Chapter 7

Rescorla’s Theory of Conditioning

Rescorla and Wagner (1972) published an influential model of classical conditioning, which has been further refined by Rescorla (2001), introducing as necessary an ogival or cubic function which departs from its original linearity. The question of interest is whether this revised model could be subsumed as a special case of nonlinear psychophysical dynamics. The original model, being linear and deterministic, was widely studied and modified in various ways to try and incorporate stochastic ideas and thus cope with the probabilistic nature of competing response behaviour and response measures (Frey & Sears, 1978; Hanson & Timberlake, 1983).

Alternative models for interaction effects between different cues in making causal judgements have evolved that vary both in their structure and the number of their free parameters, as well as in prior assumptions about what entities may be present in a decision situation. White (2005) gives a review and it is known that effects can be predicted by models better for this purpose than Rescorla’s. White’s algebra is basically built from linear weighted sums of dimensional components and is extended to cover 24 different conditions for combining evidence. He contrasts it with some Bayesian nets used to model decisions made by children (Gopnik,
Sobel, Schulz, and Glymour, 2001). We will not here go beyond considering the Rescorla algebra. Its use in causal learning is reviewed briefly at the end of this chapter.

We will keep Rescorla’s notation for his model throughout:

For a single stimulus

\[
\Delta p_n = \beta (\lambda - p_n) \tag{7.1}
\]

where \( \beta \) is the learning rate parameter, \( p_n \) is the probability of a response on trial \( n \), and \( \lambda \) is the asymptote of learning. The complications begin when compound stimuli, made up of two or more components, are used.

The model is now expressed in terms of \( V_i \), which is the strength of association to stimulus \( i \). \( V \) may be positive or negative and is thus not a probability. The compound stimulus is called \( AX \), and \( V_{AX} \) is the strength of association of the compound stimulus. The linear assumption is, then, that

\[
V_{AX} = V_A + V_X \tag{7.2}
\]

which appears from more recent data to be false. When a stimulus compound was followed by a US, the changes in the strength of each of the components \( A \) and \( X \) were taken to be a function of \( V_{AX} \). This was noted to be an important departure from linearity (Rescorla & Wagner, 1972, p. 76), but different from that in [7.2].

When a compound \( AX \) is followed by \( US_1 \) then

\[
\Delta V_A = \alpha_A \beta_1 (\lambda_1 - V_{AX}) \tag{7.3}
\]

and

\[
\Delta V_X = \alpha_X \beta_1 (\lambda_1 - V_{AX}). \tag{7.4}
\]

where \( 0 \leq \alpha, \beta \leq 1 \). The \( \alpha \) are scalar weights. If a different \( US_2 \) is employed, then the suffices 1 are replaced by 2 in all cases in [7.3] and [7.4]. It is assumed that \( \lambda \) is a function of the magnitude of the US. There are some vaguenesses or, to be more charitable, a lack of formal specification in the model concerning the boundedness of \( \lambda \) and the \( V \)s. A monotonic mapping of \( V \) onto response probabilities was assumed.
It follows that, when both $V_A$ and $V_X$ begin conditioning at zero, that after any large number of conditioning trials with $AX$ reinforced

$$V_X = \frac{\alpha_X}{\alpha_A + \alpha_X} V_{AX} \quad [7.5]$$

which resembles Luce’s version of Individual Choice Behaviour (1959), now known to be in general false.

**Extension to Discrimination Learning**

Suppose that in a series of trials where $AX$ and $BX$ are mixed randomly, $AX$ is always followed by reinforcement and $BX$ is always followed by nonreinforcement, to induce a discrimination between $A$ and $B$. Here $X$ can be thought of as a background or as contextual cues.

$$\Delta V_A = \alpha_A \beta_1 (\lambda_1 - V_{AX}) \quad [7.6]$$
$$\Delta V_X = \alpha_X \beta_1 (\lambda_1 - V_{AX}) \quad [7.7]$$

where $AX$ is reinforced, and

$$\Delta V_B = \alpha_B \beta_2 (\lambda_2 - V_{BX}) \quad [7.8]$$
$$\Delta V_X = \alpha_X \beta_2 (\lambda_2 - V_{BX}) \quad [7.9]$$

where $BX$ is nonreinforced. Together, [7.6, 7.9] is a nine-parameter model; simplifying assumptions by setting all $\alpha$ as 1.0 and $\beta$s = 0.5, and $V$s zero prior to the first learning trial were made. The main point of interest is that $AX$ increases monotonically to an upper asymptote, but $BX$ can variously rise and then fall, or fall from the onset of learning. Heath (1979) found the linear model inadequate and augmented it using stochastic decision processes.

Various probabilistic schedules of reinforcement are used in conditioning, and predictions incorporating $\pi$, the reinforcement probability on a trial, are derived. Consequentially, $\pi$ is also the proportion of reinforced
trials in a closed sequence. The asymptotic value of partially reinforced (i.e. \( \pi \simeq 0.5 \) for each compound stimulus) \( V_{AX}, V_{BX} \) is

\[
V_{asympt} = \frac{\pi \beta_1 \lambda_1 - (\pi - 1) \beta_2 \lambda_2}{\pi \beta_1 - (\pi - 1) \beta_2}
\]

[7.10]

and if \( \lambda_1 = 1.0, \lambda_2 = 0 \) and \( \pi = 0.5 \) then

\[
V_{asympt} = \frac{\beta_1}{\beta_1 + \beta_2}
\]

[7.11]

Rescorla and Wagner derived a diversity of expressions for asymptotic behaviours under assumptions about parameter values and using the linearity assumptions \( V_{AX} = V_A + V_X \) and \( V_{BX} = V_B + V_X \). An attempt to circumvent criticisms of the linearity assumption was suggested by treating the stimuli involved as sets \( V_{AX}, V_{BX}, V_{ABX} \) because \( V_A \cap V_B \neq 0 \). Then \( V_{ABX} \to \lambda_1 \) and \( V_{AX} \to \lambda_2 \) and so for \( V_{BX} \), so that \( V_X \to (2\lambda_2 - \lambda_1) \), with \( V_A, V_B \to (\lambda_1 - \lambda_2) \).

Those details constitute a sufficiency for considering if a parallel with \( \Gamma \) models in NPD is viable.

**Structural Analogies**

The conditioning theory is essentially about short trajectories leading to asymptotic values where the asymptotes are relative response probabilities. In a sense, it is about the orbits of the attractor manifold of the dynamics, but the model’s only dynamical equations are [7.1, 7.3, 7.4]. The data and theory in the source papers do not usually show confidence limits on predictions or data, although these could be presumably be derived with ancillary assumptions about binomial distributions on the \( p_n \).

The entry points for comparison are the equations for two-component mixtures in each theory, in conditioning [7.3, 7.4], and in NPD (Gregson, 1992, p. 24, eqns [2.19, 2.20]) for two continua \( h, i \), (where \( i = \sqrt{-1} \) when it is a multiplier and not a suffix)
Both models are written in difference equation form and both can induce local transient nonmonotonic trajectories in their dependent variables, the comparisons would have to be with Vs in [7.1, 7.3, 7.4] and Y(Re,Im) in [7.12, 7.13]. What is expressed as two stimuli, A, B, and a background, X, in pseudodiscrimination experiments is expressed as two complex variables, Y(Re,Im), with shared Y(Im) in 2D NPD Case 2 theory. The fundamental difference in the objectives of the two theories is that one was written for learning and the other for sensation. The possible comparison arises when responses to mixtures are mediated by nonlinear interactions between continua.

<table>
<thead>
<tr>
<th>Parameter Comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conditioning</td>
</tr>
<tr>
<td>$p_n$</td>
</tr>
<tr>
<td>$\beta$</td>
</tr>
<tr>
<td>$\alpha$</td>
</tr>
<tr>
<td>$\pi$</td>
</tr>
<tr>
<td>$\lambda_1, \lambda_2$</td>
</tr>
<tr>
<td>$V_A$</td>
</tr>
<tr>
<td>$V_{AX}$</td>
</tr>
<tr>
<td>$V_{AB}$</td>
</tr>
<tr>
<td>$[\beta_1, \beta_2]$</td>
</tr>
<tr>
<td>$[a]$</td>
</tr>
<tr>
<td>$[V_A]$</td>
</tr>
<tr>
<td>$[\pi]$</td>
</tr>
<tr>
<td>$[V_{ABX}]$</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Note in the table that $\lambda$ means an asymptote in conditioning theory, but means a cross-coupling in NPD. The two meanings are quite different, but the symbolism has been retained, so that reference to the original sources can be made simpler. We are going to drop the asymptote idea later in revising NPD to incorporate some sort of conditioning process.

To compare the two models, the parameters have to be matched, if they functionally correspond. Such correspondence does not necessarily have to exist for all parameters; that is, their role within the model may be given the same meaning, but obviously their numerical values do not match. As NPD is written as a difference equation, [7.1] may be revised to facilitate the explication and comparison of model differences as

$$V_{n+1} = (1 - \beta^*)\lambda + \beta^*V_n$$  \[7.14\]

where $\beta^* = 1 - \beta$. Equation [7.14] shows that the model is a linear weighted compromise of a postulated asymptote and the present value of $V$. There is no need to put an asymptote into $\Gamma$ [7.12,7.13] because that recursion is self-limiting in $Y(\text{Re})$ within its attractor basin of stability. In the table of Parameter Comparisons near-functional equivalences have been indicated in [..]; such equivalences are not necessarily symmetrical.

The $\Gamma$ recursions are the basis of the observable map $M_\Gamma := a \mapsto Y(\text{Re})$ and the conditioning of the map $M_{\text{cond}} := \pi \mapsto V$, as $a$ and $\pi$ are under experimenter control and functions of $Y(\text{Re})$ and $V$ are treated as response scale values. As $M_\Gamma$ is the result of a complex cubic polynomial function, it is ogival. Rescorla (2001) has now revised $M_{\text{cond}}$ to be ogival because a linear map does not fit data. One may say that end effects at zero or near asymptote exhibit nonlinearities; this is a weak convergence of the two models. The question we now explore is: can the conditioning model be rewritten as a special case of NPD? It could be argued, on the basis of a sort of biological economy, that the same brain has to support the sequential dynamics of the domains of activity, sensation and learning. Any model which is linear in its details can be treated as a local subregion of the dynamics of a more general nonlinear model on the principle of piecewise linearization.
The entry point of the argument is to note that $\Delta Y = f(e, Y(I_m))$ and $\Delta V = g(\alpha, \beta)$ are bivariate operator equations $f, g$ respectively in the two models that control rates of change of their dependent variables. But in conditioning there is an extra variable $\pi$, a variable reinforcement parameter, that has no counterpart in sensation. This is equivalent dynamically to having a nonstationary feedback loop added to the recursion. To recall the distinctions of the introduction, we move from $U$ to $C$ time series.

A simple heuristic way to parallel the reinforcement parameter in an augmented $\Gamma$ is to make $e$ in one-dimensional NPD a function of an external schedule $S_j$. Then

$$\Gamma := Y_{j+1} = -a \cdot (Y_j - 1)(Y_j - i \cdot f(S_j))(Y + i \cdot f(S_j)), \ i = \sqrt{-1} \quad [7.15]$$

The time series generated by [7.15] is

$$Y_{j+1} = f(a, S_j) \quad [7.16]$$

and [7.14] with $\pi$ incorporated as in [10] is

$$V_{n+1} = (1 - \beta')\lambda + \beta'V_n \quad [7.17]$$

where now $\beta' = (1 - \pi\beta)$.

The suggested necessary modification of [7.2] which Rescorla (2001, p. 64) considers for compound stimuli is

$$V_{AAB} = f(V_A) + f(V_B) \quad [7.18]$$

and now this $f$ is sigmoidal, so it would resemble $M_\Gamma^1$.

---

1 The theory variant called $\Gamma V 1$ in Gregson (1988) uses $e_j = f(\Delta^1 a_j)$, which means that response sensitivity is a function of a local rate of change of stimulation. This is a sort of feedback, whereas $\pi$ in [7.10] is not contingent on $\Delta V$. To modify [7.15] to accommodate schedules of reinforcement $S_j$ is therefore not the same as using $\Gamma V 1$. 

Simulation in 2D NPD

The problem now is to see if variations in the parameter values of $2\Gamma$ Case 2 equations can produce some of the same time series for reinforced and unreinforced components of a stimulus mixture. This will be attempted in two stages: first, a purely deterministic treatment of $S = f(j)$ in [7.16], and then a stochastic revision in which the shifts in parameters from trial to trial are probabilistic and not a monotone function of time.

**Deterministic model**

The qualitative pattern it is desired to reproduce, without any structural changes in the $2\Gamma$ model, is that in which the reinforced component slowly rises in output strength with successive reinforcement, but the unreinforced component first rises a little and then falls off to a low asymptote. This rise and then fall in a component which is to be extinguished is supposedly paradoxical for some linear models. It is necessary first to show that a configuration of $2\Gamma$ parameter settings which are not implausible, given the extensive results (Gregson, 1992, 2001) in other sensory (but not learning) situations can yield the apparent paradox.

The parameters in a $2\Gamma$ Case 2 model are set initially as $Y_0(Re, Im)_1, Y_0(Re, Im)_2, a_1, a_2, \lambda_1, \lambda_2, \eta$. Putting $Y_0 = (0.5, \eta)$ for both dimensions sets the process as initially stable. $\eta = 10$ has been fixed throughout, as response delay is not a variable system parameter being modelled. $a_1 = 2.4, a_2 = 3.6$ assumes that when the response to dimension 1 is initially unconditioned then the noise strength $a_2$ is greater. Both the $a$ values are left unchanged, as the actual stimuli do not change, only the relative responses to them. The only parameters available to represent an altered sensitivity to stimuli and interaction between stimuli are $\lambda_1$ and $\lambda_2$. Continuing reinforcement over a long series $J = 1, ..., N$ of trials thus implies here that $S_J = f(\lambda_1, \lambda_2, J)$.

After some exploratory analyses, it is found that setting initial values $\lambda_{0,1} = 0.2$, and $\lambda_{0,2} = 1.15$, and incrementing on each trial

$$\Delta^1\lambda_1 = +.04, \Delta^1\lambda_2 = -.04$$

[7.19, 7.20]
yields Figures 1 and 2. Bounds on $e_1, e_2$ of $e_1 < .45, e_2 > .01$ were set\(^2\), though these are not hit for $e_2$ in the series until $J = 30$.

**Figure 7.1: Time course over trials J of Y(Re)1; reinforced**

![Time course over trials J of Y(Re)1; reinforced](image1)

**Figure 7.2: Time course over trials J of Y(Re)2; extinguished**

![Time course over trials J of Y(Re)2; extinguished](image2)

Figure 1 is an analogy of $V_A$ time series and Figure 2 is an analogy of $V_X$. If an analogy of [7.2] is required then the two curves are summed at any point $J$. This would here yield an almost flat curve for the mixture until $a_1$ dominated, which is implausible unless weighted summation were to be used.

\(^2\) In **Case 2** the cross-coupling is effected by making $e_1 = \lambda_1/a_2$, and $e_2 = \lambda_2/a_1$. It is known from previous psychophysical results that plausible values of $e$ lie in the range $.01 < e < .45$. 
Stochastic models

Now the increments [7.16, 7.17] in $\lambda_1$ or $\lambda_2$ are made with associated probabilities, which adds another two parameters $\pi_1$, or $\pi_2$ to the model, compare [7.10]. There are obviously various ways in which this can be done; one, which creates slight second-order changes to the Figures 7.1 and 7.2, is to write the random variable $\kappa \sim RECT(0, 1)$, and then

\[
\begin{align*}
\text{for reinforcement: } & \quad \lambda_J = \lambda_{J-1} + \kappa \cdot .2 - .05 \quad \text{[7.19]} \\
\text{for extinction: } & \quad \lambda_J = \lambda_{J-1} - \kappa \cdot .2 + .05 \quad \text{[7.20]}
\end{align*}
\]

which produces Figures 7.3 (corresponding to 7.1) and 7.4 (corresponding to 7.2). The shifts to conditioning or extinction now arise sooner and more abruptly and the reinforcement curve runs to an asymptote.

Figure 7.3: Time course of reinforcement with stochastic perturbation

Figure 7.4: Time course of extinction with stochastic perturbation

Note that the initial rise in the extinction curve still exists. This rise exists if there is an appropriate choice of a combination of all of the parameters $a_1, a_2, \lambda_1$ and $\lambda_2$. 
Another much more erratic substitution of a stochastic component on the $\lambda$ values is shown, without any commitment to its plausibility. When conditioning experiments are run, there is no certain smooth improvement in performance, and local failures or relapses can arise, but unless these are reported in detail without averaging over replications (which is not usual) it it not possible to know if they are faithfully modelled in any simulation. However, the next two figures 7.5 and 7.6 depict what can happen. The equations corresponding to but replacing [7.19,7.20] are ($J = 1, ..., 30$)

for reinforcement: $\lambda_J = \lambda_{J-1} + .04 \cdot J(2.0 \cdot \kappa - 1.0)$ [7.21]

for extinction: $\lambda_J = \lambda_{J-1} - .04 \cdot J(2.0 \cdot \kappa - 1.0)$ [7.22]

Figure 7.5: Reinforcement with erratic stochastic effects

![Figure 7.5: Reinforcement with erratic stochastic effects](image)

Figure 7.6: Extinction with erratic stochastic effects

![Figure 7.6: Extinction with erratic stochastic effects](image)
Higher and Indeterminate Dimensionality

The Rescorla-Wagner model has been used in other contexts than the original conditioning paradigms for which it was created. One area of interest is the learning and use of causal inference in humans as a topic in developmental psychology (Gopnik, Glymour, Sobel, Schulz, Kushnir and Danks, 2004). It transpires that the Rescorla-Wagner model can not capture some processes of causal inference, but another approach, based on directed acyclic graph theory, can do better. Such graphs are called Bayes nets when combined with a recursive strategy to choose between various links between nodes. With the subsequent use of Bayes theorem to compute various posterior probabilities associated with different links and subsets of nodes in the graph. Bayes nets were developed in artificial intelligence as one way of solving what has come to be called the inverse problem (Pearl, 1988). This finding about the inadequacy of the Rescorla-Wagner model is pertinent here because, to the extent that the Rescorla-Wagner equations can be treated as analogous to NPD equations, then those also would fail if used without modification as suitable for representing inference judgements.

There are, however, some complications: in equations [7.1-7.22] we have restricted comparisons to what are two-dimensional processes and causality situations generally involve more terms, so that directed paths like $X \rightarrow Y$ may be wrong as failing to reveal that the true situation is $X \perp Y, Z \rightarrow Y, Z \rightarrow Y$ where $\perp$ means causally independent. This is like the situation in statistics where partial correlations and not first-order correlations should be used to reveal psychophysical influences like those mapped in Chapter 1, in Figure 1.3.

Gopnik et al, (2004,p.6) observe pertinently that

Causation is not just correlation, or contiguity in space, or priority in time or all three. Causal structures rarely just involve one event causing another. Instead, events involve many different causes interacting in complex ways. A system for recovering causal structure has to untangle the relations among those
causes and discount some possible causes in favour of others.

Even if the underlying causal relationship is deterministic, the occurrence of other causal factors, which may not be observed, will typically make the evidence for the relationship probabilistic.

All this is true; the a priori indeterminacy of the number of variables, and hence the dimensionality of the system, is critical and almost always greater than two, but it is still not sufficient. The Bayes net approach is acyclic, so there are no feedback loops in it, and the assumptions are that deterministic and stochastic linkages are all that may prevail; there is no provision for some of the characteristic features of nonlinear dynamics, including mixes of fast and slow dynamics, which arise naturally in NPD but not in R-W. White (2005, p. 131) suggests that the RW model can be supplemented by "an account of within-event associations", which would involve more parameters and something like adding fast dynamics. At the same time, it is clear that causal learning involves more than bare psychophysical mappings, that are necessary but not sufficient.

Both Rescorla-Wagner equations and NPD create trajectories of association or estimates of causal strength; they involve changes over time and run asymptotically to equilibria, whereas the Bayes nets need augmentation to do that. Gopnik et al (2004, p.19) show experimentally that there are problems children can solve, involving what is called "backward blocking", that are outside the scope of causal Rescorla-Wagner models but can be handled by some Bayes nets. Backward blocking is the situation where learners decide if some event is a cause of an effect by using information where the event never appears in conjunction with other trials where it does appear. The RW model can be modified by adding axioms that decrease the association of a cue with an outcome when the outcome occurs in the absence of the cue. Also, the RW model requires a prior specification of the direction of putative causal links, whereas Bayes nets (and humans) can, in some circumstances, decide between $X \rightarrow Y$ and $Y \rightarrow X$. It may be important that in using $n\Gamma$ equations we have always that

\[ \text{Cause} \rightarrow a, \ Y \rightarrow \text{Effect} \]
that is, causal direction is expressed in model structure in a different way from that in RW. If the extension to $n \times n \Gamma$ is used (Gregson, 1995), then the structure of an $n \times n$ matrix of $\lambda$ coefficients in [7.12, 7.13] is also needed to be solved for. That corresponds to restricting the (0,1) links in a directed Bayes net.

At the present state of knowledge, it can be safely asserted that none of the competing models have universal validity for explaining how and what causality inferences are actually made by human subjects. Conditioning theory, of which RW is a particular case, does not effectively deal with human contingency judgements (Shanks, 1985), nor with the role of associated implicit stimuli (Miller, Barnet and Grahame, 1995).