN: 1939-2184. DOI: 10.1037/0097-7403.2.4.313.

Drew, Michael R., Carolyn Walsh and Peter D. Balsam (2017). 'Rescaling of temporal expectations during extinction.' In: *Journal of Experimental Psychology: Animal Learning and Cognition* 43.1, pp. 1–14. ISSN: 2329-8464. DOI: 10.1037/xan0000127. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/xan0000127.

Drew, Michael R et al. (2004). 'Temporal specificity of extinction in autoshaping.' In: *Journal of experimental psychology. Animal behavior processes* 30.3, pp. 163–176. ISSN: 0097-7403. DOI: 10.1037/0097-7403.30.3.163.

Drew, Michael R et al. (2005). 'Temporal control of conditioned responding in goldfish.' In: *Journal of experimental psychology. Animal behavior processes* 31.1, pp. 31–9. ISSN: 0097-7403. DOI: 10.1037/0097-7403.31.1.31. URL: http://www.ncbi.nlm.nih.gov/pubmed/15656725.

Eshel, N. (2016). 'Trial and error'. In: *Science* 354.6316, pp. 1108–1109. ISSN: 0036-8075. DOI: 10.1126/science.aal2187. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.aal2187.

Fairhurst, S., C. R. Gallistel and J. Gibbon (2003). 'Temporal landmarks: proximity prevails'. In: *Animal Cognition* 6.2, pp. 113–120. ISSN: 1435-9448. DOI: 10.1007/s10071-003-0169-8. URL: http://link.springer.com/10.1007/s10071-003-0169-8.

Gallistel, C R, Andrew R Craig and Timothy A Shahan (2014). 'Temporal contingency.' In: *Behavioural processes* 101, pp. 89–96. ISSN: 1872-8308. DOI: 10.1016/j.beproc.2013.08.012. URL: http://www.sciencedirect.com/science/article/pii/S0376635713001873.

Gallistel, C R and J Gibbon (2000). 'Time, rate, and conditioning.' In: *Psychological Review* 107.2, p. 289.

Gallistel, C R and John Gibbon (2001). 'Computational Versus Associative Models of Simple Conditioning'. In: *Current Directions in Psychological Science* 10.4, pp. 146–150. ISSN: 0963-7214. DOI: 10.1111/1467-8721.00136. URL: http://journals.sagepub.com/doi/10.1111/1467-8721.00136.

Gallistel, C R and Louis D Matzel (2013). 'The neuroscience of learning: beyond the Hebbian synapse.' en. In: *Annual review of psychology* 64, pp. 169–200. ISSN: 1545-2085. DOI: 10.1146/annurev-psych-113011-143807. URL: http://www.annualreviews.org/doi/abs/10.1146/annurev-psych-113011-143807.

Gers, Felix A., Nicol N. Schraudolph and Jürgen Schmidhuber (2002). 'Learning Precise Timing with LSTM Recurrent Networks'. In: *Journal of Machine Learning Research* 3.Aug, pp. 115–143. ISSN: ISSN 1533-7928. URL: http://www.jmlr.org/papers/v3/gers02a.html.

Gibbon, J (1977). 'Scalar expectancy theory and Weber's law in animal timing'. In: *Psychological Review* 84.3, pp. 279–325.

— (1992). 'Ubiquity of scalar timing with a Poisson clock'. In: *Journal of Mathematical Psychology* 36.2, pp. 283–293.

Gibbon, J and R M Church (1990). 'Representation of time'. In: *Cognition* 37.1, pp. 23–54. URL: http://www.sciencedirect.com/science/article/pii/001002779090017Epapers2://publication/uuid/9AF8B274-28F2-43B1-9C15-F40779F668BA.

Gibbon, J et al. (1997). 'Toward a neurobiology of temporal cognition: advances and challenges'. In: *Current opinion in neurobiology* 7.2, pp. 170–184. URL: papers2://publication/uuid/C75085AE-2E33-4834-9AB4-0A8E74D59132.

Gibbon, John (1971). 'Scalar timing and semi-markov chains in free-operant avoidance'. In: *Journal of Mathematical Psychology* 8.1, pp. 109–138. ISSN: 00222496. DOI: 10.1016/0022-2496(71)90025-3.

— (1991). 'Origins of scalar timing'. In: *Learning and Motivation* 22.1-2, pp. 3–38. ISSN: 00239690. DOI: 10.1016/0023-9690(91)90015-Z. URL: http://www.sciencedirect.com/science/article/pii/002396909190015Z.

Gibbon, John and Peter D. Balsam (1981). 'Spreading associations in time'. In: *Autoshaping and conditioning theory*. Academic Press. Chap. 7, pp. 219–253.

Gibbon, John and Russell M. Church (1984). 'Sources of variance in an information processing theory of timing'. In: *Animal Cognition*. Ed. by H. L. Roitblat, H. S. Terrace and T. G. Bever. Hillsdale, NJ: Erlbaum. Chap. 26, pp. 465–488.

Gibbon, John, Russell M. Church and Warren H. Meck (1984). 'Scalar Timing in Memory'. In: *Annals of the New York Academy of Sciences* 423.1 Timing and Ti, pp. 52–77. ISSN: 0077-8923. DOI: 10.1111/j.1749-6632.1984.tb23417.x.

Gormezano, I., E. J. Kehoe and B. S. Marshall (1983). 'Twenty years of classical conditioning with the rabbit'. In: *Progress in psychobiology and physiological psychology*. Ed. by J. M. Sprague and A. N. Epstein. Vol. 10. New York, NY: Academic Press, pp. 197–275. ISBN: 0-12-542110-9. URL: http://webdocs.cs.ualberta.ca/~sutton/kehoepubs/00000017.PDF.

Grossberg, S and N A Schmajuk (1989). 'Neural dynamics of adaptive timing and temporal discrimination during associative learning'. In: *Neural Networks* 2.2, pp. 79–102. URL: http://www.sciencedirect.com/science/article/pii/0893608089900269papers2://publication/doi/doi:10.1016/0893-6080(89)90026-9.

Guilhardi, Paulo and Russell M. Church (2006). 'The pattern of responding after extensive extinction'. In: *Learning & Behavior* 34.3, pp. 269–284. ISSN: 1543-4494. DOI: 10.3758/BF03192883.

Guilhardi, Paulo, Linlin Yi and Russell M. Church (2007). 'A modular theory of learning and performance'. In: *Psychonomic Bulletin & Review* 14.4, pp. 543–559. ISSN: 1069-9384. DOI: 10.3758/BF03196805.

Hall, Geoffrey (2002). 'Associative Structures in Pavlovian and Instrumental Conditioning'. In: *Stevens' Handbook of Experimental Psychology*. Ed. by Hal Pashler. 3rd ed. Hoboken, NJ, USA: John Wiley & Sons, Inc. Chap. 1. ISBN: 0471214426. DOI: 10.1002/0471214426. URL: http://doi.wiley.com/10.1002/0471214426.

Hall, Geoffrey (2008). 'Pearce-Hall error learning theory'. In: *Scholarpedia* 3.2, p. 5274. ISSN: 1941-6016. DOI: 10.4249/scholarpedia.5274. URL: http://www.scholarpedia.org/article/Pearce-Hall_error_learning_theory.

Harris, Justin A. (2006). 'Elemental Representations of Stimuli in Associative Learning.' In: *Psychological Review* 113.3, pp. 584–605. ISSN: 0033-295X. DOI: 10.1037/0033-295X.113.3.584. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-295X.113.3.584.

Harris, Justin A and Evan J Livesey (2010). 'An attention-modulated associative network.' In: *Learning & behavior* 38.1, pp. 1–26. ISSN: 1543-4494. DOI: 10.3758/LB.38.1.1. URL: http://www.ncbi.nlm.nih.gov/pubmed/20065345.

Haykin, Simon (2009). *Neural networks and learning machines*. 3rd ed. Pearson. ISBN: 978-0-13-147139-9.

Hochreiter, Sepp and Jürgen Schmidhuber (1997). 'Long Short-Term Memory'. In: *Neural Computation* 9.8, pp. 1735–1780. ISSN: 0899-7667. DOI: 10.1162/neco.1997.9.8.1735. URL: http://www.mitpressjournals.org/doi/10.1162/neco.1997.9.8.1735.

Holland, Peter C (1998). 'Temporal control in Pavlovian occasion setting'. In: *Behavioural Processes* 44.2, pp. 225–236. ISSN: 03766357. DOI: 10.1016/S0376-6357(98)00051-5. URL: http://www.sciencedirect.com/science/article/pii/S0376635798000515.

Holland, Peter C. (2000). 'Trial and intertrial durations in appetitive conditioning in rats'. In: *Animal Learning & Behavior* 28.2, pp. 121–135. ISSN: 0090-4996. DOI: 10.3758/BF03200248.

Howard, Marc W. et al. (2015). 'A distributed representation of internal time.' In: *Psychological Review* 122.1, pp. 24–53. ISSN: 1939-1471. DOI: 10.1037/a0037840. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/a0037840.

Hull, Clark L. (1943). *Principles of behavior: an introduction to behavior theory*. New York: Appleton-Century-Crofts, p. 422.

Jazayeri, Mehrdad and Michael N. Shadlen (2015). 'A Neural Mechanism for Sensing and Reproducing a Time Interval'. In: *Current Biology* 25.20, pp. 2599–2609. ISSN: 09609822. DOI: 10.1016/j.cub.2015.08.038. URL: http://dx.doi.org/10.1016/j.cub.2015.08.038.

Jennings, Dómhnall and Kimberly Kirkpatrick (2006). 'Interval duration effects on blocking in appetitive conditioning'. In: *Behavioural Processes* 71.2-3, pp. 318–329. ISSN: 03766357. DOI: 10.1016/j.beproc.2005.11.007. URL: http://linkinghub.elsevier.com/retrieve/pii/S0376635705002299.

Jennings, Dómhnall J., Charlotte Bonardi and Kimberly Kirkpatrick (2007). 'Overshadowing and stimulus duration.' In: *Journal of Experimental Psychology: Animal Behavior Processes* 33.4, pp. 464–475. ISSN: 1939-2184. DOI: 10.1037/0097-7403.33.4.464. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/0097-7403.33.4.464.

Jennings, Dómhnall J et al. (2013). 'The effect of stimulus distribution form on the acquisition and rate of conditioned responding: implications for theory.' In: *Journal of experimental psychology. Animal behavior processes* 39.3, pp. 233–48. ISSN: 1939-2184. DOI: 10.1037/a0032151.

Kamin, Leon J (1968). '"Attention-like" processes in classical conditioning'. In: *Miami symposium on the prediction of behavior: Aversive stimulation*, pp. 9–31.

Kehoe, E. James and E. James (1983). 'CS–US contiguity and CS intensity in conditioning of the rabbit's nictitating membrane response to serial compound stimuli.' In: *Journal of Experimental Psychology: Animal Behavior Processes* 9.3, pp. 307–319. ISSN: 1939-2184. DOI: 10.1037/0097-7403.9.3.307. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/0097-7403.9.3.307.

Kehoe, E. James and Michaela Macrae (2002). 'Fundamental Behavioral Methods and Findings in Classical Conditioning'. In: *A Neuroscientist's Guide to Classical Conditioning*. Ed. by John W. Moore. New York, NY: Springer New York. Chap. 6, pp. 171–231. DOI: 10.1007/978-1-4419-8558-3{_}6.

Killeen, P R and J G Fetterman (1988). 'A behavioral theory of timing.' In: *Psychological Review* 95.2, pp. 274–295. URL: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=3375401&retmode=ref&cmd=prlinkspapers2://publication/uuid/8D608E96-488C-470F-AD23-7F0A65FEF58C.

Kirkpatrick, K (2002). 'Packet theory of conditioning and timing'. In: *Behavioural Processes* 57.2-3, pp. 89–106. URL: papers2://publication/uuid/E0992C65-F4DF-4CDB-A4D6-2047008D8027.

Kirkpatrick, K and R M Church (2003). 'Tracking of the expected time to reinforcement in temporal conditioning procedures'. In: *Learning & Behavior* 31.1, p. 3. URL: papers2://publication/uuid/2DA0E9C3-EB55-437C-84B8-971A77857921.

Kirkpatrick, Kimberly (2013). 'Interactions of timing and prediction error learning.' In: *Behavioural processes* 101C, pp. 135–145. ISSN: 1872-8308. DOI: 10.1016/j.beproc.2013.08.005. URL: http://www.sciencedirect.com/science/article/pii/S0376635713001800http://www.ncbi.nlm.nih.gov/pubmed/23962670.

Kirkpatrick, Kimberly and Russell M. Church (2000). 'Independent effects of stimulus and cycle duration in conditioning: The role of timing processes'. In: *Animal Learning & Behavior* 28.4, pp. 373–388. ISSN: 0090-4996. DOI: 10.3758/BF03200271.

Klopf, A. Harry (1988). 'A neuronal model of classical conditioning'. en. In: *Psychobiology* 16.2, pp. 85–125. ISSN: 0889-6313. DOI: 10.3758/BF03333113.

Komura, Yutaka et al. (2001). 'Retrospective and prospective coding for predicted reward in the sensory thalamus'. In: *Nature* 412.6846, pp. 546–549. ISSN: 0028-0836. DOI: 10.1038/35087595. URL: http://www.nature.com/doifinder/10.1038/35087595.

Lattal, K M (1999). 'Trial and intertrial durations in Pavlovian conditioning: issues of learning and performance.' In: *Journal of experimental psychology. Animal behavior processes* 25.4, pp. 433–450. ISSN: 0097-7403. DOI: 10.1037/0097-7403.25.4.433.

Leak, T M and J Gibbon (1995). 'Simultaneous timing of multiple intervals: implications of the scalar property.' In: *Journal of experimental psychology. Animal behavior processes* 21.1, pp. 3–19. ISSN: 0097-7403.

Leon, Matthew I and Michael N Shadlen (2003). 'Representation of time by neurons in the posterior parietal cortex of the macaque.' In: *Neuron* 38.2, pp. 317–327. URL: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=12718864&retmode=ref&cmd=prlinkspapers2://publication/uuid/1247779A-368C-46DE-8CAE-E39A4A2E24A2.

Lubow, R. E. and A. U. Moore (1959). 'Latent inhibition: The effect of nonreinforced pre-exposure to the conditional stimulus.' In: *Journal of Comparative and Physiological Psychology* 52.4, pp. 415–419. ISSN: 0021-9940. DOI: 10.1037/h0046700. URL: http://content.apa.org/journals/com/52/4/415.

Lubow, Robert E. (1989). *Latent inhibition and conditioned attention theory*. Cambridge University Press, p. 324. ISBN: 0521363071.

Ludvig, Elliot A, Richard S Sutton and E James Kehoe (2008). 'Stimulus representation and the timing of reward-prediction errors in models of the dopamine system.' In: *Neural computation* 20.12, pp. 3034–3054. URL: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=18624657&retmode=ref&cmd=prlinkspapers2://publication/doi/10.1162/neco.2008.11-07-654.

Ludvig, Elliot A., Richard S. Sutton and E. James Kehoe (2012). *Evaluating the TD model of classical conditioning*. DOI: 10.3758/s13420-012-0082-6.

Luzardo, Andre, Elliot A. Ludvig and François Rivest (2013). 'An adaptive drift-diffusion model of interval timing dynamics'. In: *Behavioural Processes* 95, pp. 90–99. ISSN: 03766357. DOI: 10.1016/j.beproc.2013.02.003.

Luzardo, André et al. (2017). 'A drift–diffusion model of interval timing in the peak procedure'. In: *Journal of Mathematical Psychology* 77, pp. 111–123. ISSN: 00222496. DOI: 10.1016/j.jmp.2016.10.002. URL: http://linkinghub.elsevier.com/retrieve/pii/S002224961630102X.

Machado, A (1997). 'Learning the temporal dynamics of behavior.' In: *Psychological Review* 104.2, pp. 241–265. URL: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=9127582&retmode=ref&cmd=prlinkspapers2://publication/uuid/087AAE03-52E2-498A-8A18-92A70CE44EF5.

Machado, Armando, Maria Teresa Malheiro and Wolfram Erlhagen (2009). 'Learning to Time: a perspective.' In: *Journal of the experimental analysis of behavior* 92.3, pp. 423–58. ISSN: 1938-3711. DOI: 10.1901/jeab.2009.92-423. URL: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=20514171&retmode=ref&cmd=prlinkspapers2://publication/doi/10.1901/jeab.2009.92-423http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2771665&tool=pmcentrez&rendertype=abstra.

Mackintosh, N. J. (1975a). *A theory of attention: Variations in the associability of stimuli with reinforcement*. DOI: 10.1037/h0076778.

— (1975b). 'Blocking of conditioned suppression: Role of the first compound trial.' In: *Journal of Experimental Psychology: Animal Behavior Processes* 1.4, pp. 335–345. ISSN: 1939-2184. DOI: 10.1037/0097-7403.1.4.335. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/0097-7403.1.4.335.

Matell, Matthew S, Melissa Bateson and Warren H Meck (2006). 'Single-trials analyses demonstrate that increases in clock speed contribute to the methamphetamine-induced horizontal shifts in peak-interval timing functions.' In: *Psychopharmacology* 188.2, pp. 201–12. ISSN: 0033-3158. DOI: 10.1007/s00213-006-0489-x. URL: http://www.ncbi.nlm.nih.gov/pubmed/16937099.

Matell, Matthew S and Alexandra M Henning (2013). 'Temporal memory averaging and post-encoding alterations in temporal expectation.' In: *Behavioural processes* 95, pp. 31–9. ISSN: 1872-8308. DOI: 10.1016/j.beproc.2013.02.009.

Matell, Matthew S, Jung S. Kim and Loryn Hartshorne (2014). 'Timing in a variable interval procedure: Evidence for a memory singularity'. In: *Behavioural Processes* 101, pp. 49–57. ISSN: 03766357. DOI: 10.1016/j.beproc.2013.08.010. URL: http://www.sciencedirect.com/science/article/pii/S037663571300185X.

Matell, Matthew S and Allison N. Kurti (2014). 'Reinforcement probability modulates temporal memory selection and integration processes'. In: *Acta Psychologica* 147, pp. 80–91. ISSN: 00016918. DOI: 10.1016/j.actpsy.2013.06.006. URL: http://linkinghub.elsevier.com/retrieve/pii/S000169181300139X.

Matell, Matthew S and W H Meck (2000). 'Neuropsychological mechanisms of interval timing behavior.' In: *BioEssays : news and reviews in molecular, cellular and developmental biology* 22.1, pp. 94–103. URL: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=10649295&retmode=ref&cmd=prlinks(null).

Matell, Matthew S and Warren H Meck (2004). 'Cortico-striatal circuits and interval timing: coincidence detection of oscillatory processes.' In: *Brain research. Cognitive brain research* 21.2, pp. 139–170. URL: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=15464348&retmode=ref&cmd=prlinkspapers2://publication/doi/10.1016/j.cogbrainres.2004.06.012.

McClelland, James L. and David E. Rumelhart (1985). 'Distributed memory and the representation of general and specific information.' In: *Journal of Experimental Psychology: General* 114.2, pp. 159–188. ISSN: 1939-2222. DOI: 10.1037/0096-3445.114.2.159. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/0096-3445.114.2.159.

McLaren, I. P. L. and N. J. Mackintosh (2000). 'An elemental model of associative learning: I. Latent inhibition and perceptual learning'. In: *Animal Learning & Behavior* 28.3, pp. 211–246. ISSN: 0090-4996. DOI: 10.3758/BF03200258. URL: http://www.springerlink.com/index/10.3758/BF03200258.

— (2002). 'Associative learning and elemental representation: II. Generalization and discrimination'. In: *Animal Learning & Behavior* 30.3, pp. 177–200. ISSN: 0090-4996.

DOI: 10.3758/BF03192828. URL: http://www.springerlink.com/index/10.3758/BF03192828.

Meck, W H and Russell M Church (1984). 'Simultaneous temporal processing.' In: *Journal of experimental psychology. Animal behavior processes* 10.1, pp. 1–29. ISSN: 0097-7403.

Miller, R R, R C Barnet and N J Grahame (1995). 'Assessment of the Rescorla-Wagner model.' In: *Psychological bulletin* 117.3, pp. 363–86. ISSN: 0033-2909. URL: http://www.ncbi.nlm.nih.gov/pubmed/7777644.

Minsky, Marvin and Seymour. Papert (1988). *Perceptrons : an introduction to computational geometry*. MIT Press, p. 292. ISBN: 0262631113.

Mnih, Volodymyr et al. (2015). 'Human-level control through deep reinforcement learning'. In: *Nature* 518.7540, pp. 529–533. ISSN: 0028-0836. DOI: 10.1038/nature14236. URL: http://www.nature.com/doifinder/10.1038/nature14236.

Mondragón, Esther et al. (2014). 'SSCC TD: a serial and simultaneous configural-cue compound stimuli representation for temporal difference learning.' In: *PloS one* 9.7, e102469. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0102469. URL: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0102469#pone-0102469-g011.

Moore, John W., ed. (2002). *A Neuroscientist's Guide to Classical Conditioning*. New York: Springer-Verlag, p. 323. ISBN: 9781441985583.

Moore, John W., June-Seek Choi and Darlene H. Brunzell (1998). 'Predictive Timing under Temporal Uncertainty: The Time Derivative Model of the Conditioned Response'. In: *Timing of Behavior: Neural, Psychological, and Computational Perspectives*. Ed. by David A. Rosenbaum and Charles E. Collyer. The MIT Press. Chap. 1, pp. 3–34.

Niv, Y (2009). 'Reinforcement learning in the brain'. In: *Journal of Mathematical Psychology*. URL: http://www.sciencedirect.com/science/article/pii/S0022249608001181papers2://publication/uuid/92FE83EF-3E6F-483D-B774-4782DC93D48E.

Ohyama, T and M Mauk (2001). 'Latent acquisition of timed responses in cerebellar cortex.' In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 21.2, pp. 682–90. ISSN: 1529-2401.

Ohyama, Tatsuya et al. (1999). 'Temporal control during maintenance and extinction of conditioned keypecking in ring doves'. In: *Animal Learning & Behavior* 27.1, pp. 89–98. ISSN: 0090-4996. DOI: 10.3758/BF03199434. URL: http://www.springerlink.com/index/10.3758/BF03199434.

Pavlov, I. P. (1927). *Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex*. Oxford: Oxford Univ. Press, p. 430.

Pearce, J M and M E Bouton (2001). 'Theories of associative learning in animals.' en. In: *Annual review of psychology* 52, pp. 111–39. ISSN: 0066-4308. DOI: 10.1146/annurev.psych.52.1.111. URL: http://www.annualreviews.org/doi/abs/10.1146/annurev.psych.52.1.111.

Pearce, J M and G Hall (1980). 'A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli.' In: *Psychological review* 87.6, pp. 532–52. ISSN: 0033-295X. URL: http://www.ncbi.nlm.nih.gov/pubmed/7443916.

Pearce, J M, H Kaye and G Hall (1982). 'Predictive accuracy and stimulus associability: Development of a model for Pavlovian learning'. In: *Quantitative analyses of behavior* 3, pp. 241–256.

Pearce, John M. (1987). 'A model for stimulus generalization in Pavlovian conditioning.' In: *Psychological Review* 94.1, pp. 61–73. ISSN: 0033-295X. DOI: 10.1037/0033-295X.94.1.61. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-295X.94.1.61.

Pearce, John M., David N. George and Aydan Aydin (2002). 'Summation: Further assessment of a configural theory'. In: *The Quarterly Journal of Experimental Psychology: Section B* 55.1, pp. 61–73. ISSN: 0272-4995. DOI: 10.1080/02724990143000171. URL: http://www.informaworld.com/openurl?genre=article&doi=10.1080/02724990143000171&magic=crossref%7C%7CD404A21C5BB053405B1A640AFFD44AE3.

Rakitin, B C et al. (1998). 'Scalar expectancy theory and peak-interval timing in humans.' In: *Journal of Experimental Psychology: Animal Behavior Processes* 24.1, pp. 15–33. ISSN: 0097-7403. DOI: 10.1037/0097-7403.24.1.15. URL: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=9438963&retmode=ref&cmd=prlinkspapers2://publication/uuid/89530A59-6A09-4439-B83F-61EB28CEB739.

Ratcliff, R (1978). 'A theory of memory retrieval.' In: *Psychological Review* 85.2, p. 59. URL: http://psycnet.apa.org/journals/rev/85/2/59/papers2://publication/uuid/C23DBAF7-0C7E-434A-8A3C-D9F0945328B8.

Rescorla, R A (1967). 'Pavlovian conditioning and its proper control procedures.' In: *Psychological review* 74.1, pp. 71–80. ISSN: 0033-295X. DOI: <adata-auto="ep{_}link"href="http://0-dx.doi.org.wam.city.ac.uk/10.1037/h0024109"target="{_}blank"id="linkhttp:dx.doi.org10.1037h0024109"title="http://0-dx.doi.org.wam.city.ac.uk/10.1037/h0024109"data-title="http://0-dx.doi.org.wam.city.ac.uk/10.1037/h0024109">http://0-dx.doi.org.wam.city.ac.uk/10.1037/h0024109.

— (1968). 'Probability of shock in the presence and absence of CS in fear conditioning.' In: *Journal of comparative and physiological psychology* 66.1, pp. 1–5. ISSN: 0021-9940. DOI: <adata-auto="ep{_}link"href="http://0-dx.doi.org.wam.city.ac.uk/10.1037/h0025984"target="{_}blank"id="linkhttp:dx.doi.org10.1037h0025984"title="http://0-dx.doi.org.wam.city.ac.uk/10.1037/h0025984"data-title="http://0-dx.doi.org.wam.city.ac.uk/10.1037/h0025984">http://0-dx.doi.org.wam.city.ac.uk/10.1037/h0025984.

Rescorla, R A and A R Wagner (1972). 'A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement'. In: *Classical Conditioning II Current Research and Theory*. Ed. by A H Black and W F Prokasy.

Vol. 21. 6. Appleton-Century-Crofts. Chap. 3, pp. 64–99. ISBN: 0390718017. DOI: 10.1101/gr.110528.110. URL: http://homepage.mac.com/sanagnos/rescorlawagner1972.pdf.

Rescorla, Robert A. (1969). 'Conditioned inhibition of fear resulting from negative CS-US contingencies.' In: *Journal of Comparative and Physiological Psychology* 67.4, pp. 504–509. ISSN: 0021-9940. DOI: 10.1037/h0027313. URL: http://content.apa.org/journals/com/67/4/504.

— (1988). 'Pavlovian conditioning: It's not what you think it is.' In: *American Psychologist* 43.3, pp. 151–160. ISSN: 1935-990X. DOI: 10.1037/0003-066X.43.3.151.

— (1997). 'Summation: Assessment of a configural theory'. In: *Animal Learning & Behavior* 25.2, pp. 200–209. ISSN: 0090-4996. DOI: 10.3758/BF03199059. URL: http://www.springerlink.com/index/10.3758/BF03199059.

Rescorla, Robert A. and Susan E. Coldwell (1995). 'Summation in autoshaping'. In: *Animal Learning & Behavior* 23.3, pp. 314–326. ISSN: 0090-4996. DOI: 10.3758/BF03198928. URL: http://www.springerlink.com/index/10.3758/BF03198928.

Ricker, Sean T. and Mark E. Bouton (1996). 'Reacquisition following extinction in appetitive conditioning'. In: *Animal Learning & Behavior* 24.4, pp. 423–436. ISSN: 0090-4996. DOI: 10.3758/BF03199014.

Rivest, F and Y Bengio (2011). 'Adaptive Drift-Diffusion Process to Learn Time Intervals'. In: *Arxiv preprint arXiv:1103.2382*. URL: http://arxiv.org/abs/1103.2382.

Rivest, Francois, John F. Kalaska and Yoshua Bengio (2014). 'Conditioning and time representation in long short-term memory networks'. In: *Biological Cybernetics* 108.1, pp. 23–48. ISSN: 03401200. DOI: 10.1007/s00422-013-0575-1.

Rosenblatt, F. (1958). 'The perceptron: A probabilistic model for information storage and organization in the brain.' In: *Psychological Review* 65.6, pp. 386–408. ISSN: 1939-1471. DOI: 10.1037/h0042519. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/h0042519.

Rumelhart, David E., G. E. Hinton and R. J. Williams (1986). 'Learning Internal Representations by Error Propagation'. In: *Parallel distributed processing: explorations in the microstructure of cognition. v. 1, Foundations*. Ed. by David E. Rumelhart, James L. McClelland and PDP Research Group. MIT Press. Chap. 8, pp. 318–362. ISBN: 9780262291408. URL: http://0-cognet.mit.edu.wam.city.ac.uk/book/parallel-distributed-processing.

Savastano, Hernán I. and Ralph R. Miller (1998). 'Time as content in Pavlovian conditioning'. In: *Behavioural Processes* 44.2, pp. 147–162. ISSN: 03766357. DOI: 10.1016/S0376-6357(98)00046-1. URL: http://www.sciencedirect.com/science/article/pii/S0376635798000461.

Schmajuk, Nestor A. and James J. DiCarlo (1992). 'Stimulus configuration, classical conditioning, and hippocampal function.' In: *Psychological Review* 99.2, pp. 268–305. ISSN: 0033-295X. DOI: 10.1037/0033-295X.99.2.268. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-295X.99.2.268.

Schmajuk, Nestor A. and John W. Moore (1988). 'The hippocampus and the classically conditioned nictitating membrane response: A real-time attentional-associative model'. In: *Psychobiology* 16.1, pp. 20–35. ISSN: 0889-6313.

Schneider, Bruce A. (1969). 'A two-state analysis of fixed-interval responding in the pigeon'. In: *Journal of the Experimental Analysis of Behavior* 12.5, pp. 677–687. ISSN: 0022-5002. DOI: 10.1901/jeab.1969.12-677. URL: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1338670&tool=pmcentrez&rendertype=abstracthttp://www.pubmedcentral.gov/articlerender.fcgi?artid=1338670.

Schreurs, B. G. and R. F. Westbrook (1982). 'The effects of changes in the CS-US interval during compound conditioning upon an other wise blocked element'. In: *The Quarterly Journal of Experimental Psychology Section B* 34.1, pp. 19–30. ISSN: 0272-4995. DOI: 10.1080/14640748208400887.

Schultz, W, P Dayan and P R Montague (1997). 'A neural substrate of prediction and reward.' In: *Science (New York, N.Y.)* 275.5306, pp. 1593–1599. URL: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=9054347&retmode=ref&cmd=prlinkspapers2://publication/uuid/B1F98CBD-DA0A-409A-91DC-F842ED2C09F1.

Shankar, Karthik H. and Marc W. Howard (2012). 'A Scale-Invariant Internal Representation of Time'. In: *Neural Computation* 24.1, pp. 134–193. ISSN: 0899-7667. DOI: 10.1162/NECO{_}a{_}00212.

Simen, Patrick et al. (2011). 'A model of interval timing by neural integration.' In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* 31.25, pp. 9238–9253. ISSN: 0270-6474. DOI: 10.1523/JNEUROSCI.3121-10.2011.

Simen, Patrick et al. (2013). 'Timescale Invariance in the Pacemaker-Accumulator Family of Timing Models'. In: *Timing & Time Perception* 1.2, pp. 159–188. ISSN: 2213-445X. DOI: 10.1163/22134468-00002018. URL: http://wrap.warwick.ac.uk/58690/http://booksandjournals.brillonline.com/content/journals/10.1163/22134468-00002018.

Skinner, B. F. and C. B. Ferster (2015). *Schedules of Reinforcement*. B. F. Skinner Foundation, p. 740. ISBN: 0989983951.

Smith, Marius C. (1968). 'CS-US interval and US intensity in classical conditioning of the rabbit's nictitating membrane response.' In: *Journal of Comparative and Physiological Psychology* 66.3, Pt.1, pp. 679–687. ISSN: 0021-9940. DOI: 10.1037/h0026550. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/h0026550.

Staddon, J E R and J J Higa (1999). 'Time and memory: towards a pacemaker-free theory of interval timing.' In: *Journal of the experimental analysis of behavior* 71.2, pp. 215–251. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1284701/papers2://publication/uuid/BDD2DD5B-324A-4E12-B7A6-B8A5964C5643.

Stout, Steven C. and Ralph R. Miller (2007). 'Sometimes-competing retrieval (SOCR): A formalization of the comparator hypothesis.' In: *Psychological Review* 114.3, pp. 759–783. ISSN: 1939-1471. DOI: 10.1037/0033-295X.114.3.759. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-295X.114.3.759.

Sutton, R S and A G Barto (1981). 'Toward a modern theory of adaptive networks: expectation and prediction.' In: *Psychological review* 88.2, pp. 135–70. ISSN: 0033-295X.

Sutton, Richard S (1992). 'Adapting Bias by Gradient Descent: An Incremental Version of Delta-Bar-Delta'. In: *Proceedings of the Tenth National Conference on Artificial Intelligence*, pp. 171–176.

Sutton, Richard S. and Andrew G. Barto (1990). 'Time-Derivative Models of Pavlovian Reinforcement'. In: *Learning and Computational Neuroscience: Foundations of Adaptive Networks*. Ed. by M. Gabriel and J. Moore. The MIT Press. Chap. 12, pp. 497–537.

— (1998). *Reinforcement Learning: An Introduction*. MIT Press, p. 322. ISBN: 0262193981.

Swanton, Dale N, Cynthia M Gooch and Matthew S Matell (2009). 'Averaging of temporal memories by rats.' In: *Journal of Experimental Psychology: Animal Behavior Processes* 35.3, pp. 434–439. ISSN: 1939-2184. DOI: 10.1037/a0014021. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/a0014021.

Swanton, Dale N. and Matthew S Matell (2011). 'Stimulus compounding in interval timing: the modality-duration relationship of the anchor durations results in qualitatively different response patterns to the compound cue.' In: *Journal of Experimental Psychology: Animal Behavior Processes* 37.1, pp. 94–107. ISSN: 0097-7403. DOI: 10.1037/a0020200.

Vogel, Edgar H., Susan E. Brandon and Allan R. Wagner (2003). 'Stimulus representation in SOP: II. An application to inhibition of delay'. In: *Behavioural Processes* 62.1-3, pp. 27–48. ISSN: 03766357. DOI: 10.1016/S0376-6357(03)00050-0. URL: http://www.sciencedirect.com/science/article/pii/S0376635703000500http://linkinghub.elsevier.com/retrieve/pii/S0376635703000500.

Vogel, Edgar H, María E Castro and María A Saavedra (2004). 'Quantitative models of Pavlovian conditioning.' In: *Brain research bulletin* 63.3, pp. 173–202. ISSN: 0361-9230. DOI: 10.1016/j.brainresbull.2004.01.005. URL: http://www.sciencedirect.com/science/article/pii/S0361923004000383.

Voss, Andreas, Markus Nagler and Veronika Lerche (2013). 'Diffusion models in experimental psychology: a practical introduction.' In: *Experimental psychology* 60.6, pp. 385–402. ISSN: 1618-3169. DOI: 10.1027/1618-3169/a000218. URL: http://www.ncbi.nlm.nih.gov/pubmed/23895923.

Wagner, Allan R. (1981). 'SOP: A Model of Automatic Memory Processing in Animal Behavior'. In: *Information Processing in Animals: Memory Mechanisms*. Ed. by Norman E. Spear and Ralph R. Miller. Hillsdale: Psychology Press. Chap. 1, pp. 5–47. ISBN: 0898591570.

Wagner, Allan R and Susan E Brandon (2001). 'A componential theory of associative learning'. In: *Handbook of Contemporary Learning Theories*. Ed. by Robert R Mowrer and Stephen B Klein. New York: Psychology Press. Chap. 2, pp. 23–64.

Whitaker, S, C F Lowe and J H Wearden (2003). 'Multiple-interval timing in rats: Performance on two-valued mixed fixed-interval schedules.' In: *Journal of experimental psychology. Animal behavior processes* 29.4, pp. 277–291. ISSN: 0097-7403. DOI: 10.1037/0097-7403.29.4.277.

— (2008). 'When to respond? And how much? Temporal control and response output on mixed-fixed-interval schedules with unequally probable components.' In: *Behavioural Processes* 77.1, pp. 33–42.

Widrow, Bernard and Marcian E. Hoff (1960). 'Adaptive switching circuits.' In: *1960 IRE WESCON Convention Record*. 4, pp. 96 –104. ISBN: 0-262-01097-6.

Williams, B. A. and M. A. McDevitt (2002). 'Inhibition and Superconditioning'. In: *Psychological Science* 13.5, pp. 454–459. ISSN: 0956-7976. DOI: 10.1111/1467-9280.00480. URL: http://pss.sagepub.com/content/13/5/454.short.

Williams, Douglas A. (2014). 'Building a Theory of Pavlovian Conditioning From the Inside Out'. In: *The Wiley Blackwell Handbook of Operant and Classical Conditioning*. Ed. by Frances K. McSweeney and Eric S. Murphy. John Wiley & Sons. Chap. 2, pp. 27–52. ISBN: 1118468171. URL: https://books.google.com/books?id=ZmWjAwAAQBAJ&pgis=1.

Williams, Douglas A, Kenneth W Johns and Mirna Brindas (2008). 'Timing during inhibitory conditioning.' In: *Journal of experimental psychology. Animal behavior processes* 34.2, pp. 237–46. ISSN: 0097-7403. DOI: 10.1037/0097-7403.34.2.237.

Williams, Douglas A. et al. (2016). 'Intertrial unconditioned stimuli differentially impact trace conditioning'. In: *Learning & Behavior*, pp. 1–13. ISSN: 1543-4494. DOI: 10.3758/s13420-016-0240-3. URL: http://link.springer.com/10.3758/s13420-016-0240-3.

Wittmann, Marc (2013). 'The inner sense of time: how the brain creates a representation of duration'. In: *Nature Reviews Neuroscience* 14.3, pp. 217–223. ISSN: 1471-003X. DOI: 10.1038/nrn3452.

Zimmer-Hart, C L and R A Rescorla (1974). 'Extinction of Pavlovian conditioned inhibition.' In: *Journal of comparative and physiological psychology* 86.5, pp. 837–45. ISSN: 0021-9940.