CARDIFF UNIVERSITY

Perceptual and functional categorisation in associative learning

Christopher S. Grand

A thesis submitted for the degree of
Doctor of Philosophy

School of Psychology, Cardiff University, Cardiff CF10 3AT

CARDIFF UNIVERSITY

Perceptual and functional categorisation in associative learning

Christopher S. Grand

A thesis submitted for the degree of
Doctor of Philosophy

School of Psychology, Cardiff University, Cardiff CF10 3AT

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

ABSTRACT

This thesis investigated the theoretical processes that underlie perceptual and functional categorisation: perceptual categorisation refers to the process of forming an integrated representation of a pattern of stimulation and functional categorisation refers to the process of integrating otherwise equivalent patterns of stimulation according to their uses or consequences. Investigation of perceptual categorisation in people and of functional categorisation in rats provided results that place important constraints on the nature of the involvement of elemental and configural processes.

DECLARATION

This work has not previously been accepted in any substance for any degree and is not being currently submitted in candidature for any degree.

Signed 4/10/07 (candidate)

Date C.S. Grand

STATEMENT 1

This thesis is the result of my own investigation, except where otherwise stated. Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed 4/10/07 (candidate)

Date C.S. Grand

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and inter-library loan, and for the title and summary to be made available to outside organisations.

Signed 4/10/07 (candidate)

Date C.S. Grand

CONTENTS

ACKNOWLEDGEMENTS 1

CHAPTER 1: INTRODUCTION 3

1.2. Perceptual categorisation 5

1.3. Functional Categorisation 12

CHAPTER 2: PERCEPTUAL CATEGORISATION 22

2.1. Introduction 22

2.2. Experiments 1 and 2 24

2.3. Experiment 3 35

2.4. Experiment 4 42

2.5. General Discussion 49

CHAPTER 3: FUNCTIONAL CATEGORISATION I 52

3.1. Introduction 52

3.2. Experiment 5 54

3.3. Experiment 6 66

3.4. Experiment 7 76

3.5. General Discussion 85

CHAPTER 4: FUNCTIONAL CATEGORISATION II 90

4.1. Introduction 90

4.2. Experiments 8 and 9 90

4.4. Discussion 100

CHAPTER 5: GENERAL DISCUSSION 102

5.1. Summary of rationale for the new research presented in this thesis 102

5.2. Perceptual categorisation 103

5.3. Functional categorisation 106

5.4. Concluding comments 110

REFERENCES 112

APPENDIX 1 117

APPENDIX 2 118

FIGURE LIST

Figure 1. Summary of the design of Rescorla and Furrow's (1977) experiment. 7

Figure 2. An elemental analysis of how similarity can promote association. 9

Figure 3. A configural account for how similarity can promote association. The associative structures for first-order conditioning (left) and second-order conditioning (right). 10

Figure 4. A representation mediated account of acquired equivalence. 14

Figure 5. A connectionist account for the results observed by Honey and Watt (1998). 16

Figure 6. A configural analysis of patterning (A+, B+, AB-; left-hand panel) and the same model following a revaluation procedure in which A is paired with shock after initial negative patterning training (right-hand panel). 19

Figure 7. Patterns used in Experiments 1–4. Each pattern was constructed from two squares and each square contained a pattern that was a mirror image of the other. The grey internal features of both squares served as A, B, C and D and the black external features served as X, Y and Z (or S, T and U in Experiment 4). The letter combinations above each pattern (i.e., AX–DZ) illustrate one counterbalanced sub-condition from Experiments 1 and 2. 26

Figure 8. Experiment 2. Mean percentage correct for the first patterns (A and C; left panel) and second patterns (B and D; right panel). 33

Figure 9. Experiment 3. Mean percentage correct for the first patterns (A and C; left panel) and second patterns (B and D; right panel). 38

Figure 10. An analysis of elemental (left) and configural (right) models of learning in human similarity experiments. 41

Figure 11. The additional patterns used in Experiment 4. Each pattern was constructed in the same way as those used in Experiments 1-3 and the grey internal

features again served as A, B, C and D and the black external features as S, T and U (or X, Y and Z). The letter combinations above each pattern (i.e., AS–DU), together with those in Figure 1 (i.e., AX–DZ), illustrate one counterbalanced sub-condition from Experiment 4. 44

Figure 12. Experiment 4. Mean percentage correct for the first patterns (S1 and D1; left panel) and second patterns (S2 and D2; right panel). 48

Figure 13. Predictions derived from configural and elemental accounts of acquired equivalence/distinctiveness. 53

Figure 16. A possible modification to a configural model that allows it to provide a potential account for the results of Experiment 5. The dotted link illustrates the reciprocity of one of the links that are presumed to form between the hidden layer units and the output units. 65

Figure 17. Modified configural and elemental accounts of the structures formed after patterning training (summarised in Table 10). Note that for presentational purposes, the patterning and control conditions are presented separately. 68

Figure 18. Experiment 6: Magazine entry scores (in responses per minute; RPM) during training for groups Positive patterning (top two panels) and Negative patterning (bottom two panels). 71

Figure 19. Experiment 6. Mean levels of activity when rats in groups Positive and Negative patterning were placed in element contexts B and D (upper two panels) and compound contexts AB and CD (lower two panels). 74

Figure 20. Application of elemental and configural models for Experiment 7. U1 and U2 refer to different unique elements. Direct links between the various input units and shock are not shown in the left-hand panel. 78

Figure 21. Experiment 7: Magazine entry scores (in responses per minute; RPM) during training for groups Positive patterning (Top two panels) and Negative patterning (bottom two panels). 81

Figure 22. Experiment 7. Mean levels of activity when rats in group Positive and Negative patterning were placed in contexts B and D (upper two panels) and AB and CD (lower two panels). 84

Figure 23. A dual unit model. A potential neural network model for functional grouping after patterning training with standard (sometimes AND-type) and XOR-type hidden-layer units. 87

Figure 24. Application of dual unit model to Experiments 6 and 7. 89

Figure 26. Experiment 8. Mean levels of test activity when rats in group Patterning and Control were placed in contexts B/D and AB/CD. 96

Figure 27. Experiment 9. Magazine entry scores (in responses per minute; RPM) during training. 98

Figure 28. Experiment 9. Mean levels of activity when rats were placed in contexts B/D and AB/CD. 99

TABLES LIST

Table 1: Design of the study reported by Rescorla and Gillan (1980)........................8

Table 2: Design of the study conducted by Honey and Watt (1998). 15

Table 3: Design of negative and positive patterning..18

Table 4: Design of Experiments 1 and 2. 25

Table 5: Mean reaction times in Experiment 2. 34

Table 6: Mean reaction times for Experiment 4..39

Table 7: Sequence types used in Experiment 4..43

Table 8: Mean reaction times for Experiment 4..48

Table 9: Design of Experiment 5. 54

Table 10: Design of Experiment 6. 66

Table 11: Design of Experiment 7. 76

Table 12: Design for Experiment 8. 91

Table 13: Design of Experiment 9. 91

Table 14: Design of proposed habituation experiment. 105

Table 15: Assessment of performance to second patterns. 106

PUBLICATIONS

Paper:

Grand, C. Close, J. Hale, J. and Honey, R.C. (2007). The role of similarity in human associative learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *33*, 64-71.

Talks:

Grand, C., Close, J., & Honey, R.C. (2005). Similarity and association. *Associative Learning Symposium*. Gregynog, Wales.

Grand, C., Close., J, Hale, J., & Honey, R.C. (2006). Similarity and association in simple and configural discriminations. *Associative Learning Symposium*. Gregynog, Wales.

Grand, C., & Honey, R.C. (2006). How are nonlinearly separable discriminations acquired? *Associative Learning Symposium*. Gregynog, Wales.

Grand, C., & Honey, R.C. (2006). How are nonlinearly separable discriminations acquired? Associative Learning and Reinforcement Learning workshop at the *Adaptation in Artificial and Biological Systems* meeting. Bristol, England.

ACKNOWLEDGEMENTS

Particular debt is owed to my supervisors Rob Honey and Simon Killcross for supporting the research both with their knowledge and enthusiasm, to my parents for listening and debating for many hours of arguments on the subjects of this thesis and to James Close, Jessica Hale and Melissa Allman for helping to conduct various parts of the experiments described within.

Perceptual and functional categorisation in associative learning

CHAPTER 1: INTRODUCTION

1.1. General Introduction

The ability of an animal to adapt to its environment confers obvious benefits for its survival. A powerful yet simple form of adaptation can be generated by allowing the central processes activated by the stimuli in an animal's environment to become linked to motor programs that underlie that animal's behavioural repertoire. That is, adaptation can be achieved by allowing the processes activated by a stimulus to become linked, by association, to those processes responsible for generating behaviour. Provided the animal is either operating only in a very simple environment or is in possession only of very limited senses (e.g., *Aplysia*) simple, stimulus→response (S→R), learning will suffice. Once more complex senses have evolved, this simple form of learning will become a less efficient means of adapting to an environment. For example, as the number of sensations that an animal can discriminate increases, so too does the number of potential associative connections that can form from a given learning episode. Indeed, the number of connections increases in a non-linear fashion: for example, in an animal which can discern only two elements (e.g., light and dark) and possesses only two motor programs (withdraw or approach), six single (i.e., bi-directional) links can form, but if a single further element and motor program is added, then fifteen links can form, and so on.

In addition to the issue identified immediately above, S→R models of learning fail to categorise stimuli in sensible ways under certain circumstances. To take a concrete example, a banana is 'an' object but it has many elements (colour, curvature, texture, smell), some of which can vary greatly depending on vantage point. According to an S→R analysis, learning to respond to a banana simply involves attaching each of these elements to an appropriate response (e.g., approach, consume). This form of analysis has undoubtedly enjoyed a great deal of success (Hull, 1943; Spence, 1936), as have more recent elemental theories of learning (Rescorla & Wagner, 1972; Wagner, 1981). However, as general accounts of behavioural adaptation such models have limitations. For example, following exposure to an object (e.g., a banana) there has been no re-organisation of its sensory elements that would constitute the formation of what might be considered to be an integrated representation. This observation has important consequences. Thus, if a subset of the elements of a banana was experienced and the ingestional consequences were adverse, an S→R animal would, nevertheless, consume the remaining parts of the banana if those parts activated a set of elements that differed from those experienced prior to the adverse consequences. That is, there is no sense in which the animal has a representation of banana that can enter into *other* associations independently of the responses that have become established to its elements. It would clearly be advantageous for animals to form such an integrated representation to allow further learning about the elements of that object (e.g., that it is no longer edible) to generalise to its remaining elements.

Recent analyses of associative learning have placed greater emphasis on the process of "category formation" (see, e.g., Gluck & Myers, 1993; Honey & Ward-Robinson, 2002; McLaren, Kaye & Mackintosh, 1989; McLaren & Mackintosh, 2002) identified in the previous paragraph. In the following sections, I distinguish between two sorts of categorisation: perceptual and functional. Herein, perceptual categorisation refers to the process of forming an integrated representation of a pattern of stimulation and functional categorisation refers to the process of integrating otherwise equivalent patterns of stimulation according to their uses or consequences.

1.2. Perceptual categorisation

It is well established that if a response is conditioned to one stimulus, or stimulus-complex, then other stimuli will, to a certain extent, also elicit that response (Pavlov, 1927; for a review, see Mackintosh, 1974). The functional utility of stimulus generalisation is clear: having learnt that one plant was edible and nutritious, it is likely that other plants that look similar (that is, have elements in common) will also prove to be edible and nutritious. An experimental example of stimulus generalisation is provided by Guttman and Kalish (1956) who trained pigeons to peck a key with a light of a given wavelength (to get a reward) and then received test trials with differing wavelengths. Guttman and Kalish (1956) observed that the closer the test wavelength was to the original training stimulus, the more vigorous was the response.

Stimulus generalisation gradients of the kind described by Guttman and Kalish (1956) are open to just the form of S→R analysis described above: to the extent that two patterns of input excite the same elements, generalisation will occur. However, this analysis does not anticipate the fact that the amount of stimulus generalisation can be influenced by the animal's experience with the stimuli. For example, rats that have received exposure to a stimulus compound (AB) will subsequently show greater generalisation between A and B than those who have not; an effect known as 'sensory preconditioning' (e.g., Brogden, 1939; Rescorla & Cunningham, 1978). Also, Gibson and Walk (1956) demonstrated that rats which had received a period of exposure to two similar stimuli subsequently learnt to discriminate between them more rapidly than those that had not; an effect known as 'perceptual learning'. Perceptual learning has been the subject of a great deal of recent interest (e.g., Hall, 2003; McLaren & Mackintosh, 2000, 2002). These exposure learning phenomena clearly suggest that experience with a stimulus results in the formation of some internal representation that is simply not captured by an S→R analysis. Sensory preconditioning, in particular, suggests that simultaneous exposure to the elements of a stimulus results in them coming to activate the same representation. Sensory preconditioning can also be observed when the stimulus elements are presented successively rather than simultaneously (e.g., Brogden, 1939; Rescorla, 1980). Successive sensory preconditioning can be allied to the process of binding together successive views of the same object; for example, when one aspect of an object is occluded, or when the object subtends a large visual angle. In both cases, one needs to bind elements that are only successively presented.

There are a variety of possible accounts of sensory preconditioning. For example, during exposure to a compound (e.g., AB or A→B) an association between A and B might form. When B is subsequently paired with an aversive event, this will, at test, allow A to activate a memory of the same aversive event by virtue of the associative chain A→B→aversive event. A second possibility, most often entertained in the context of simultaneous exposure, is that A and B might come to activate the same 'configural' representation and that this representation allows what is learnt about B to transfer to A (e.g., Rescorla, 1980). In fact, there is other evidence that is consistent with such a configural analysis from the conditions under which successive associations are formed. In particular, there is some evidence that similarity promotes (associative) learning. However, it should be acknowledged that such evidence and the theoretical analysis of the role of similarity in learning is limited. One of the aims of the empirical work presented in this thesis (in Chapter 2) is to examine the role of similarity in learning in a novel preparation. This work, conducted in human participants, builds on work conducted using pigeons by Rescorla and his colleagues (Rescorla & Furrow, 1977; Rescorla & Gillan, 1980).

1.2.1. Similarity and association

Rescorla and Furrow (1977) used a second-order autoshaping procedure to study the role of similarity in association formation. In their studies, the design of which is summarised in Figure 1, pigeons received first-order conditioning with green and vertical stimuli, in which both stimuli were separately presented and paired with food. They then received either similar second-order pairings (blue→green and horizontal→vertical) or dissimilar pairings (blue→vertical and horizontal→green).

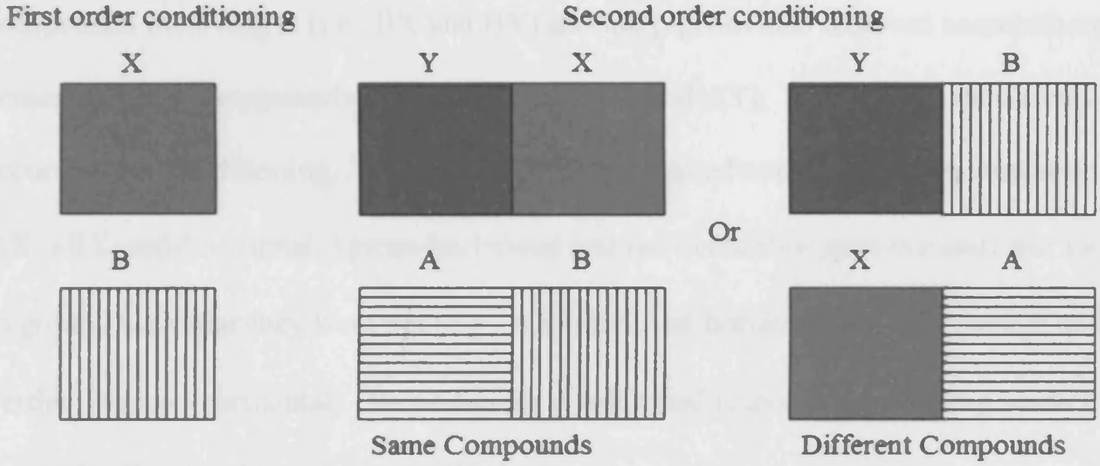


Figure 1. Summary of the design of Rescorla and Furrow's (1977) experiment.

Second-order conditioned responding was more evident when similar stimuli were paired than when dissimilar stimuli were paired. These results suggest that similarity promotes learning (see also, Holland & Ross, 1981; Testa & Ternes, 1977; see for a review, Hall, 1994). The pattern of results reported by Rescorla and Furrow (1977) has been confirmed by Rescorla and Gillan (1980), who made use of stimuli that could be artificially constructed to be more or less similar to one another (for the experimental design that they employed, see Table 1).

Table 1: Design of the study reported by Rescorla and Gillan (1980).

Group	**First order**	**Second order**
Similar	BX+, BY+ AX-, AY-	AX → BX AY → BY
Dissimilar	BX+, BY+ AX-, AY-	AX → BY AY → BX

Note: A, B, X and Y were stimuli and + and - refer to food and no food, respectively.

Thus, during first-order conditioning conditioned responding was established to compounds involving B (i.e., BX and BY) and the pigeons also received nonreinforced presentations of compounds involving A (i.e., AX and AY). The pigeons then received second-order conditioning. For group Similar, the paired compounds were similar (e.g., AX→BX; red-horizontal→green-horizontal and red-vertical→green-vertical) and for those in group Dissimilar they were not (e.g., AX→BX; red-horizontal→green-vertical and red-vertical→green-horizontal). Second-order conditioned responding was more marked in group Similar than it was in group Dissimilar.

1.2.2. Elemental and configural accounts

There are two classes of model of associative learning that can account for the fact that similarity promotes association: elemental and configural. The elemental analysis presented by Rescorla and Gillan (1980) is depicted in Figure 2.

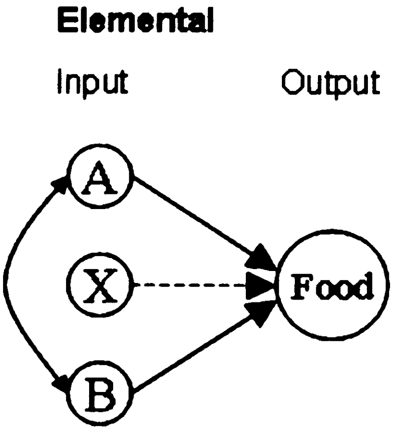


Figure 2. An elemental analysis of how similarity can promote association.

Rescorla and Gillan (1980) argued that when two similar stimulus compounds are paired (e.g., red-horizontal→green-horizontal; AX→BX), the fact that horizontal (X) is present during the first compound (AX) might render it poorly processed (dotted line) and green particularly well processed during the second compound (BX; cf. Wagner, 1981). The operation of this (short-term habituation) process should allow A to be selectively associated with the well-processed B. Given the observation that Rescorla and Gillan's first-order conditioning phase would have left B with greater associative strength than X (again represented by a continuous as opposed to dotted line), it is to be expected that a stimulus (in this case A) that has an effective association with B will elicit a great deal of second-order conditioned responding. One straightforward reason why B will have greater associative strength is because it is continuously reinforced and X is partially reinforced. The situation will be quite different during the pairing of dissimilar compounds (e.g., red-horizontal→green-vertical; AX→BY). Without the bias in the processing of the elements of the second compound (now both elements will be equally well or poorly processed), both will be processed to an equal extent and A will not become especially well associated with B. This form of analysis receives support from a series of supplementary observations made by Rescorla and Gillan (1980); namely, that the tendency for A to elicit responding was greater in group 'similar' than in 'dissimilar' when it was arranged that B elicited more responding than X, but that this pattern of results was reversed when it was arranged that X elicited more responding than B.

The conclusion that Rescorla and Gillan (1980) reached was that the influence of similarity upon association formation can be reduced to the operation of another well known principle of association, that of intensity; albeit one in which intensity is not a simple function of the physical intensity of a stimulus. There is, however, an alternative configural analysis for the basic finding that similarity promotes association (although not Rescorla and Gillan's supplementary observations, but see below) that is depicted in Figure 3 and has not been the subject of any experimental scrutiny.

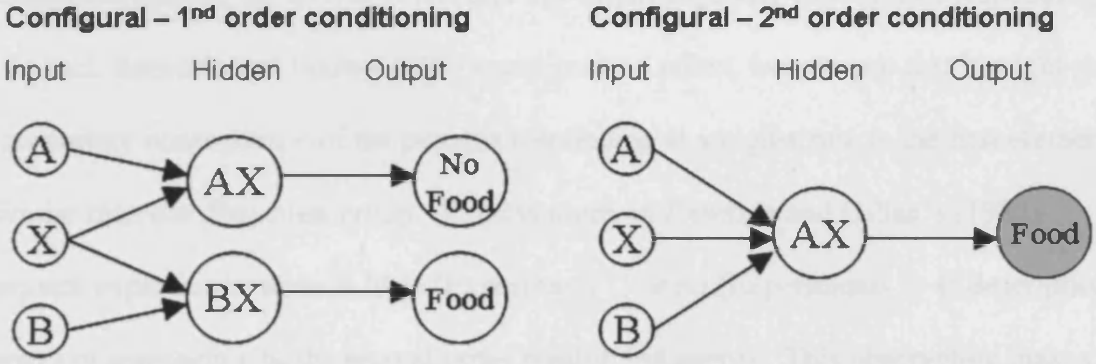


Figure 3. A configural account for how similarity can promote association. The associative structures for first-order conditioning (left) and second-order conditioning (right).

According to this analysis, during first-order conditioning, the input units activated by B and X become linked to the same configural, or hidden unit; one that differs from that activated by A and X. During second-order conditioning, when AX is presented before BX, AX will activate its unit (AX) and also, to some degree, the unit activated by BX. This will mean that BX will come to fully activate its hidden unit rapidly and the residual activity in input unit A will allow A to become linked to BX. This effect will be less likely to operate in the dissimilar case, where presenting AX cannot increase the ease with which BY will activate its unit and thereby foster an association between A and the BY unit; indeed the presentation of AX might tend to interfere with the ability of BY's hidden unit to become active.

It is perhaps worth noting that while Rescorla and Gillan's (1980) elemental analysis relies on one modified principle of association (namely that of intensity), the configural analysis that I have just offered relies on the effects of residual activity within a configural system. It should be immediately apparent that the configural analysis predicts that there might well be differences not only in responding to the first elements of the sequences presented during second-order conditioning, but also to the second elements. Thus, if AX and BX are linked to the same representation (i.e., ABX) then AX will activate this representation prior to the arrival of BX and this might engender conditioned responding to BX. In fact, Rescorla and Furrow (1977) note such an effect, but assume that it might simply be a secondary consequence of the pigeons responding at a higher rate to the first elements in the similar than the dissimilar group. What is more, in Rescorla and Gillan's (1980) subsequent experiments there is little (Experiment 1) or no (Experiments 2–4) description of the levels of responding to the second-order reinforcing agents. This observation makes interpretation of their findings difficult. It also leaves open the possibility that the supplementary observations that support an elemental analysis (that similarity impairs learning under some conditions) are open to alternative, less interesting interpretations (see below).

For example, in Rescorla and Gillan's (1980) Experiment 4, pigeons first received the following training trials: AX→no food, BX→no food, A→food and B→food. During second-order conditioning group 'similar' received AX→A and BX→B, whereas group 'dissimilar' received AX→B and BX→A. It was observed that pigeons showed greater responding to AX and BX in the dissimilar than in the similar group. This result is consistent with the suggestion that when A and B are poorly processed after AX and BX respectively, little second-order conditioning is observed. It is also inconsistent, taken at face value, with the configural analysis outlined above. Of course, it is possible that there is more than one process contributing to the effect of similarity on association (an issue that will be revisited in Chapters 2 and 5). Moreover, the results of Experiment 4 could be a simple artefact of there being much less responding to A and B in the similar than the dissimilar group: under these circumstances one would anticipate, for all kinds of reasons, less second-order conditioning in the similar than dissimilar case.

1.2.3. Rationale for further research

The elemental and configural analyses described above suggest the need for further work of an empirical nature. What is more, it is undoubtedly possible that the results of the experiments described above might have limited generality and the theoretical analyses offered for them might not apply to other experimental preparations and species. To this end, I conducted a series of four experiments with human participants to examine the role of similarity in associative learning, using a novel procedure. These experiments will be reported in Chapter 2 and should reveal important information about one of the general topics that I am interested in (i.e., perceptual categorisation). For now, I will proceed by discussing another important type categorisation: functional categorisation.

1.3. Functional Categorisation

One example of a functional category, which presumably arises as the result of experience, is clothing: such items can be perceptually distinct (e.g., hats and Wellington boots) but have the same use (to keep the elements at bay) or consequences (the promotion of dryness and warmth). There is also the complementary possibility that perceptually similar stimuli (e.g., different red wines) can become part of different functional categories ("burgundy" and "claret"; see James, 1890). Experimental evidence of functional categorisation comes from demonstrations of the acquired equivalence and distinctiveness of cues.

1.3.1. Acquired equivalence and distinctiveness

A clear empirical demonstration of the operation of the acquired equivalence and distinctiveness of cues process has been reported by Honey and Hall (1989; see also, for example, Zentall, Steirn, Sherburne, & Urcuioli, 1991; for a review, see Hall, 1991). In their study, rats were presented with two auditory stimuli (A and B), paired with food, and a third auditory stimulus (C), paired with no food. Subsequently, A was paired with the delivery of shock and the rats were then tested with B and C. As a consequence of this revaluation treatment, B elicited more fear than C. This result clearly demonstrates that the extent of generalisation between two stimuli is modifiable by experience and, in particular, supports the suggestion that stimuli with a common outcome (i.e., A and B) become treated as one category and stimuli associated with a different outcome (i.e., C) become treated as another category.

1.3.2. Elemental and configural accounts

There has been a variety of attempts to capture this type of effect in associative terms. Application of such analyses is best understood with reference to the study by Honey and Hall (1989). One of the earliest associative accounts relied on the possibility that response-produced cues (e.g., the sight of the food magazine when A is presented) acquire associative strength (e.g., become connected to the response elicited by shock) and thereby mediate generalisation between A and B (e.g., Hull, 1943). That is, response-produced cues simply add to the number of common elements that two stimuli activate. This possibility is less plausible when there is no obvious reason for the training procedure to engender different responses to A, B and C; and acquired equivalence and distinctiveness effects have been observed using a sensory preconditioning procedure using flavours (see, for example, Honey & Hall, 1991). A related, elemental account to that described by Hull (1943), presented by Honey and Hall (1989), was based on the idea of representation-mediated conditioning (Holland, 1981). This account is depicted in Figure 4.

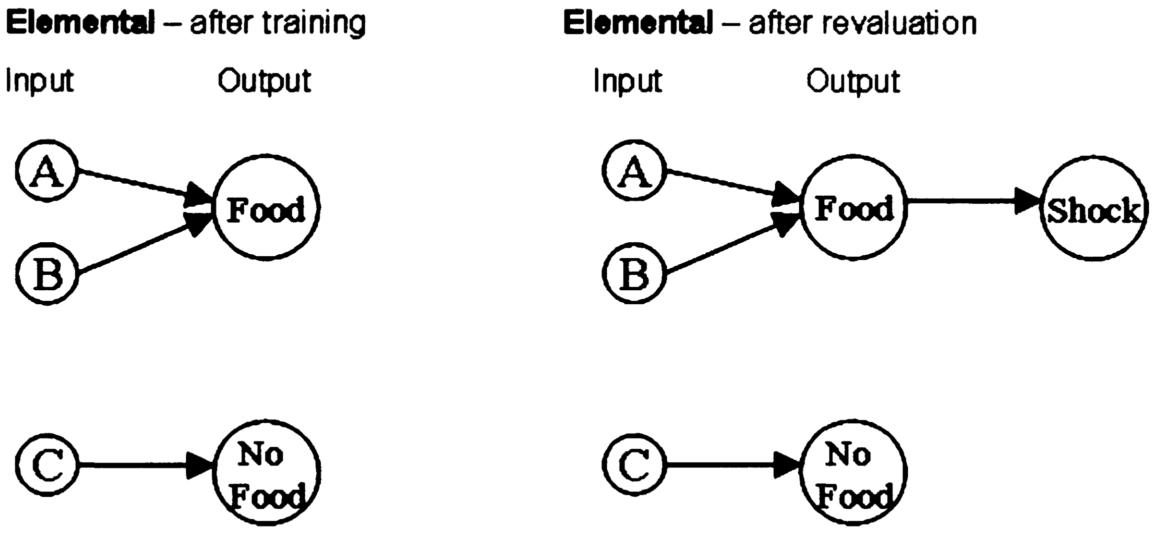


Figure 4. A representation mediated account of acquired equivalence.

According to this analysis, appetitive conditioning endows A and B (but not C) with the ability to activate a representation of food and during aversive conditioning this representation becomes associated with that of shock. At test, this will result in B, but not C, being able to activate a representation of shock via B→food→shock associative chain.

Although the analysis outlined in the previous paragraph seems reasonable, there is little direct evidence to support it. In particular, there is only correlational evidence linking mediated conditioning to the acquired equivalence and distinctiveness of cues. Thus, Ward-Robinson and Hall (1999) replicated the effect demonstrated by Honey and Hall (1989) and showed that after A was paired with shock, there appeared to be evidence of mediated conditioning in the form of the reluctance of the rats to perform an instrumental response for food. However, acquired equivalence/distinctiveness and this reluctance might not be causally related to one another. In an unpublished experiment, presented in M. Allman's (2004) doctoral thesis, that used the same design as Honey and Hall (1989), multiple presentations of food after aversive conditioning but before testing did not influence the magnitude of the acquired equivalence/distinctiveness effect. That is, extinction of the supposed mediating association between the representation of food and shock did not undermine the acquired equivalence/distinctiveness effect.

While there is little direct evidence to support the mediated conditioning account, there is direct evidence that it cannot provide a complete account for all instances of functional categorisation. For example, Honey and Watt (1998, 1999) first gave their rats a complex, configural, discrimination in which some similar compounds were followed by the same outcome (e.g., AX→food, BX→food) and other similar compounds were followed by a different outcome (e.g., CX→no food, DX→no food; see Table 2)

Table 2: Design of the study conducted by Honey and Watt (1998).

Training	**Revalue**	**Test**
AX+, AY-	A → Shock	
BX+, BY-		B vs. D
CX-, CY+	C→ No Shock	
DX-, DY+		

The important thing to note in this procedure is that each of the target stimuli (contexts A, B, C and D) are equally often paired with X and Y and with food and no food. This means that there will be no simple basis upon which mediated conditioning can result in greater generalisation between A and B than between A and C. However, this is exactly what Honey and Watt (1998, 1999) observed. These results, and others subsequently reported by Honey and his colleagues (e.g., Allman, Ward-Robinson & Honey, 2004; Allman & Honey, 2005, 2006; Honey & Watt, 1998, 1999; Honey & Ward-Robinson, 2001, 2002), provide support for a connectionist analysis. This connectionist account is shown in Figure 5.

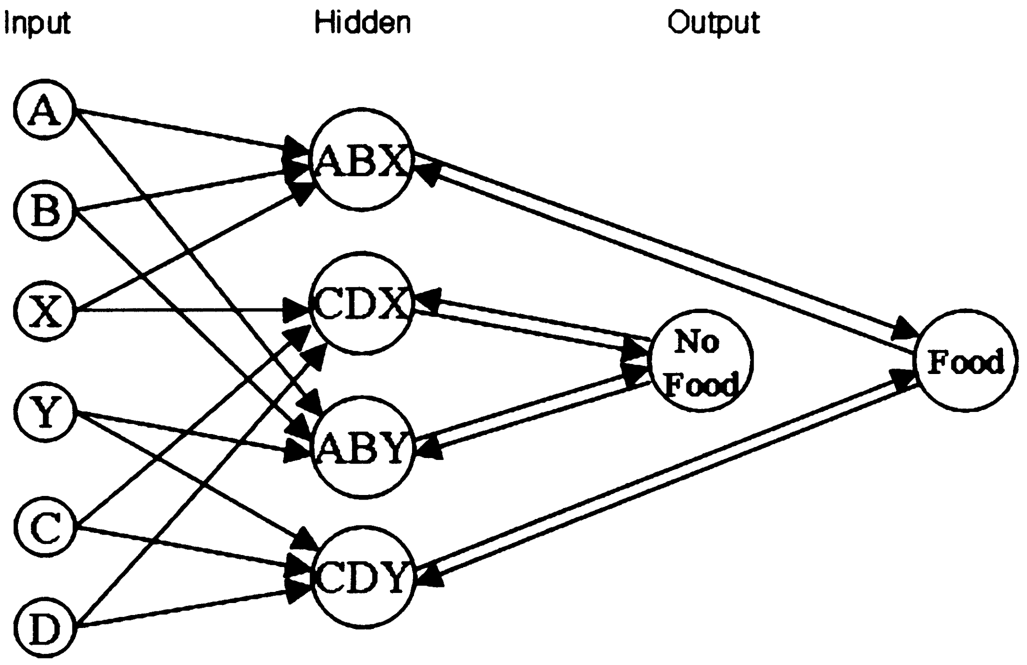


Figure 5. A connectionist account for the results observed by Honey and Watt (1998).

The connectionist structure depicted in Figure 5 is generated in the following manner. On an initial AX→food trial, both the A and X input units and the food output unit will be active. These units will activate (based on an initially random pattern of connections) a certain hidden node or unit ("AX") based on a winner-takes-all rule. Feed-forward links between the input units and this hidden unit will then be strengthened, as will reciprocal (feedback) links between the same hidden and the food output unit. On a subsequent BX→food trial, there are already strengthened connections between both X and "AX" and between this unit and food. This will increase the probability of "AX" being selected as the 'winner' and allow the B's input unit to become linked into the "AX" (now "ABX") unit.

This integrative process will be less likely to operate on a CX→no food trial, because activation of the no-food unit (unlike activation of the food unit on a BX→food trial) will not bias the selection of the BX unit through the reciprocal output-to-hidden layer links. Similarly, on a CY→food trial, Y will not bias the system to select the BX hidden unit. In summary, on trials where there is both a common input unit active and a common output unit active (e.g., when AX and BX are presented), the confluence of these two influences increases the likelihood that the same hidden unit will be selected.

The connectionist structure depicted in Figure 5 enables generalisation to occur between A and B in the following manner: during A→shock trials, the ABX and ABY nodes can become active and whichever wins will then become connected to shock. When B is then presented at test it will also activates both ABX and ABY and the animal will show fear to B. This will not be the case when D is presented because neither CDX nor CDY has gained an association with shock.

Connectionist networks of the type described above provide one way of integrating perceptual and functional categorisation: initial perceptual similarity is coded by the pattern of activation across the set of input units and is modified by experience through the patterns of weighted links involving the hidden units. Studies of the type described by Honey and Watt (1998) indicate that when stimuli are both similar and followed by the same outcome they become linked to the same hidden-layer node. This is an interesting suggestion and is one that I intend to explore in my second set of experiments, presented in Chapters 3 and 4, that examine functional categorisation in patterning problems. These are interesting problems in their own right (see below), but in the present context provide a useful test-bed within which to assess the accounts of acquired equivalence and distinctiveness outlined in the previous sections.

1. 3. 3. Patterning and categorisation

An interesting problem that rats and people alike are capable of solving is patterning. In a negative patterning task, when two elements (A or B) are presented separately, they are followed by one outcome (e.g., food) and when they are presented together (AB) they have a different outcome (e.g., no food; see Delamater, Sosa, & Katz, 1999; Woodbury, 1943; Forbes & Holland, 1980; Rescorla, 1974). In a positive patterning task, the pattern (AB) is paired with food while the individual elements are followed by no food (see Table 3).

Table 3: Design of negative and positive patterning.

	Elements		Compound
Negative patterning	A→Food	B→Food	AB→No Food
Positive patterning	A→No Food	B→No Food	AB→Food

The fact that organisms can learn such discriminations (particularly negative patterning) has been thought to pose a problem for some (elemental) models of learning (e.g., Minsky & Papert, 1969). However, there are a variety of ways in which such accounts can be embellished to allow them to solve a negative patterning discrimination. For example, if two stimuli have unique elements (A and B) and common elements (X), then X will acquire excitatory associative strength during a negative patterning (AX+, BX+, ABX-) and the unique elements will acquire inhibitory strength. The inhibitory associative strength will combine on compound trials and thereby result in less conditioned responding (see McLaren & Mackintosh, 2000, 2002). Alternatively, presenting A and B together might generate a unique configural cue or element that can enter into association with the outcome of a trial on which it is generated (see Rescorla, 1974; see also, Wagner, 2003). According to this account, the unique cue will acquire inhibitory associative strength during a negative patterning discrimination (A+, B+, ABX-). There is no reason for an elemental analysis to suppose that patterning training will not result in functional categorisation as measured by acquired equivalence. According to such an elemental analysis, following negative patterning (A+, B+, AB-), A will activate a representation of food when it is paired with shock and when B is presented at test it too will be able to activate food, and thereby shock.

The elemental form of analysis, however, might be considered to be rather contrived. Configural models provide a more intuitively plausible (and internally consistent) explanation of the acquisition of patterning discriminations: A, B and AB will each come to activate different configural units that become, in turn, linked to the outcome with which they are paired (see left-hand panel of Figure 6; e.g., Pearce, 1994); some support for this form of analysis comes from studies conducted by Williams, Metha, Poworoznyk, Orihel, George and Pearce (2002; see also, Redhead & Pearce, 1995; Williams, Dumont & Metha, 2004).

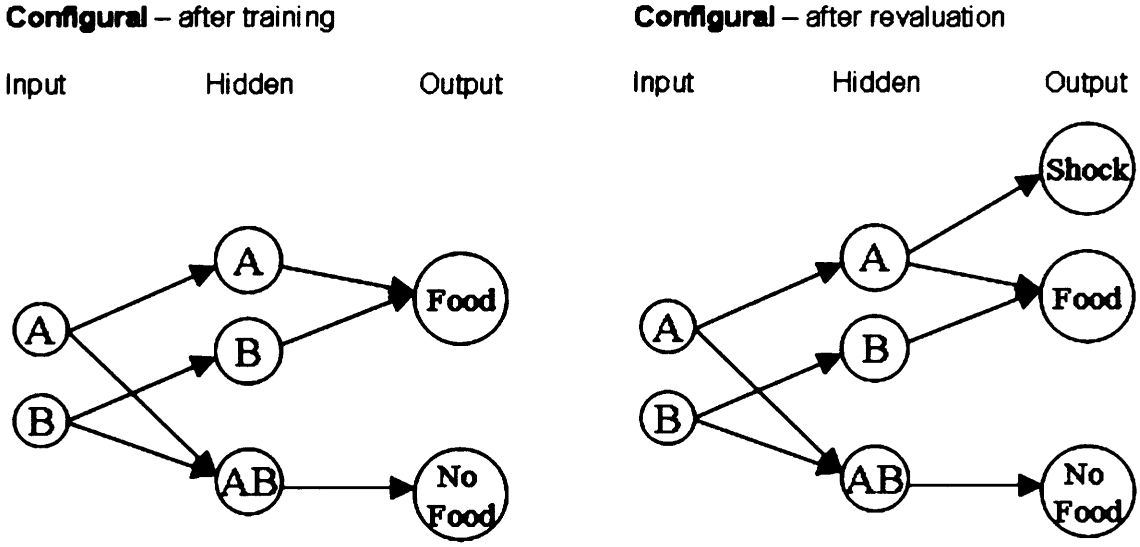


Figure 6. A configural analysis of patterning (A+, B+, AB-; left-hand panel) and the same model following a revaluation procedure in which A is paired with shock after initial negative patterning training (right-hand panel).

In terms of functional categorisation, patterning poses something of a problem for configural analyses. According to the configural analysis, patterning training should result in A and B being linked to separate configural representations (e.g., hidden-layer unit or nodes; see right-hand panel of Figure 6) and each of these units should become linked to food. Following revaluation training in which A is paired with shock, the configural unit activated by A will become connected to shock. This analysis would predict that generalisation might well occur to the presentation of AB (by virtue of similarity), but it should be unlikely to occur to B. In short, a configural analysis (one that precludes direct links forming between units in a given layer or even between input and output layer units) does not predict that a functional category to be formed as the result of separate pairings of A and B with food. Thus, the immediate question that this line of reasoning poses is: will the phenomena that are indices of functional categorisation be observed following patterning training?

1.3.4. Rationale for further research

The immediately-preceding analysis provides a clear rationale for further experimental work investigating whether or not patterning training results in functional categorisation. More specifically, I am interested in whether or not acquired equivalence/distinctiveness effects are observed after patterning training. If they are not observed, this would provide support for configural/connectionist analyses of functional categorisation (e.g., Honey & Ward-Robinson, 2002) and of associative learning more generally. If they are observed, then an elemental account of acquired equivalence (based on mediated conditioning) will gain a measure of support. The experiments presented in Chapters 3 and 4 examine whether animals, in this case rats that have received patterning discriminations, show acquired equivalence/distinctiveness effects. If these experiments do not reveal evidence of functional categorisation, then I will examine in detail why they do not, and if they do, then I will examine the nature of this functional categorisation. Before I present the results of Chapters 3 and 4, however, I will describe the work that I have conducted on perceptual categorisation.

CHAPTER 2: PERCEPTUAL CATEGORISATION

2.1. Introduction

The principal concern for any analysis of learning is to characterise the conditions under which it occurs. Empirical analysis of the conditions under which an association forms between the memories of two events, in both human and nonhuman animals, provides support for the three principles or laws of association identified with the British associationists of the nineteenth century (for reviews, see Hall, 1994; Warren, 1921): the principles of frequency, intensity and contiguity. A general restatement of these principles is that in order for the memories of two stimuli to become strongly associated they should, in the past, have been repeatedly and fully activated in close temporal proximity. It is widely accepted that any theoretical treatment of associative learning needs to capture these three principles. However, there is also evidence from nonhuman animals that association formation is influenced by stimulus similarity (see, for example, Rescorla & Furrow, 1977) and it is the theoretical analysis of this influence that is the primary focus of this chapter. In particular, this chapter is concerned with the nature of the associative structures (elemental or configural) that might underlie the finding that similarity promotes associative learning and indeed whether or not similarity should be accorded an independent principle of association.

To briefly recap on previous findings, in Rescorla and Furrow's (1977) seminal paper on similarity and association, pigeons initially received first-order conditioning to establish responding to two different stimuli (green and vertical) by pairing them with food. The pigeons then received second-order conditioning in which the trials involved stimuli that were either similar (e.g., blue→green and horizontal→vertical) or dissimilar (e.g., blue→vertical and horizontal→green). There was greater second-order conditioned responding to red and horizontal in the similar than in the dissimilar condition. Rescorla and Gillan's (1980) replication of this finding is more relevant to the experiments presented in this chapter where I manipulated the similarity of the stimuli rather than relying on their presumed similarity. In one of Rescorla and Gillan's studies, pigeons received pairings of similar compound stimuli (e.g., AX→BX and AY→BY) or dissimilar compound stimuli (e.g., AX→BY and AY→BX). Second-order conditioning again proceeded more readily in the similar than in the dissimilar condition.

Taken at face value, the evidence described in the previous paragraph seems to pose a problem for elemental theories of associative learning (e.g., Rescorla & Wagner, 1972) within which similarity *per se* has no independent influence on the process of association – independent from the influences of frequency, intensity and contiguity. However, Rescorla and Gillan (1980) presented an analysis of their findings based on a relatively simple modification to such an elemental associative analysis that was later given formal expression within Wagner's (1981) elemental model of associative learning – a modification that is now considered to be relatively uncontroversial. Rescorla and Gillan (1980) argued that when two similar stimulus compounds are paired (i.e., AX→BX), the fact that X is present during the first compound might render it poorly processed and B particularly well processed during the second compound. The operation of this (short-term habituation) process should allow A to be selectively associated with the well-processed B feature. Given the observation that Rescorla and Gillan's first-order conditioning phase would have left B with greater associative strength than X, then it is to be expected that a stimulus that has an effective association with B will elicit a great deal of second-order conditioned responding. The conclusion that Rescorla and Gillan reached was that the influence of similarity upon association formation can be reduced to the operation of the principle of intensity – albeit one in which intensity is not a simple function of the physical intensity of a stimulus. This analysis was couched in elemental terms.

In Chapter 1 of this thesis I presented an alternative configural analysis of Rescorla and his colleagues' findings. Briefly, according to this analysis, the fact that there is a common element on similar trials results in the three elements that are presented (e.g., A, X and B) coming to share the ability to activate a hidden unit. Whatever the associative structures (elemental or configural) that underlie Rescorla and his colleagues' findings, it is possible that their results might have limited generality and the theoretical analyses offered for them might not apply to other experimental preparations and species. The central aim of Experiments 1-4, therefore, was to examine the generality of Rescorla and Gillan's (1980) results and to assess their theoretical treatment of the role of similarity in associative learning. To this end, in Experiments 1-3 I examined the role of similarity in associative learning in humans using a novel experimental procedure that was superficially different, but formally similar to that used by Rescorla and Gillan (1980). In Experiment 4, I assessed the possibility that similarity has an influence on learning that is not captured by the elemental analysis outlined above, but is rather supported by configural processes.

2.2. Experiments 1 and 2

2.2.1. Introduction

The experimental design used in Experiments 1 and 2 is summarised in Table 4. The participants sat before a computer screen on which different sequences of greyscale patterns were presented. The six sequences (AX→BX, AY→BY, AZ→BZ, CX→DY, CY→DZ, and CZ→DX) each consisted of two patterns constructed from grey internal features (A, B, C and D) and black external features (X, Y and Z). The sequences involving A and B were similar to one another by virtue of the fact that both were presented with the same external features (e.g., AX→BX) and those involving C and D were dissimilar by virtue of the fact that C and

D were presented with different external features (e.g., CX→DY). Examples of the patterns used are shown in Figure 7. The participants were informed that these patterns represented alien bugs and that they were required to learn which of the two insecticide sprays (red or blue), depicted on the left- and right-hand sides of the screen, eliminated which bugs. Patterns with feature B could be eliminated using one of the sprays (e.g., red) and those with feature D could be eliminated with the other spray (e.g., blue). Patterns with features A and C could be eliminated using either spray. The issues of principal interest were (1) whether the fact that the first and second patterns were temporally contiguous would promote the transfer of performance (i.e., spray use) between them and (2) whether this transfer would be more marked between the similar sequences (e.g., AX→BX) than the dissimilar sequences (e.g., CX→DY). That is, when participants were presented with bugs with features A and C would they use the spray that eliminated bugs containing feature B and D, respectively; and if so, then would this effect be more evident between the similar patterns (e.g., AX and BX) than between the dissimilar patterns (e.g., CX and DY)?

Table 4: Design of Experiments 1 and 2.

Condition	**Sequence Type**		
Similar	AX→BX→red	AY→BY→red	AZ→BZ→red
Dissimilar	CX→DY→blue	CY→DZ→blue	CZ→DX→blue

Note: Participants received six sequences that each consisted of two patterns constructed from internal features (A, B, C and D) and external features (X, Y and Z). The patterns involving A and B were similar to one another (e.g., AX→BX) and those involving C and D were dissimilar (e.g., CX→DY). The patterns were portrayed as alien bugs and the participants needed to learn to use one insecticide spray (red) to eliminate bugs with feature B and another spray (blue) to eliminate bugs with feature D. The participants could use either spray to eliminate bugs with features A and C.

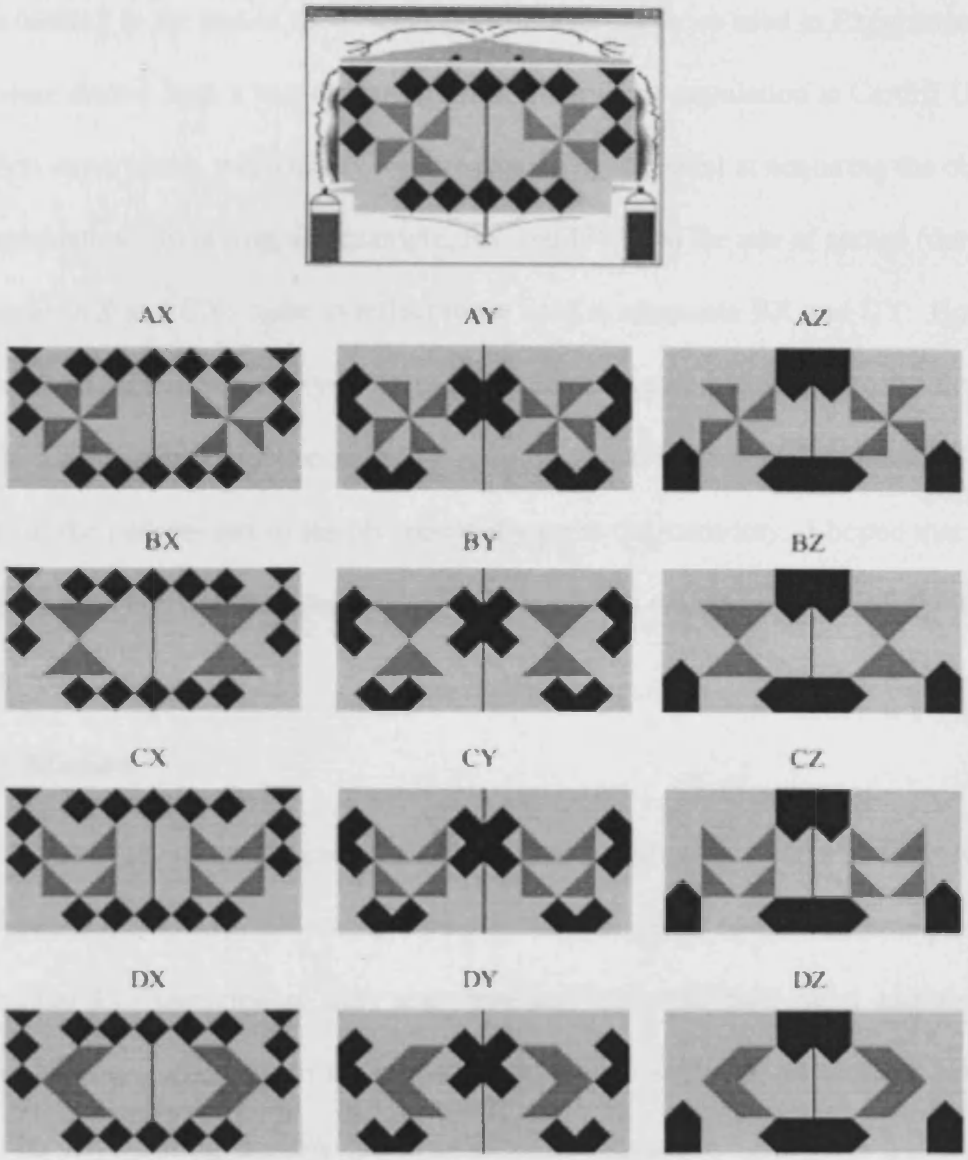


Figure 7. Patterns used in Experiments 1–4. Each pattern was constructed from two squares and each square contained a pattern that was a mirror image of the other. The grey internal features of both squares served as A, B, C and D and the black external features served as X, Y and Z (or S, T and U in Experiment 4). The letter combinations above each pattern (i.e., AX–DZ) illustrate one counterbalanced sub-condition from Experiments 1 and 2.

Experiments 1 and 2 were run in an identical fashion with the exception that the participants used in Experiment 1 were unpaid students in the School of Psychology who were coming to the end of the academic year, whereas those used in Experiment 2 were paid and were drawn from a broader spectrum of the student population at Cardiff University. In the first experiment, participants were reasonably successful at acquiring the obligatory discrimination (involving, for example, BX and DY) and the use of sprays (during, for example, AX and CX) came to reflect those used to eliminate BX and DY. However, in Experiment 1 there was no systematic difference between responding to the first patterns of similar and dissimilar sequences (see Appendix 1) and there was a pronounced tendency for many of the participants to simply repeatedly press the same key. I hoped that the change in remuneration and the characteristics of the population would avoid the aforementioned problems.

2.2.2. Method

Participants and apparatus. Thirty-two participants were solicited using the School of Psychology, Cardiff University, paid panel and given £5.00 for their participation. An additional four participants were also recruited using the paid panel and also given £5.00. (These four replaced four of the original sample who exhibited an extreme bias to use one of the two keys during the first patterns of each sequence.) All participants had normal or corrected-to-normal vision. A CRT monitor in an evenly lit, quiet room was used to display the stimuli, using custom-written software written in Microsoft Visual C#. Participants used the 'T' key on a computer keyboard to activate the red spray (on the left-hand side of the screen) and the 'V' key to activate the blue spray (on the right-hand side of the screen).

Stimuli and counterbalancing. The stimuli were created using custom-written software written in Microsoft Visual C#. Each pattern was constructed from two adjacent squares, separated by a black line that contained patterns that were reflections of one another along the vertical axis. The total extent of the pattern was 800×400 pixels (W×H; approximately 25×12.5cm). Each pattern was also surrounded by elements intended to reinforce the cover story that the patterns represented alien bugs (see Figure 7). The internal features of each square were grey and occupied a central square (200×200 pixels or approximately 6.25×6.25cm) and these features served as A, B, C and D. The external features were black and occupied the remaining area surrounding the central square. For half of the participants, one pair of internal features (depicted in the first and second rows of three patterns in conjunction with the outer features that served as X, Y and Z) served as A and B and the remaining pair (depicted in the third and fourth rows of three patterns) served as C and D; for the other half, this arrangement was reversed. Each of the internal features (A, B, C and D) was presented with each of the external features (X, Y and Z). The resulting patterns were presented on a screen that was otherwise illuminated with white light.

Procedure. After signing the informed consent form, participants were seated approximately 80cm in front of the computer screen and read the following instructions presented on a sheet of paper before beginning the experiment:

For participants:

Evil 'Space Bugs' have invaded Britain! To make matters worse, these bugs have mutated into a number of differing types. However, British scientists have developed two different insecticide sprays which have been found to be effective in destroying different bug types. Using these two different sprays, your task is to destroy these 'Space Bugs' as quickly as possible before they have a chance to spread throughout the world.

As said above, the two spray cans, which you will see on the computer screen, contain different insecticide sprays which are effective in destroying different bug types. A Bug's *type is denoted by the pattern on its back. These spray cans can only be fired once they have fully charged, and if allowed to discharge completely, can only be sprayed again once the spray cans have fully charged again. The charge of the spray cans can be identified from the rising and falling levels inside the spray cans. Feedback as to whether you have or have not picked the correct spray to destroy a particular 'Space Bug' will only be given for every other encounter with a 'Space Bug'. This feedback will be presented in the top-centre of the computer screen, and you should correct your original decision accordingly.*

The fate of the world rests in your hands, and speed is of the essence. In this task speed is heavily linked to accuracy, so please try to be as accurate and fast as possible.

Competition:

A luxury bottle of wine will be given to the participant with the highest score at the end of testing. As such, your details will be recorded confidentially, until the prize winner has been identified, after which your details will be anonymised. In the event of a tie, these participants' names will be put into a hat and a winner drawn. If you do not wish to take part in the competition and have your details be anonymous from the point of testing, please inform the experimenter.

Keys:

Key T/F fires the left-hand spray.

Key V/H fires the right-hand spray.

The first sequence was initiated by the participant pressing the return key (for flowchart of events see appendix 2). Each sequence started with the presentation of a pattern with feature A (accompanied by X, Y or Z) or feature C (accompanied by X, Y or Z). Once the pattern was presented, an animated 2 s screen display began, in which the spray cans were filled with red and blue spray and then emptied over the course of 5 s. During these first patterns, pressing either key at any point resulted in the pattern disappearing and being replaced by a blank white screen. There then followed an inter-stimulus interval of 1 s and the presentation of the second pattern in the designated sequence. If the first pattern contained feature A and an internal feature X, Y or Z, then the second pattern would contain feature B and the same internal feature (i.e. sequences AX→BX, AY→BY, AZ→BZ); if the first pattern contained feature C and internal feature X, Y or Z, then the second pattern contained feature D and the internal feature Y, Z and X, (i.e., sequences CX→DY, CY→DZ, CZ→DX). During the second patterns, there was again a 2-s animated display in which both cans filled. During this period, the participants' responses had no consequences and the message "Wait, preparing sprays!" appeared in black text within a white banner immediately above the pattern; this message was replaced by "Ready to zap". Over the next 5 s the contents of the cans emptied: this period constituted the response window. If participants did not respond during this window, a message in black text was displayed: "Answer faster!", the contents of the cans were replenished over a 5-s period and the participants could not respond. There followed another 5 s response window identical to that just described. This process continued until participants responded. For half of the participants, the second patterns in similar sequences (i.e., BX, BY and BZ) could be eliminated by the red spray and the second patterns in dissimilar sequences could be eliminated by the blue spray; for the remaining participants, this assignment of condition (similar or dissimilar) to spray (red or blue) was reversed. If participants used the correct spray, the pattern remained on the screen for a further 2 s together with the message "You pressed T/V and are CORRECT" in green text on a white banner centrally located and immediately above the pattern. The screen then became blank and there followed an inter-trial interval (ITI) of a random value between 1 and 6 s. If participants used the incorrect spray, both cans emptied and the pattern remained on the screen with the message “You pressed T/V and were WRONG”, in red. The correct can was replenished over the course of 5 s. Once the can was replenished, a further 5 s response window started. This process continued until the correct spray was selected, at which point the same procedure was started as if the correct key had been selected when subjects were first given the choice. If participants failed to respond during the response window then the message “Respond faster!” appeared and the contents of the cans were replenished over a 5 s period, during which the participants could not respond. During the session there were 20 presentations of each of the six sequences (i.e., 60 similar sequences and 60 dissimilar sequences). These sequences were arranged in a pseudo-random order that differed across participants, with the constraint there were no more than two sequences of the same type (similar or dissimilar) in succession.

Data analysis. For the first and second pattern of each type of sequence (similar and dissimilar) I recorded the percentage of sequences in which the correct spray was chosen as the first choice. For the second patterns, ‘correct’ simply referred to using the spray that was designated to eliminate the pattern. For the first patterns, ‘correct’ indicated that participants used the same spray that eliminated the second pattern of the same sequence. Thus, an increase in correct scores for the second patterns indicated that participants were learning (using trial and error) which spray (red or blue) to use to eliminate patterns with features B and D and correct scores for the first patterns indicated transfer of what was learnt about the second patterns to the first patterns by dint of their temporal contiguity. I also measured reaction times for the first and second patterns (measuring reaction times from pattern onset for the first patterns and 2 s from pattern onset for the second patterns; see above). These scores were pooled across the counter-balanced factors of assignment of stimulus pairs to condition (similar or dissimilar) and assignment of spray to condition. The percentage correct scores and reaction times were presented in 10-sequence blocks. As I have already mentioned, four of the participants showed an extreme spray bias, almost exclusively using one of the keys (i.e. sprays) during the first but not the second patterns. These participants were replaced with four new participants (run through the same counterbalanced sub-conditions as those they replaced) who did not show such a bias.

2.2.3. Results

Figure 8 shows the mean percentage correct scores for the first patterns (involving A and C) and second patterns (involving B and D). Inspection of the left-hand panel reveals that, over the course of training, there was an increase in the percentage correct for A but not for C and during the final three blocks the percentage correct scores for A were higher than those for C. ANOVA revealed no effect of block (1-6), $F(5,155)=2.02, p>.07$, an effect of similarity (A versus C), $F(1,31)=4.58, p<.05$ and an interaction between these factors, $F(5,155)=2.42, p<.05$. Analysis of simple main effects revealed an effect of block for stimulus A, $F(5,155)=4.54, p=.001$ but not C, $F<1$, and effects of similarity on blocks 4 and 6, minimum $F(1,31)=5.31, p<.05$. Moving to the scores in the right-hand panel, it is clear that there was a dramatic increase in the scores over the course of training and this increase was similar for patterns involving both B and D. ANOVA confirmed that there was an effect of block, $F(5,155)=14.61, p<.001$, no effect of similarity, $F<1$ and no interaction between these factors, $F(5,155)=1.87, p>.10$.

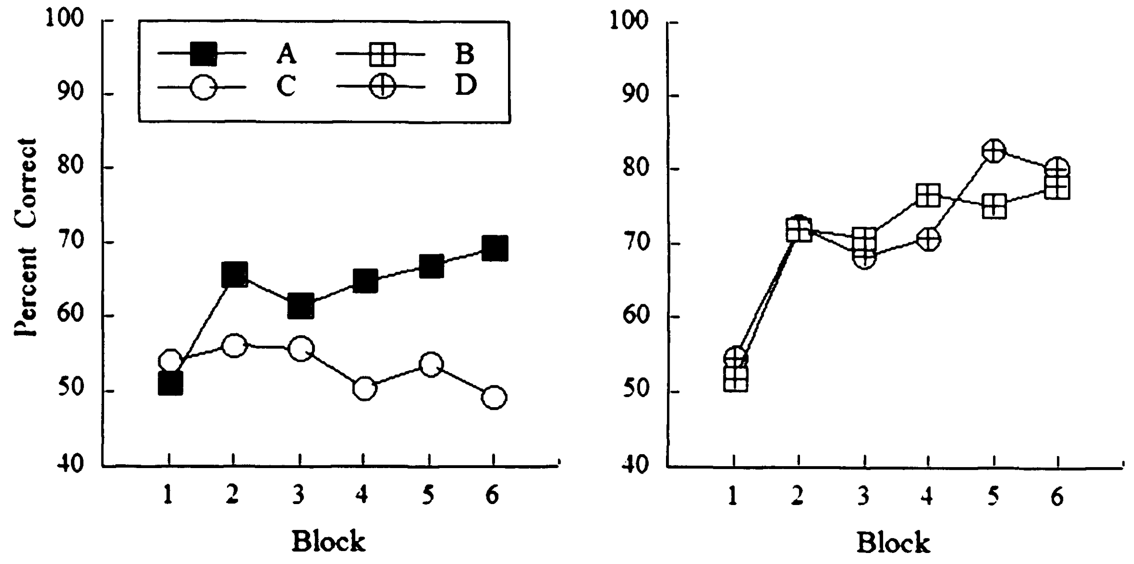


Figure 8. Experiment 2. Mean percentage correct for the first patterns (A and C; left panel) and second patterns (B and D; right panel).

The reaction times for each of the four stimuli across training are shown in the top panel of Table 5. Inspection of this panel reveals that reaction times were initially similar for the first and second patterns and that over the course of training reaction times became shorter; markedly more so for the first patterns than the second. This difference probably reflects the fact that reaction times for the first patterns were measured from pattern onset, including the period in which the spray cans were filling but during which time participants could respond, whereas the reaction times for the second patterns included the 2 s period in which participants' responses had no consequences and the message "Wait, preparing sprays!" was displayed. Inspection of Table 5 also indicates that there were no differences between the pairs of stimuli serving as first or second patterns. ANOVA conducted on the reaction times for the first patterns revealed an effect of block, $F(5,155)=19.64, p<.001$, no effect of similarity and no interaction between these factors, $Fs<1$. Similarly, ANOVA conducted on the reaction times for the second patterns revealed an effect of block, $F(5,155)=16.84, p<.001$, no effect of stimulus and no interaction between these factors, $Fs<1$.

Table 5: Mean reaction times in Experiment 2.

Block	1	2	3	4	5	6
A	3066.41	2007.50	1842.13	1747.39	1743.64	1725.46
C	3169.61	2098.22	1916.59	1722.54	2038.86	1643.78
B	3376.12	2798.63	2735.36	2627.88	2614.58	2600.36
D	3253.22	2836.33	2773.70	2618.77	2624.73	2599.35

Note: Mean reaction times to the first patterns (A and C) and second patterns (B and D) in Experiment 2. Responses during the first 2 s of each pattern were ineffective, but the reaction times included these periods.

2.2.4. Discussion

The results of Experiment 2 allow me to confirm the fact that similarity promotes association in humans using a novel within-participants procedure employing novel stimuli whose similarity can be readily manipulated (cf. Rescorla & Gillan, 1980; see also Holland and Ross, 1981, for analogous pattern of results and procedure). In this particular case, participants were exposed to both similar (e.g., AX→BX) and dissimilar (e.g., CX→DY) sequences and learnt to make one response (red spray) in the presence of BX and another response (blue spray) in the presence of DY. In spite of the fact that the participants could use either spay during AX and CX, their spray use during AX (i.e., red spray) came to match that used during BX; an effect that was not evident for CX. Given the fact that the responses to BX and DY were themselves similarly high, then the observed differences in responding to AX and CX must reflect the fact that on similar trials there is a common element whereas on dissimilar trials there is not.

There are, of course, a variety of ways in which the similarity effect observed in Experiment 2 might have been generated, some of which were foreshadowed in Chapter 1. One possibility that was not mentioned in the Introduction, but that is readily applicable to the results of Experiment 2, is that any direct associations between the first patterns and the outcomes might have been augmented by the presence of similar, but not by dissimilar, second patterns. Indeed, there is historical precedence for this type of analysis, from studies of serial conditioning in pigeons (e.g., Rescorla, 1982; Thomas, Robertson, & Lieberman, 1990). However, this form of analysis predicts that similarity should always facilitate learning in the type of procedure used in Experiment 2 and evidence that is inconsistent with this prediction will be presented in Experiment 4.

2.3. Experiment 3

2.3.1. Introduction

Experiment 3 was identical in design to Experiment 2 with three amendments: first, the participants were forced to wait 2 s before being able to respond to the first patterns, to prevent participants developing the strategy of responding very rapidly to patterns preceded by a long interval. Second, the length of the training was doubled from 120 to 240 trials, to enable me to assess whether the effect of similarity observed during the latter blocks of Experiment 2 was maintained. Third, the number of participants was doubled from 32 to 64, to reduce the impact of individual variability on the results without excluding participants. These changes should help to confirm the reliability of the effects observed in Experiment 2.

2.3.2. Method

Participants and apparatus. Sixty-four participants were solicited using the School of Psychology, Cardiff University, paid panel and given £5.00 for their participation. All participants had normal or corrected-to-normal vision. The stimuli and apparatus were the same as for Experiment 2. The procedure was identical to Experiment 2 with the following exceptions: the number of sequences of each of the six types was doubled from 20 to 40; thus there were now 120 similar sequences and 120 dissimilar sequences, the procedure that was used to stop participants responding during the first 2 s of the second patterns was also applied to the first patterns and while half of the participants used the 'T' key to activate the red spray and the 'V' key to activate the blue spray, as in Experiment 2, the remainder used the 'F' key to activate the red spray and the 'H' key to activate the blue spray. The instructions were the same, except the prize for the participant that performed most accurately was £10.00 and the participants who were to use keys F and H were instructed to do so. Statistical analysis showed that the manipulation of key assignment made no significant difference to the pattern of the results that was observed and so it will henceforth be ignored. One participant's mean reaction time on block 10 was extremely long (consistent with an interruption) and this participant's score on this block was replaced with the mean of the remaining 63 participants; aside from this (and as in Experiment 2), no other smoothing operations were used on the reaction time scores.

2.3.3. Results

The mean percentage correct scores for the first patterns (involving A and C) and second patterns (involving B and D) are shown in Figure 9. Inspection of the left-hand panel reveals that over the course of training there was an increase in the percentage correct for A but not for C. ANOVA revealed an effect of block (1-12), $F(11,693)=5.19, p<.001$, no effect of similarity (A or C), $F(1,63)=2.84, p>.09$ and an interaction between these factors,

$F(11,693)=3.25$, $p<.001$. Analysis of simple main effects revealed an effect of block for stimulus A, $F(11,693)=8.27$, $p<.001$ but not C, $F<1$ and effects of stimulus on blocks 8-11, minimum $F(1,63)=4.05$, $p<.05$. Moving to the scores shown in the right-hand panel, it is clear that there was a dramatic increase in the scores over the course of training and that this increase was similar for patterns involving both B and D. ANOVA confirmed that there was an effect of block, $F(11,693)=53.38$, $p<.001$, no effect of similarity and no interaction between these factors, *F*s<1.

Finally, the reaction times for each of the four stimuli across training are depicted in the middle panel of Table 6. Inspection of this panel reveals that the reaction times, now measured in the same way for the first and second patterns (i.e. excluding the 2-s period in which participants could not respond), were similar for A–D, and became shorter as training proceeded. ANOVA conducted on the reaction times for the first patterns revealed an effect of block, $F(11,693)=14.92$, $p<.001$, no effect of similarity, $F<1$ and no interaction between these factors, $F(11,693)=1.09$, $p>.36$. Similarly, ANOVA conducted on the reaction times for the second patterns revealed an effect of block, $F(11,693)=29.67$, $p<.001$, no effect of similarity, $F(1,63)=1.13$, $p>.29$ and no interaction between these factors, $F(11,693)=1.24$, $p>.25$.

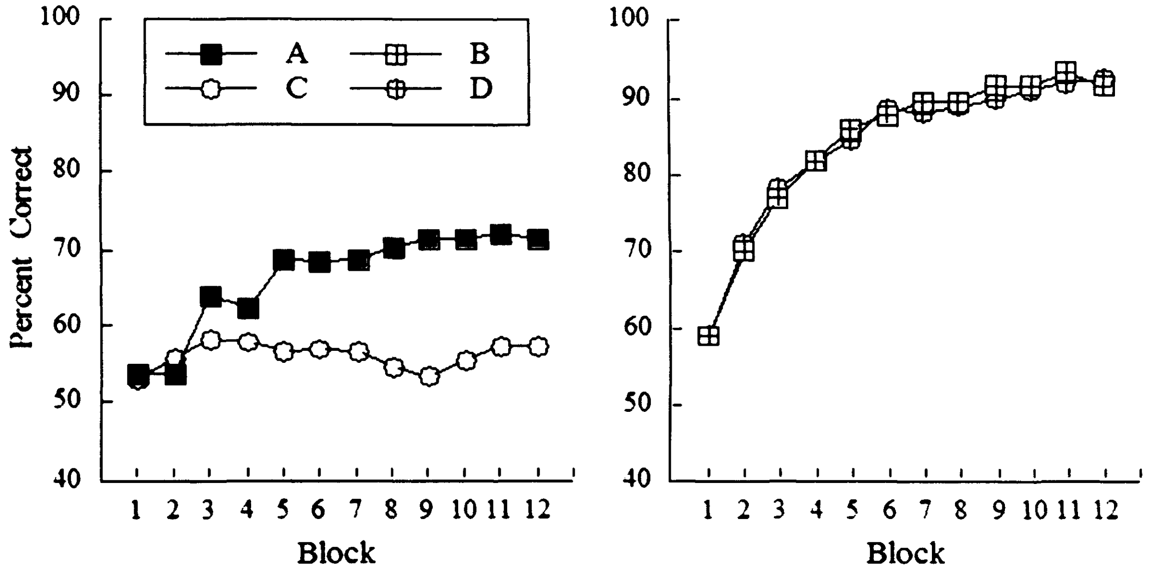


Figure 9. Experiment 3. Mean percentage correct for the first patterns (A and C; left panel) and second patterns (B and D; right panel).

Table 6: Mean reaction times for Experiment 4.

Block	**1**	**2**	**3**	**4**	**5**	**6**
A	1580.22	927.86	734.08	663.24	571.61	532.61
C	1655.14	1278.94	733.12	560.12	496.07	541.35
B	1368.56	817.51	742.30	615.19	561.05	536.87
D	1211.77	829.06	707.73	615.24	555.54	561.54
Block	**7**	**8**	**9**	**10**	**11**	**12**
A	444.81	497.21	402.95	539.37	535.22	539.38
C	515.52	402.09	551.39	497.64	378.75	450.84
B	508.52	504.03	513.57	505.24	501.06	511.60
D	514.41	517.28	530.55	497.65	479.14	489.58

Note: Mean reaction times to the first patterns (A and C) and second patterns (B and D) in Experiment 4. In both experiments, responses during the first 2 s of each pattern were ineffective and the reaction times excluded these periods

2.3.4. Discussion

Experiment 3 successfully replicated the results from Experiment 2, again showing that when two similar patterns are paired (e.g., AX→BX) the response established to the second pattern (BX) is more likely to be elicited by the first pattern (AX) than when the two patterns are dissimilar (e.g., CY→DZ). In Chapter 1 (and indeed in the Introduction to Chapter 2) I identified two general ways of interpreting evidence showing that similarity promotes association: an elemental analysis and a configural analysis. It is now time to consider possible ways of distinguishing between them in the present experimental paradigm.

The elemental analysis developed by Rescorla and Gillan (1980) can be readily applied to the results of Experiments 2 and 3. During similar trials (AX→BX; see upper left-hand panel of Figure 10) the fact that X is present during AX will, due to a process of short--term habituation, render it poorly processed (dotted line) during the presentation of BX and will allow B to be particularly well processed. This change in the distribution of processing between B and X should allow A to be better associated with B than with X. Given the fact that in the procedures used in Experiments 2 and 3, it is the relevant B as opposed to the irrelevant X that is associated with the use of red spray, then associating A with B will result in substantial transfer of responding from BX to AX. During dissimilar compounds (CY→DZ; see lower left-hand panel of Figure 10) both components of the second pattern will attract equivalent levels of processing and C will be no more likely to be associated with D than Z. Given the fact that, in this example, it is the relevant D as opposed to the irrelevant Z that is associated with the use of blue spray, then the fact that D is not particular well processed should leave C association with D relatively weak; relative, that is, to the A-B association.

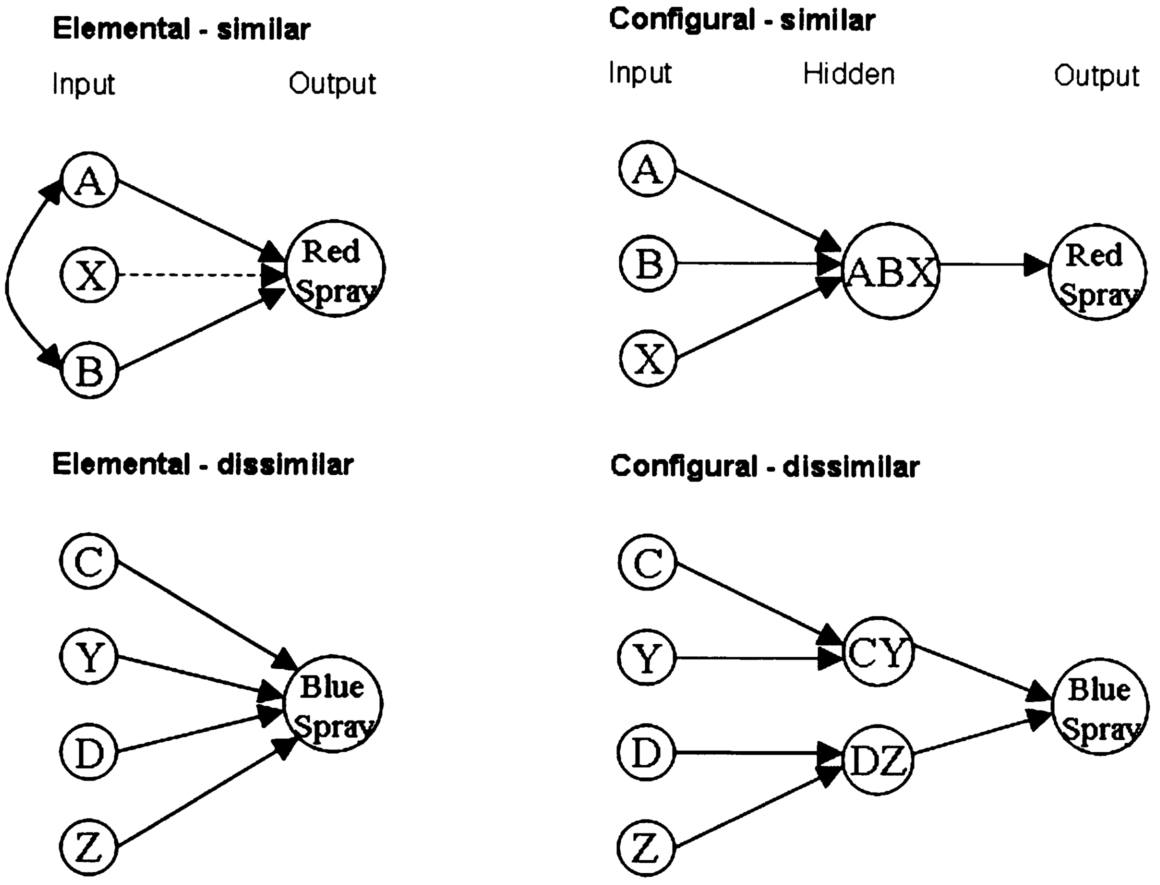


Figure 10. An analysis of elemental (left) and configural (right) models of learning in human similarity experiments.

A configural account for the results of Experiments 2 and 3 is slightly different to the configural analysis developed for Rescorla and Gillan's (1980) results in Chapter 1. On the one hand, during a similar trial (e.g., AX→BX→red spray trial) when AX is presented, its components will activate a random hidden-layer unit (AX), and the participant will use the red or blue spray to eliminate the bug. When BX is presented, the fact that X is present and that there is residual activity in the AX node, then B will become connected to this node (now ABX). The participant will generate another choice of spray use until the correct one is chosen. This will result in the pattern of links shown in the top right-hand panel of Figure 10 and allow the presentation of AX to initiate the same choice as BX. On the other hand, during a dissimilar (e.g., CY→DZ→blue spray), when CY is presented its components will activate a random hidden-layer unit (CY), and the participant will use the red or blue spray to eliminate the bug. When DZ is then presented, the sole basis for the components of DZ to become linked to CY is residual activity in CY. This residual activity might not be sufficiently marked to prevent DZ activating its own, different hidden unit. The participant will generate another choice of spray use on DZ until the correct one is chosen. This will result in the pattern of links shown in the lower right-hand panel of Figure 10 and will not allow the presentation of CY to initiate the same choice as DZ. The design used in Experiment 4 will allow me to determine which of these two types of account for the results of Experiments 2 and 3, elemental or configural, is to be preferred.

2.4. Experiment 4

2.4.1. Introduction

The experimental design and (additional) stimuli used in Experiment 4 are summarised in Table 7 and Figure 11, respectively. Participants received two types of similar sequence (e.g., AX→BX and CS→DS) and two types of dissimilar sequence (e.g., AS→BT and CX→DY). Which of four sprays eliminated bugs with features B and D depended upon the (external) features that accompanied them: when B was accompanied by either X, Y or Z, one spray (e.g., red) eliminated the pattern, whereas when B was accompanied by either S, T or U, a different spray (e.g., green) was effective. Similarly, when D was accompanied by X, Y or Z, the blue spray was effective, whereas when D was accompanied by S, T or U, a different spray was effective (e.g., purple). Under these conditions, participants could not simply process the internal features of the second patterns (i.e., B or D) to guide their choice of response (i.e., red, blue, green or purple). For example, if participants processed only B and D, it would be possible for them to learn that using either red or green is effective during

B (and using purple or blue is effective during D), but this would result in an overall level of accuracy during the second patterns of only fifty per cent. If the participants were able to learn to use the appropriate sprays during the various compounds involving features B and D (i.e., score above 50%), then the question of interest becomes: would what is learned about these compounds (e.g., BX→red and DY→blue) be more or less likely to transfer to the first patterns of similar trials (e.g., as a result of AX→BX→red trials) than to the first patterns of dissimilar trials (e.g., as a result of CX→DY→blue trials).

Table 7: Sequence types used in Experiment 4.

Condition	**Sequence Types**		
Similar	AX→BX→red	AY→BY→red	AZ→BZ→red
Similar	CS→DS→purple	CT→DT→purple	CU→DU→purple
Dissimilar	AS→BT→green	AT→BU→green	AU→BS→green
Dissimilar	CX→DY→blue	CY→DZ→blue	CZ→DX→blue

Note: Participants received 12 sequences that each consisted of two patterns constructed from internal features (A, B, C and D) and two sets of external features (set 1: X, Y and Z; set 2: S, T and U). The patterns involving A and B were similar to one another when accompanied by set 1 (e.g., AX→BX) and those involving C and D were similar when accompanied by set 2 (e.g., CS→DS); the patterns involving A and B were dissimilar to one another when accompanied by set 2 (e.g. AS→BT) and those involving C and D were dissimilar when accompanied by set 1 (e.g., CX→DY). Participants learnt to use different insecticide sprays (e.g. red and purple) to eliminate the second patterns in similar sequences (e.g., BX and DS, respectively) to those (e.g., blue and green) used to eliminate the second patterns in dissimilar sequences (e.g., DY and BT, respectively). Participants could use any spray to eliminate patterns with features A and C.

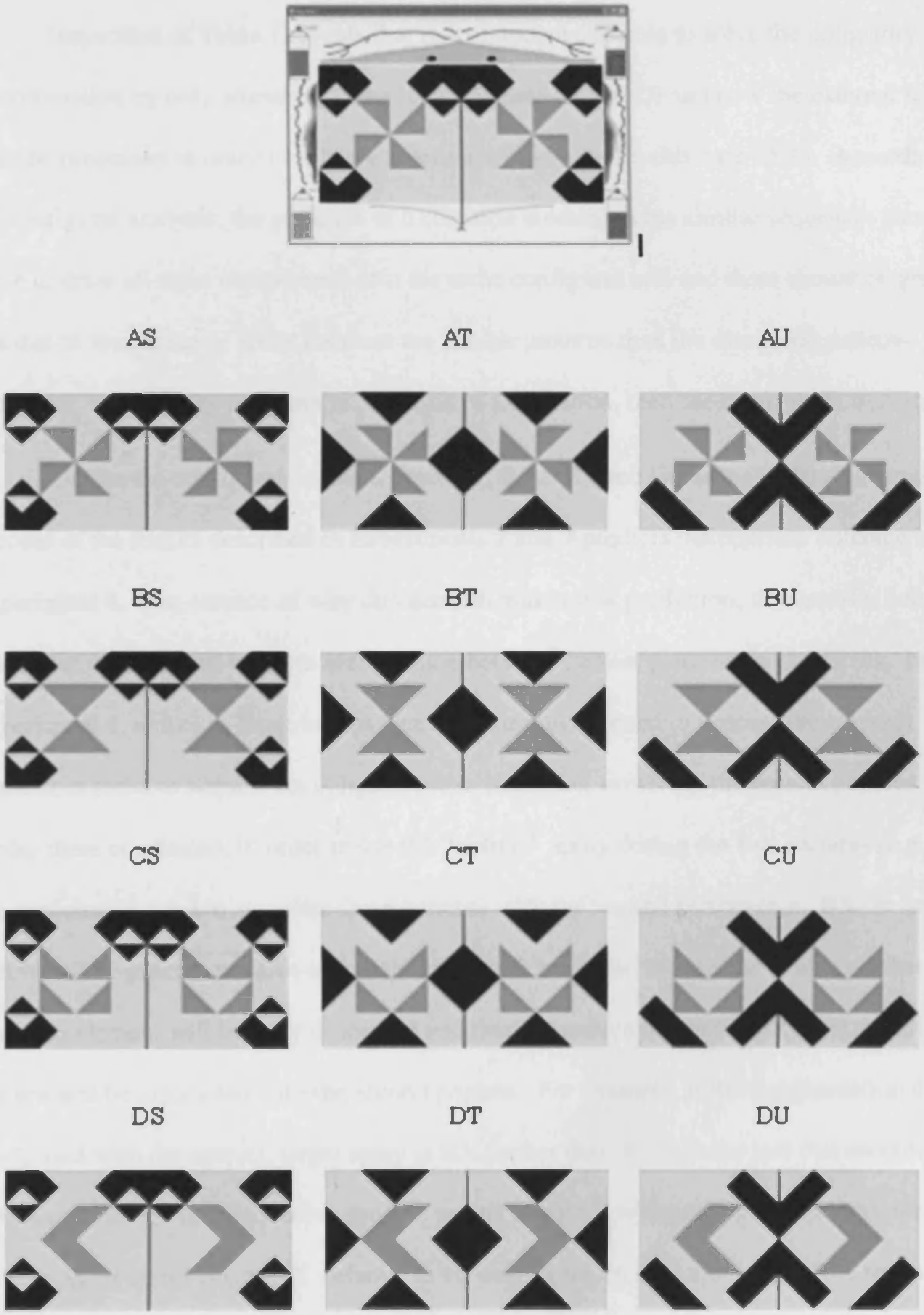


Figure 11. The additional patterns used in Experiment 4. Each pattern was constructed in the same way as those used in Experiments 1-3 and the grey internal features again served as A, B, C and D and the black external features as S, T and U (or X, Y and Z). The letter combinations above each pattern (i.e., AS–DU), together with those in Figure 1 (i.e., AX–DZ), illustrate one counterbalanced sub-condition from Experiment 4.

Inspection of Table 7 reveals that it is no longer possible to solve the obligatory discrimination by only attending to the inner features (B and D) and now the external features must be processed in order to achieve scores above chance (in this case 25%). According to the configural analysis, the presence of a common element in the similar sequences should serve to draw all three components into the same configural unit and there should be greater transfer of the choice of spray between the similar patterns than the dissimilar patterns. Moreover, if similarity is a general principle of association, then the same prediction holds.

Unlike the configural account, however, Rescorla and Gillan's (1980) elemental account of the results described in Experiments 2 and 3 predicts the opposite outcome in Experiment 4. The essence of why this account makes this prediction, that transfer between dissimilar patterns will be more marked than between similar patterns, is as follows. In Experiment 4, unlike in Experiments 2 and 3, participants need to process the external features in order to acquire the obligatory discrimination involving the second patterns. Under these conditions, in order to use the "correct" spray during the first patterns (e.g., AX) the participants need to associate these patterns with the second pattern (e.g., BX) in its entirety. The process of short-term habitation will, however, reduce the likelihood that the common element will be fully processed and thereby reduce the probability that the first pattern will be associated with the second *pattern*. For example, if the representation that is associated with the correct, target spray is BX (rather than B), then the fact that short-term habituation biases the associative process to link A with B will now reduce the transfer of what is learnt about BX to AX, relative to the case in which both elements of the second pattern are fully processed (i.e., on a CX→DY trial). That is, according to Rescorla and Gillan's (1980) form of analysis, similarity should hinder learning when X and Y are relevant to the task that participants are undertaking. These different predictions of the configural and elemental analyses were assessed in Experiment 4.

2.4.2. Method

Participants and apparatus. Sixty-four participants were solicited using the School of Psychology, Cardiff University, paid panel and given £5.00 for their participation. All participants had normal or corrected-to-normal vision. The apparatus was as that used in experiments 1 to 3, with the exception that participants used the 'T' key on the computer keyboard to activate the red spray, 'U' to activate the blue spray, 'B' to fire the green spray and 'M' to fire the purple spray. The red and green sprays were positioned in the top and bottom left-hand corners of the screen, respectively, and the blue and purple sprays were positioned in the top and bottom right-hand corners of the screen, respectively.

Stimuli, counterbalancing and procedure. The additional set of external patterns needed for experiment 4 are depicted in Figure 10 and were created in the same way as those used in experiments 1 to 3. In Experiment 4, the counterbalancing was achieved in the same way as experiments 1 to 3, with the exception that the set of external patterns (used as X, Y and Z) in experiment 4 was supplemented by an additional set of external patterns that could serve as S, T and U. The instructions that participants received were the same as in experiments 1 to 3, with the exception of the fact that there were four sprays (red, blue, green and purple) and that these sprays could be activated by four keys (T, U, B and M, respectively). The feedback was changed to reflect the fact that there were now four sprays. For half of the participants, the second patterns in the similar sequences were eliminated by the red spray (for BX, BY and BZ) and purple spray (for DS, DT and DU) and the second patterns in dissimilar sequences were eliminated by the blue spray (for DX, DY and DZ) and green spray (BS, BT and BU). For the remaining half, the assignment of pairs of sprays (red and purple or blue and green) that would eliminate the second patterns in similar and dissimilar sequences was reversed. There were 20 presentations of each of the twelve sequences and there were no more than two sequences of the same type (similar or dissimilar) in succession. To assess the effect of similarity on learning, I combined the percentage correct scores from both types of similar trials and from both types of dissimilar trial.

2.4.3. Results and Discussion

The mean percentage correct scores for the first patterns (S1 and D1) and second patterns (involving S2 and D2) are shown in Figure 12. Inspection of the left-hand panel reveals that over the course of training there was an increase in the percentage correct for S1 and S2 but that the scores for D1 were consistently higher than for S1. ANOVA revealed an effect of block, $F(5,315)=9.42$, $p<.001$, an effect of similarity (S1 or D1), $F(1,63)=6.54$, $p<.05$ and no interaction between these factors, $F(5,315)=1.24$, $p>.28$. Inspection of the right-hand panel of Figure 12 indicates that the scores for S2 and D2 increased over the course of training and that this increase was similar in magnitude. ANOVA confirmed that there was an effect of block, $F(5,315)=129.88$, $p<.001$, no effect of similarity, $F<1$ and no interaction between these factors, $F(5,315)=1.82$, $p>.10$. Finally, the reaction times for S1, D1, S2 and D2 are shown in the lower panel of Table 8. Inspection of this panel reveals that reaction times became shorter as training proceeded but were similar for the different types of first and second patterns (i.e. S1 and D1 and S2 and D2) and reaction times became faster as training proceeded. ANOVA conducted on the reaction times for S1 and D1 revealed an effect of block, $F(5,315)=24.36$, $p<.001$, no effect of similarity and no interaction between these factors, $F\text{s}<1$. Similarly, ANOVA conducted on the reaction times for the second patterns revealed an effect of block, $F(5,315)=20.62$, $p<.01$, no effect of similarity and no interaction between these factors, $F\text{s}<1$.

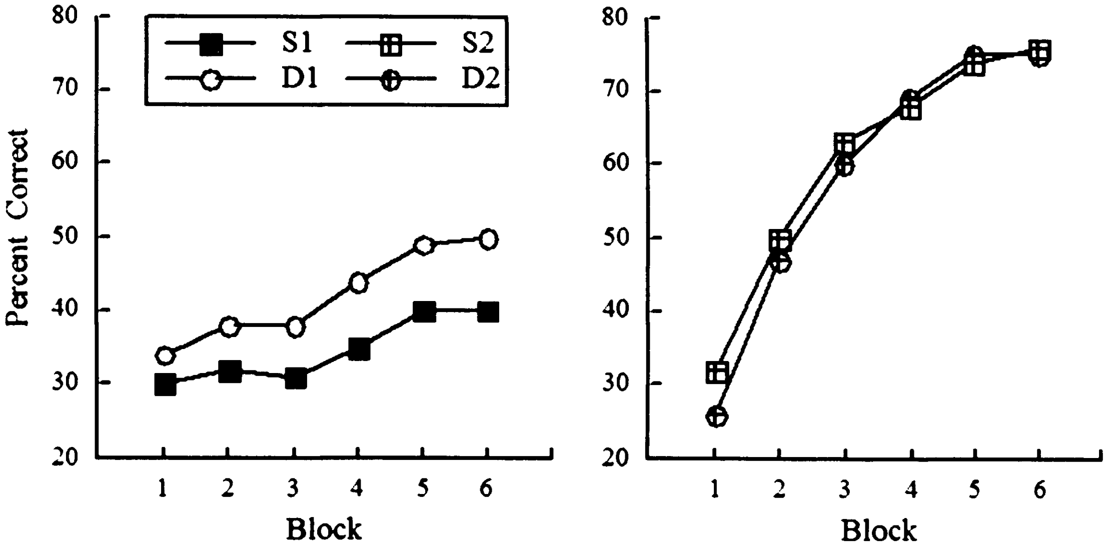


Figure 12. Experiment 4. Mean percentage correct for the first patterns (S1 and D1; left panel) and second patterns (S2 and D2; right panel).

Table 8: Mean reaction times for Experiment 4.

Block	**1**	**2**	**3**	**4**	**5**	**6**
S1	1008.53	611.75	573.33	514.54	496.60	419.46
D1	1005.20	682.12	479.09	466.60	450.89	434.48
S2	1311.48	1393.49	1337.76	1198.68	1050.33	926.08
D2	1342.85	1458.59	1432.59	1268.19	1101.48	992.83

Note: Mean reaction times to the first patterns (S1 and S2) and second patterns (S2 and D2) in Experiment 4.

The implications of the different influences of similarity on performance to the first patterns in the sequences in Experiments 2, 3 and 4 will be discussed in detail in the General Discussion that follows. However, before doing so, brief consideration should be given to another feature of the results of Experiments 2–4 that might appear puzzling; namely, that in none of the experiments was there any (significant) indication from either the percent correct scores or latencies that performance to the second patterns in the sequences varied as a consequence of whether they were preceded by a similar pattern or a dissimilar pattern. Of course, the latency scores might simply be an insensitive index of the associative processes that are in operation in Experiments 2–4. However, the same argument does not apply to the accuracy scores. In fact, Rescorla and Gillan's (1980) brief descriptions of the levels of responding shown by their pigeons to the first-order conditioned stimuli during second-order conditioning (equivalent to, for example, BX and DX in Experiments 2–4), there also appears to have been little difference depending on the similarity (see also, Holland & Ross, 1981). One possibile account of this puzzle is that the influence that the presentation of AX has on the processing of BX (cf. Rescorla & Gillan, 1980) has a much greater impact on learning (in this case to associate the components of AX with those of BX) than it does on performance generated by those aspects of BX that are linked to the (correct) response (e.g., red spray).

2.5. General Discussion

The aims of the experiments reported in this paper were twofold: to evaluate the effect of similarity on associative learning in humans using a novel procedure and to examine the origin of any influence of similarity on association formation. These aims were developed in the service of examining the processes involved in perceptual categorisation. In Experiments 2 and 3, patterns with feature A preceded similar patterns with feature B (AX→BX, AY→BY, AZ→BZ) and patterns with feature C preceded dissimilar patterns with feature C (CX→DY, CY→DZ, CZ→DX). Participants were required to learn that patterns with feature B (i.e., BX, BY, BZ) could be eliminated with one insecticide spray (e.g., red) and those with feature D (i.e., DX, DY, DZ) could be eliminated with a second spray (e.g., blue) and either spray was effective in eliminating patterns with features A and C. In Experiments 2 and 3, the participants learned which spray to use during patterns with features B and D and this learning was mirrored in their choice of which spray to use during patterns with features A and C. Critically, participants' spray choice for bugs with feature A came to match that used to eliminate bugs with feature B, but there was no such associative transfer between C and D. It is worth emphasising, at this point, that within this experimental design the effect of similarity is on learning rather than on simple generalisation between the patterns (which should be equivalent between, for example, BX and AX and DX and AX), or other nonspecific factors (see Hall, 1994).

As I have already noted, neither the elemental analysis of the effect of similarity on association formation offered by Rescorla and Gillan (1980) nor the configural analysis introduced in Section 1.2.2. of Chapter 1 have any difficulty explaining the results of Experiments 2 and 3 (in which the elements used to manipulate similarity, X, Y and Z, were irrelevant). However, in Experiment 4, where X, Y and Z were now relevant, transfer of 'correct' responding was greater between dissimilar than similar patterns. These results are inconsistent with the configural analysis and, what is more, are also inconsistent with the more general suggestion that similarity promotes association. However, the analysis provided by Rescorla and Gillan (1980), that gives no special role to similarity *per se* in the process of association can be readily applied not only to the results of Experiments 2 and 3, but also to those of Experiment 4 (see Section 2.4.1.).

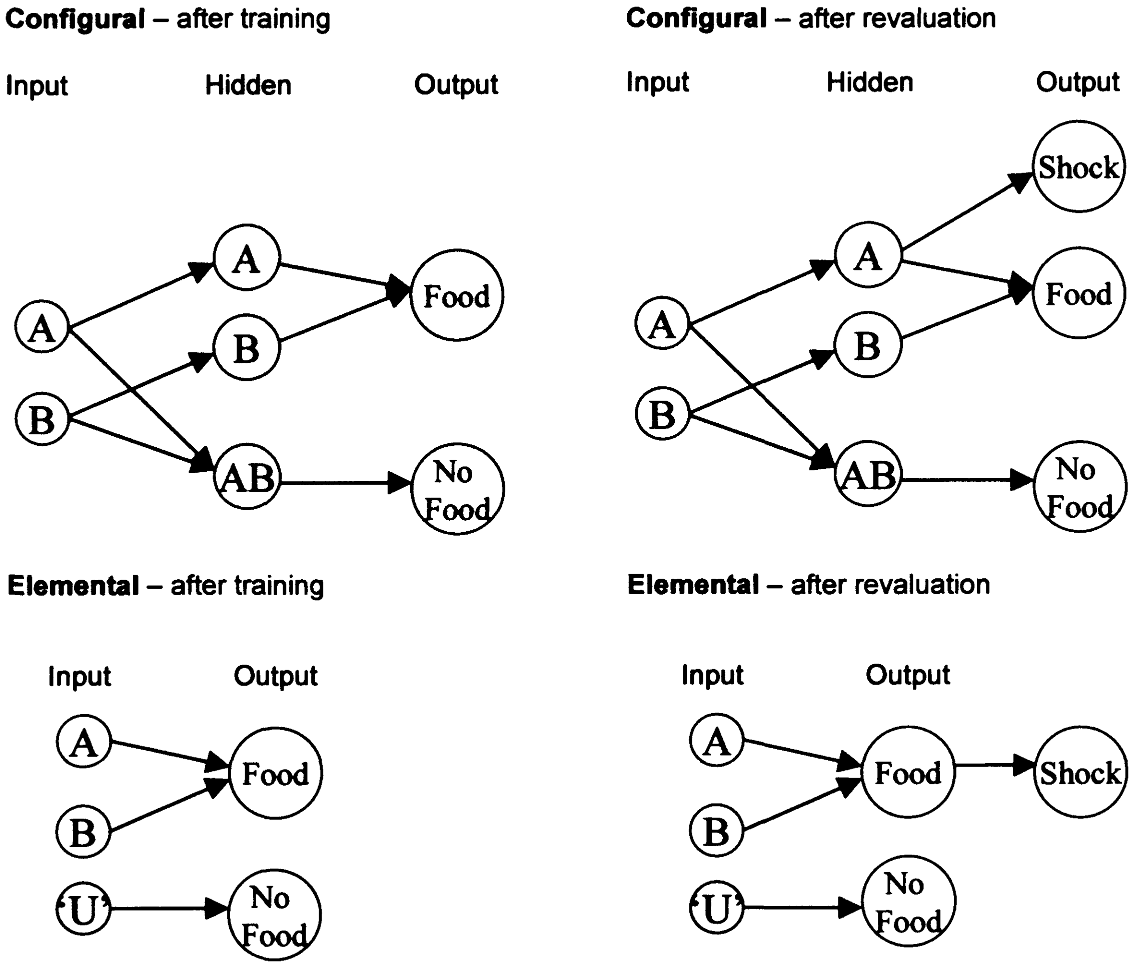
The fact that there is both empirical and theoretical convergence between studies of human and nonhuman animals offers general encouragement to those who imagine that the results from nonhuman animals might be of relevance to understanding human learning and memory. The results of Experiments 2–4, observed in humans, and those reported by Rescorla and Gillan (1980), observed in pigeons, provide no good reason to suppose that similarity should be given an independent status as a principle or law of (associative) learning. It remains to be seen whether under other, yet to be determined, conditions similarity might have an independent influence on learning. However, at least under those conditions that have been assessed, the influence of similarity on associative learning is readily interpreted as a secondary consequence of the operation of a principle of intensity – albeit one in which the relative effective intensity of the components of a stimulus compound can be altered by whether or not one of them has been recently presented (cf. Wagner, 1981).

Insofar as perceptual categorisation is concerned, the elemental analysis has proven to be rather powerful – it provides a general account for the results from studies of second-order conditioning in pigeons and the parallel studies described in this chapter. In the following chapter, I assess the adequacy of both the elementary and configural analyses in the context of studies of functional categorisation.

CHAPTER 3: FUNCTIONAL CATEGORISATION I

3.1. Introduction

In Chapter 1, I discussed several possible models of the nature of the associative structures that underlie the acquired equivalence and distinctiveness of cues, which are my indices of the processes that underlie functional categorisation. To briefly recap, I suggested that both elemental models (supplemented by a process of mediated conditioning) and configural models of the form developed by Pearce (1994) (supplemented by the suggestion that different patterns of stimulation can be grouped when they have similar consequences; cf. Honey & Ward-Robinson, 2002) could account for simple demonstrations of the acquired equivalence of cues (e.g., Honey & Hall, 1989). In considering these analyses I suggested that the investigation of patterning discriminations might be particularly illuminating. For example, in acquiring a negative patterning discrimination (A+, B+, AB-), a standard configural system will be forced to develop three configural units, one for each pattern, and these will be linked to the outcomes that they precede. If one were now to present A and pair it with shock, then the configural unit activated by A might well become linked to the outcome unit activated by shock. However, the presentation of B will not be able to activate the shock unit within the associative structures just described (see Figure 13). The situation is rather different for the elemental account (which includes mediated conditioning) in that it will allow B to contact a representation of shock through the association established between food and shock during A→shock training (see Figure 13). These distinct predictions apply equally readily to the case of positive patterning. Experiment 5 was designed to assess these different predictions concerning whether or not patterning training generates an acquired equivalence/distinctiveness effect. The remaining two experiments in Chapters 3 and the two experiments in Chapter 4 seek to examine the nature of the effects observed in Experiment 5; and, in particular, the associative structures that are involved.



Configural – after training

Configural – after revaluation

Elemental – after training

Elemental – after revaluation

Figure 13. Predictions derived from configural and elemental accounts of acquired equivalence/distinctiveness.

3.2. Experiment 5

3.2.1. Introduction

Experiment 5 was designed with a simple question in mind; namely, to evaluate whether or not an acquired equivalence/distinctiveness effect can be observed following patterning discriminations. A summary of the experimental design used in Experiment 5 is depicted in Table 9.

Table 9: Design of Experiment 5.

Group	Training	Revaluation	Test
Patterning	A+, B+, AB– C–, D–, CD+	A→shock C→no shock	B? D?
Control	A+, B+, AB+ C–, D–, CD–	A→shock C→no shock	B? D?
Or			
Patterning	A-, B-, AB+ C+, D+, CD-	A→shock C→no shock	B? D?
Control	A-, B-, AB C+, D+, CD+	A→shock C→no shock	B? D?

Note: A, B, C, and D refer to elemental contexts, AB and CD denote compound contexts, + and - indicate food and no food, respectively, and shock refers to the delivery of shock. Whether the revalued context, A, has been previously used as part of negative patterning (upper panel) or positive patterning (lower panel) was counterbalanced in this experiment and subsequent experiments.

In Experiment 5, rats initially received training in four element contexts (A, B, C and D) and two compound contexts (AB and CD). For group Patterning, when rats were placed in contexts A and B they received food pellets and when they were placed in AB no food was presented (i.e., a negative patterning discrimination); when the same rats were placed in C and D they received no food and when they were placed in CD they received food (i.e., a positive patterning discrimination). For group Control, whenever the rats were placed in A, B or AB they received food and when they were placed in C, D, or CD they did not receive food. The inclusion of this control group was intended to assess whether or not my procedures were capable of generating a basic acquired equivalence/distinctiveness effect, that could be explained in a variety of more or less interesting ways. Following the acquisition of these discriminations, rats were placed in context A, where they received shocks, and context C, where they received no shocks. Finally, rats were placed in contexts B and D and their activity levels were video-recorded and then scored in order to assess the level of generalised fear. In particular, I assumed that greater generalisation would be evident in less activity.

3.2.2. Method

Subjects. Charles River Ltd UK, supplied 32 naïve Lister hooded rats that were maintained at 80% of their *ad lib* weight (*M*=367g; range=332–389g) by receiving restricted amounts of food supplied by Harlan Tekland (Bicester, Oxfordshire UK). The rats were housed in pairs in a colony room illuminated between 8 am and 8 pm, where they received unrestricted access to water when in their home cages. Testing began at, approximately, 11 am.

Apparatus. Eight operant chambers (Campden Instruments Ltd, Loughborough, England: Test Chamber CI-410) were used. These chambers were arranged in two separate 2 × 2 grids housed in different experimental rooms. Each chamber (24.5 cm × 23 cm × 21 cm;

W×D×H) was positioned within a box having three aluminium walls and an aluminium ceiling, the door of which remained open. The front wall of the chamber was made of transparent Perspex and also served as the door to the chamber. There was a food well in the left-hand wall (5cm × 4cm × 6cm: W×D×H) into which 45 mg food pellets (supplied by PJ Noyes, Lancaster, NH) could be delivered. A top-hinged transparent plastic flap guarded access to this food well. When this flap was moved by approximately 2 mm, a magazine entry response was recorded. A 3W light bulb, positioned in the centre of the ceiling panel of each chamber provided local illumination and the chambers received additional ambient illumination from the striplight in the experimental room. Both experimental rooms also had a Panasonic movie camera (model no. NV-M40) mounted on a tripod that was positioned, approximately, 1.5 m from the chambers.

In one room, each experimental chamber served as a single element context (A, B, C and D). The right-hand pair of chambers (visual contexts) were lined with either spotted (top) or checked (bottom) wallpaper, mounted behind transparent plastic panels (for details of the wallpapers, see Honey & Watt, 1999). The left-hand pair of chambers (thermal contexts) were created by replacing the standard 16-bar grid floor fitted in the visual contexts' chambers (stainless steel bars, diameter 0.47 cm, spacing between bar centres 0.93 cm), with an aluminium sheet floor (24.5 cm × 23.5 cm). This floor had a bracket attached to the underside into which two Thermos picnic blocks, either heated or cooled, could be inserted thus creating the "warm" (35°C, top chamber) or "cool" (10°C, lower chamber) contexts (for a full description, see Ward-Robinson & Honey, 2000). The walls of these thermal contexts remained undecorated. The second experimental room contained the compound contexts (AB and CD). The top right chamber contained the spot and cool elements, the bottom right chamber contained the check and warm elements, the top left chamber contained the spot and warm elements and the bottom left chamber contained the check and cool elements. The element and compound contexts were placed in separate rooms because limitations in space prevented all contexts being placed in the same room. As it transpires, the fact that the element and compound contexts were in different rooms does not affect, in a theoretically interesting way, the interpretation of the results that follow (see Sections 4.3.1 and 4.3.2).

Appetitive training. On days 1 and 2, rats received magazine training consisting of two 20-min sessions, in which 20 food pellets were delivered on a variable time (VT), 60-s schedule. In the first session, the door to the food well was taped up, allowing the rat unrestricted access to the pellets and during the second session the door was returned to its normal position. During these two days the chambers were undecorated and had grid floors.

The rats were then split into two equal-sized groups for the next 16 days of training. For all rats, A and C were visual contexts and B and D were thermal contexts. For half of the rats in each group, A was 'spot' and C was 'check' and for the remainder this arrangement was reversed. For half of the rats created by the previous counterbalancing operation, B was 'warm' and D was 'cool' and for the remainder this arrangement was reversed. Group Patterning received both a negative patterning discrimination and a positive patterning discrimination. For half the rats in group Patterning, A, B and AB were employed as the stimuli for a negative patterning discrimination and C, D, and CD were used as the stimuli in a positive patterning discrimination (see Table 9); and for the remaining participants this arrangement was reversed. Similarly, for half of the rats in group Control, A, B and AB were reinforced and C, D, and CD were not (see Table 9); and for the remainder this arrangement was reversed.

Rats received six 5-min trials on each day of training, four in the elemental training room and two in the compound training room. The elemental trials were run in one session (e.g., in the morning, first session) and the compound trials in a different session (e.g., in the afternoon, second session); with the sessions being separated by, on average, three hours.

These two types of session were run alternately across days, so that if on day one a rat received the elemental session in the first session and compound training in the second session, on day two it received the reverse order of element and compound training. On day one, half of the rats received their elemental session first and the remainder received the compound session first. Each 5 min session was split into 30 s periods. On a reinforced trial, two food pellets were delivered at the end of each 30 s period; there were thus ten periods that each ended in the delivery of two food pellets. No food pellets were delivered on non-reinforced trials. The rate of magazine entries during the first 30 s period of each trial type (in which no food pellets were delivered) served as the index of discrimination learning. In each successive 4-day block, the orders in which the rats were placed in contexts A-D were pseudo-random with the constraints that each context could occupy each position (1st, 2nd, 3rd and 4th) and that each context was equally likely to be immediately preceded or succeeded by each of the other three contexts. Similarly, the order in which the compound contexts were presented within a session alternated across days (AB, CD, CD, AB ... for half of the rats and CD, AB, AB, CD for the remainder).

Revaluation and test trials. On each of the next two days, rats received two 3 min sessions, one in A and the other in C. On each occasion on which rats were placed in A, they received three 0.5 s presentations of a .5mA shock, the first of which was delivered one minute after the rat was placed in context A; the remaining two shocks followed at one-minute intervals. When rats were placed in C, they received no shock. On day one, half of the rats received A first and C second and for the remainder this arrangement was reversed. On day two, those rats that had received the sequence AC on day one received the sequence CA on day two. On each of the next two test days the rats again received two 3 min sessions, one in B and the other in D. On the first test day, half of the rats received the test order BD and the other half received DB; on the second day, those that had received BD received DB and vice versa. Within a day the revaluation and test sessions were separated by, on average, three hours. These test sessions were videotaped to enable us to gain a measure of generalised fear.

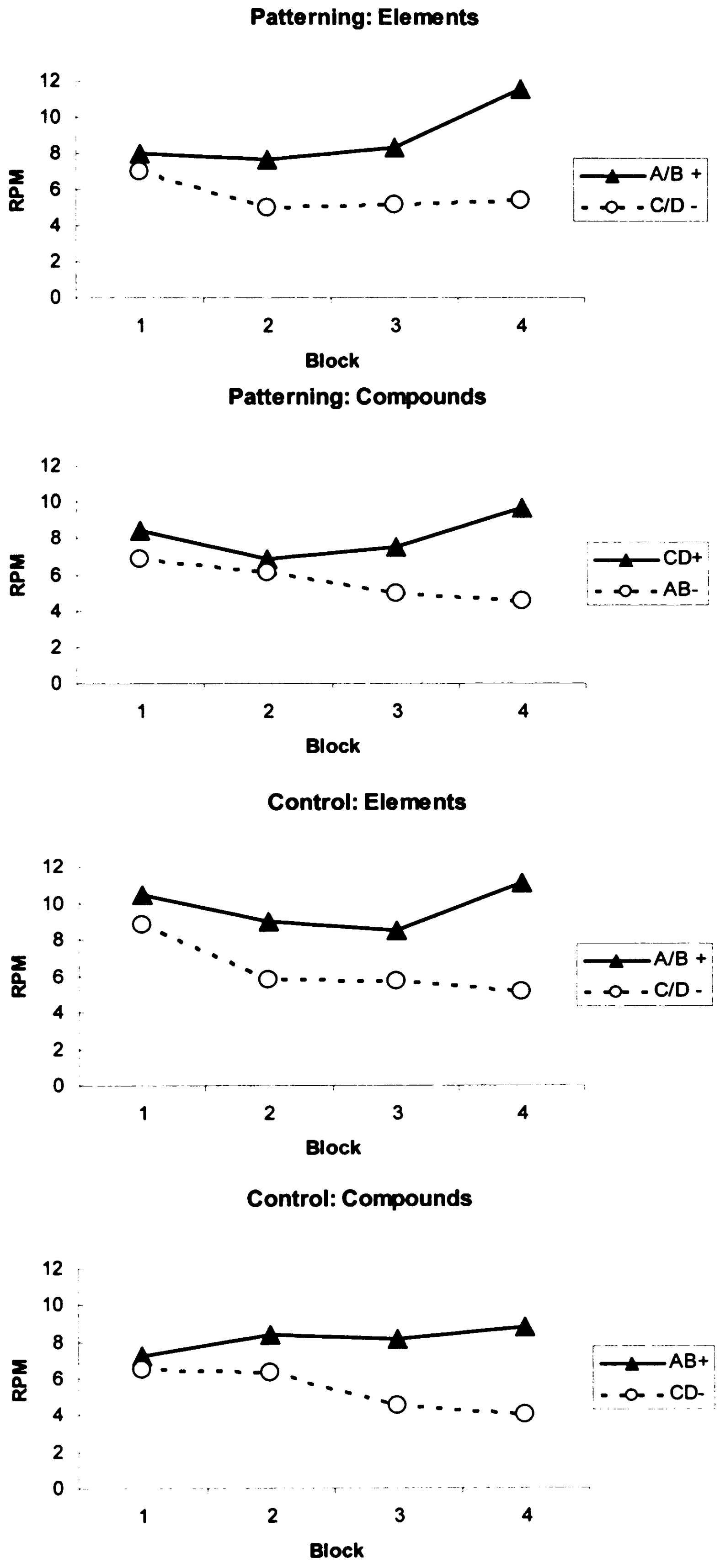
Behavioural measures. To assess the effects of appetitive training I used the rates of magazine entries on the various trial types, whereas to assess generalised aversive conditioning, the levels of activity that B and D provoked were monitored. In this experiment, and subsequent experiments, the results of appetitive training were pooled into 4 successive blocks of 4 days. It was assumed that the lower the level of activity the greater the amount of generalised fear. Activity was defined as any movement of the rat that was not necessary to support breathing (e.g., fluctuation in the volume of the thorax). To do so, while watching each video, I held a computer mouse button depressed while the rat was active and released this button when it was inactive. The computer generated a trace of this activity in two-second bins recorded as 1s (activity) and 0s (inactivity), which were converted into rates of activity per minute (RPM; with a maximum of 30 rpm). The scores from a given test in this experiment and subsequent experiments were pooled into 9 successive blocks of 10 bins (i.e., 9, 20 s bins).

I assessed the validity of this new method by correlating the scores gained from it with those from an independent scorer who used a traditional method (i.e., in which a metronome indicates the start and end of each bin and the scorer records bins within which activity occurs). The Pearson product moment correlation for the scores using the 2 methods for 16 rats was .894, $p<.001$.

3.2.3. Results

Figure 14 shows the mean rates of magazine entries in groups Patterning (upper two panels) and Control (lower two panels). Inspection of the 1st and 3rd panels reveals that both groups of rats came to respond at a higher rate to the reinforced elements (A and B in Figure 14) than to the non-reinforced elements (C and D in Figure 14) and that there was little difference in the levels of responding between the groups. Turning to the 2nd and 4th panels, it is apparent that groups Patterning and Control responded at higher rate to the reinforced compounds (CD for group Patterning and AB for group Control in Figure 14) than to the nonreinforced compounds (AB for group Patterning and CD for group Control in Figure 14); there was also some indication that the reinforced compound reached asymptote earlier in training in the Control than in the Patterning group. ANOVA with group (Patterning or Control), stimulus type (elements or compounds), training set (A/B and AB versus C/D and CD) and block (1-4) revealed no effect of group, $F<1$, and effects of stimulus type, $F(1,30)=16.63, p<.001$, training set, $F(1,30)=26.85, p<.001$, and block, $F(3,90)=5.29, p<.005$. This analysis also revealed an interaction between group and stimulus type, $F(1,30)=4.93$, $p<.05$, group by training set, $F(1,30)=16.47, p<.001$, stimulus type and training set, $F(1,30)=25.60, p<.001$, and training set and block, $F(3,90)=3.62, p<.05$. There was no interactions between group and block, $F(3,90)=1.59, p>.19$, or stimulus type and block, $F(3,90)=2.40, p>.05$. There were also three-way interactions between group, training set and stimulus type, $F(1,30)=17.32, p<.001$, and training set, stimulus type, and block, $F(3,90)=3.53, p<.05$. There were no other significant three-way interactions, largest $F(1,30)=2.13, p>.10$. There was also a four-way interaction, $F(3,90)=3.31, p<.05$. All the interactions described above involving block are of little or no interest, with the notable exception of the three-way interaction-involving group. To analyse this interaction in more detail, separate ANOVAs were conducted on the two groups.

The separate ANOVA for group Patterning revealed a main effect of block, $F(3,45)=3.15, p<.05$, but no other main effects, largest $F(1,15)=1.58, p>.20$. This analysis also revealed an interaction between stimulus type and training set, $F(1,15)=47.91, p<.001$, but no other two-way interactions, largest *F*(1,15)=1.33, *p*>.27. There was also a three-way interaction, *F*(3,45)=8.03, *p*=.001. The interaction between stimulus type and training set was explored using an analysis of simple main effects which revealed that the elements and the compounds differed between the training sets, and that there was a difference between the elements and the compound for both training sets, smallest *F*(1,15)=21.55, *p*<.001. The separate ANOVA for group Control revealed that each of the main effects were significant, smallest *F*(3,45)=3.92, *p*<.05. This analysis also revealed no interaction between stimulus type and training set, *F*<1, but interactions between stimulus type and block, *F*(3,45)=2.98, *p*<.05, and a significant interaction between training set by block, *F*(3,45)=4.79, *p*<.01. There was no three-way interaction, *F*<1.



Patterning: Elements

Patterning: Compounds

Control: Elements

Control: Compounds

Figure 14. Experiment 5. Magazine entry scores (in responses per minute; RPM) during training for groups Patterning (upper two panels) and Control (lower two panels).

The results from the test are shown in Figure 15. Inspection of the two panels reveals that in both groups Patterning and Control, rats showed less activity (i.e., more evidence of fear) during B than D. ANOVA confirmed that there was a main effect of stimulus type, *F*(1, 30)=8.03, *p*<.005, an effect of block, *F*(8, 240)=2.89, *p*<.05, and an interaction between group and block, *F*(8,240)=2.19, *p*<.05, which merely reflects the fact that the scores for group Patterning decline across the session whereas those for group control do not. There were no other significant effects or interactions, largest *F*(8,240)=1.26, *p*>.26.

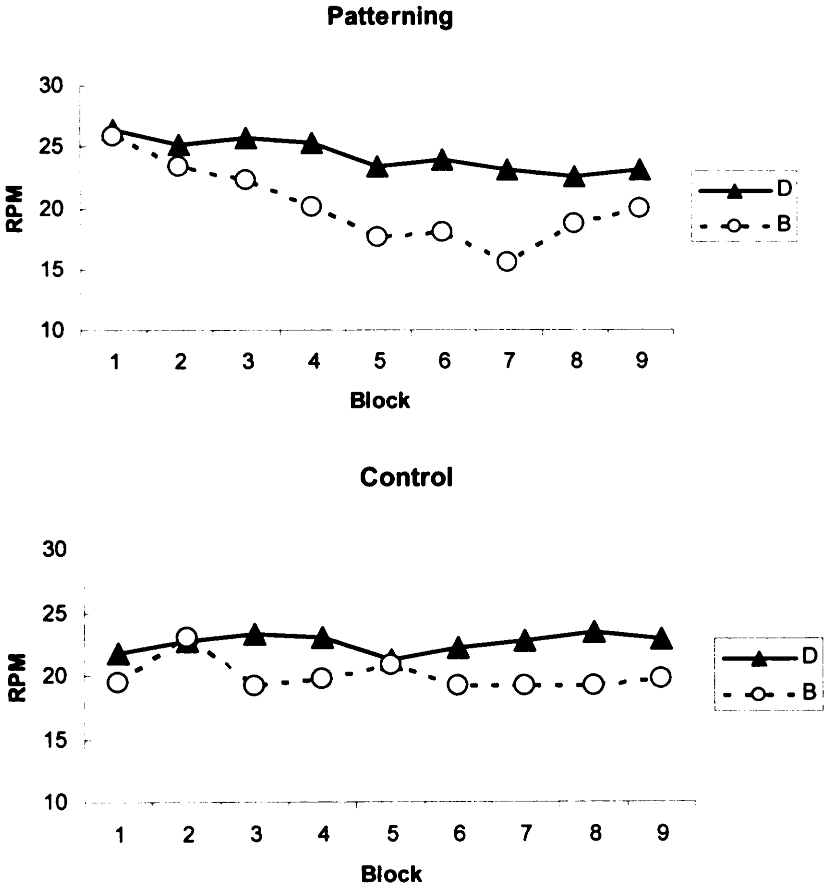


Figure 15. Experiment 5. Mean levels of activity when rats in groups Patterning and Control were placed in contexts B and D.

3.2.4. Discussion

Experiment 5 examined whether an acquired equivalence/distinctiveness effect could be observed after a patterning discrimination. Rats first received either patterning training (e.g., A+, B+, AB-, C-, D-, CD+) or control training (e.g., A+, B+, AB+, C-, D-, CD-). After the rats had acquired these discriminations, they received shocks in context A, but not in context C. During the critical test, rats in both groups were more likely to show evidence of generalised fear in context B than in context D. In short, an acquired equivalence effect of a similar magnitude was observed irrespective of the nature of the training (patterning or control) that the rats had received. These results are of theoretical interest, because whereas a simple configural account of acquired equivalence fails to predict that the effect will be observed after patterning training (see Figure 13), an elemental model (supplemented by mediated conditioning) predicts that such an effect will be observed. Briefly, the elemental analysis of the results of Experiment 5 is that when A is paired with shock, it will activate a representation of food. As a result, this representation of food will become linked to that of shock and permit other stimuli that can also activate a memory of food (e.g., B) to elicit fear.

It might, however, be deemed premature to abandon the configural approach at this point, if for no other reason than it provides an account of other, complex instances of acquired equivalence/distinctiveness that are beyond an elemental analysis. In fact, the modification of a strict configural theory (e.g., Pearce, 1994), which provides a possible account of complex instances of acquired equivalence/distinctiveness (e.g., Honey, 2000) can also provide an account for the results of Experiment 5. The feature of this type of model that allows it to do so is that there are feedback or reciprocal links between the outcome and hidden-layer nodes. In the present case, these links will have the consequences depicted in Figure 16.

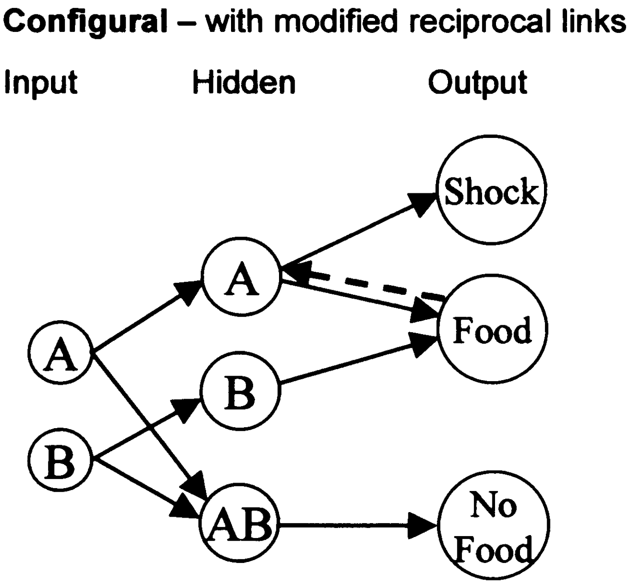


Figure 16. A possible modification to a configural model that allows it to provide a potential account for the results of Experiment 5. The dotted link illustrates the reciprocity of one of the links that are presumed to form between the hidden layer units and the output units.

Once the network has been trained it will have the pattern of links shown in Figure 16 involving the input units (A and B), the configural units (A, B, and AB), and the outcome units (food and no food); importantly, reciprocal links are assumed to form between hidden-layer units and output-layer units (only one of which is shown). When A is then paired with shock, hidden unit A will become linked with the output layer unit activated by shock. Now, when B is presented it will activate hidden unit B, which will then activate the output unit for food. The presence of the reciprocal link between food and the A hidden unit will allow this unit to become active and to thereby activate the shock output unit.

Although the analysis presented in the preceding paragraph provides, at least in principle, an analysis of the results of Experiment 5, there are several problems with it: first, if one has a winner-take-all learning and/or performance rule (for the hidden-layer units) then it is rather difficult to imagine that A's hidden unit can become active through B's influence on its hidden layer unit. Second, there is the issue of signal loss: it is a common assumption that the propagation of activity through a network of links will decline as the set of links increases (e.g., Wagner, 1981). The circuitous nature of the links that are supposed to underpin the results from group Patterning in Experiment 5, coupled with the strength of the effect seen in Experiment 5 do not sit comfortably with one another. This is not to say, of course, that the analysis is fatally flawed, but the problems identified above have already raised some concerns about its plausibility.

3.3. Experiment 6

3.3.1 Introduction

In Experiment 6, I attempted to establish the reliability of the results in Experiment 5, in particular to establish, using a within-subjects design, that the extent of acquired equivalence/distinctiveness is similar following control training (e.g., C+, D+, CD+) and patterning training (e.g., A-, B-, AB+; see Table 10). Briefly, following these two forms of training, rats received a revaluation procedure in which both contexts A and C were paired independently with shock. During the test rats received presentations of B and D. On the basis of the results of Experiment 5, rats should show equal levels of fear to B and D.

Table 10: Design of Experiment 6.

Group	Training	Revaluation	Test
Positive	A-, B-, AB+ C+, D+, CD+	A→shock C→shock	B versus D AB versus CD
Negative	A+, B+, AB- C-, D-, CD-	A→shock C→shock	B versus D AB versus CD

A second aim of Experiment 6 was to assess the two accounts outlined in Section 3.2.4 for the results of Experiment 5: the mediated conditioning account and the modified configural account. Both accounts would have no difficulty in explaining the fact that B and D would evoke similar levels of fear during the test. However, in Experiment 6 there were also test trials with both of the original test compounds (i.e., AB and CD). According to the mediated conditioning view, the revaluation procedure will result in both the food output unit and the no food output unit becoming connected to the output unit activated by shock (see Figure 17). This will mean that when the compounds are presented they should elicit fear of equal or greater magnitude than the elements: not only will the compounds elicit fear by virtue of direct generalisation from A and C to AB and CD, respectively, but because the unique elements of the compounds will activate an output unit that has become linked to shock (see Figure 17). The modified configural account, however, makes a different prediction. On the one hand, following control training (e.g., C-, D-, CD-) and revaluation with C, when CD is presented it will elicit more marked levels of fear than D both because (1) the CD node is directly linked to shock and (2) that node is receiving two sources of input. On the other hand, following patterning training (e.g., A+, B+, AB-) and revaluation with A, when AB is presented it is possible for it to elicit fear, but the level of fear might be expected to be somewhat less than to CD, because of the circuitous nature of the links involved: AB should activate the AB node, and in turn the no food output unit, which can then activate the CD hidden-layer node, and finally the shock node. The obvious additional implication of these two statements is that CD should elicit more fear than AB. This route from the presentation of AB to shock is clearly more elliptical than that between CD and shock. Moreover, this configural analysis seems rather implausible (see Section 3.2.4). Both analyses, of course, ignore the possibility that AB and CD will activate shock more directly, through A and C's more direct link (e.g., Kehoe, 1988) to shock that should contribute equally to the fear elicited by AB and CD, respectively.

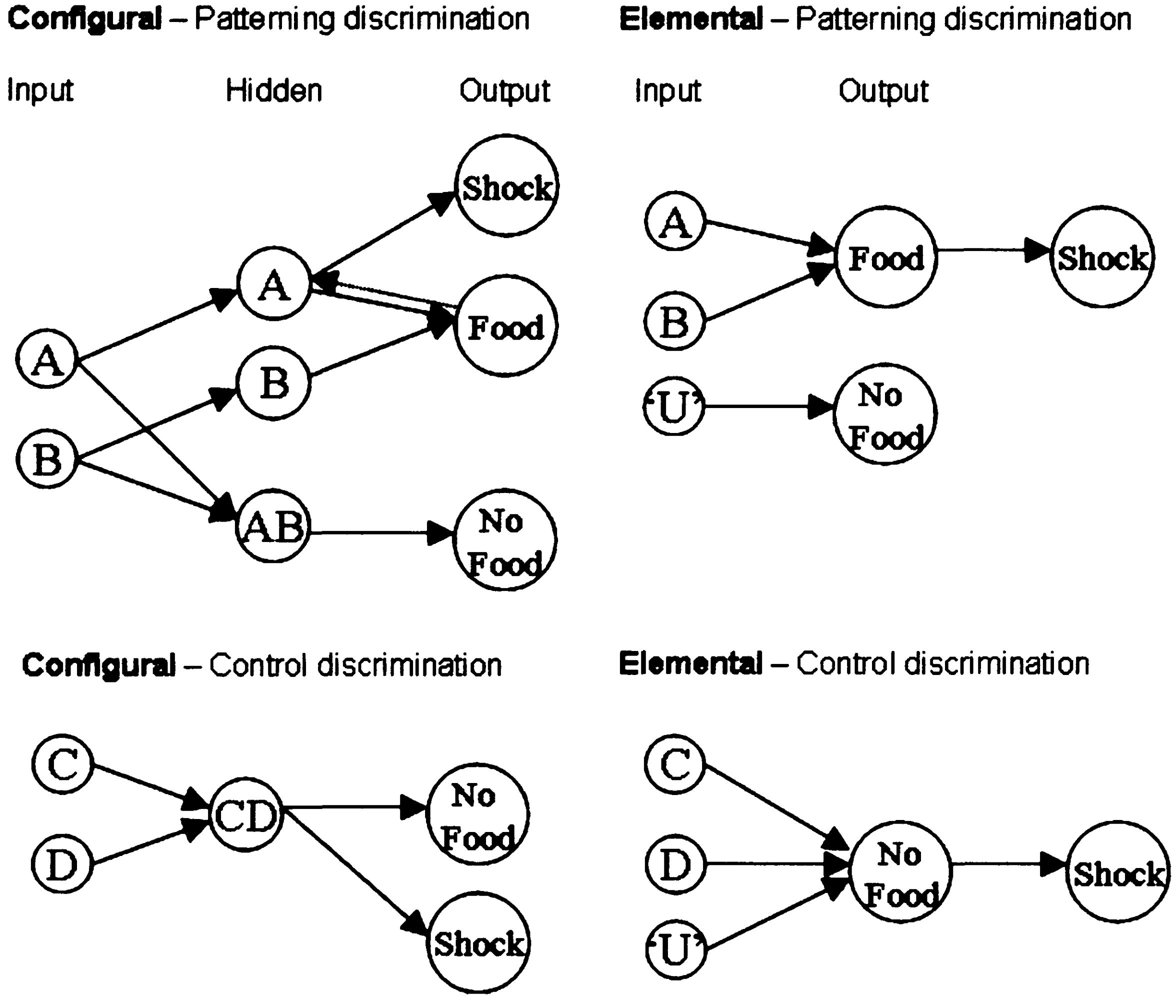


Figure 17. Modified configural and elemental accounts of the structures formed after patterning training (summarised in Table 10). Note that for presentational purposes, the patterning and control conditions are presented separately.

3.3.2. Method

Subjects and apparatus. 32 naïve Lister hooded rats from the same source as Experiment 5 were maintained at 80% of their *ad lib* weights (*ad lib* weight M=387g; range=346–391g). The rats were housed in the same way and the experiment was conducted between the same times of day as Experiment 5. The experimental apparatus was that used in Experiment 5.

Appetitive training. All rats received magazine training in the same way as in Experiment 5 and over the following 16 days they received appetitive conditioning. Group Positive (see Table 10) received positive patterning with one pair of contexts (e.g., A-, B-, AB+) and a control treatment with a second pair of contexts (e.g., C+, D+, CD+). Group Negative (see Table 10) received negative patterning with one pair of contexts (e.g., A+, B+, AB-) and a control treatment with the other pair (e.g., C-, D-, CD-). For half of the rats in each group, A was the spotted context and C was the checked context and for the remainder this arrangement was reversed. For half of the rats created by the previous counterbalancing operation, B was warm and D was cool and for the remainder this arrangement was reversed. The procedure was the same as Experiment 5, with the notable exception that one pair of contexts (e.g., A and B) was used in a patterning discrimination (positive or negative) whereas the other pair of contexts received a control treatment in which the contexts, whether presented alone or in compound, were followed by the same outcome. Other details of appetitive training were identical to those described in Experiment 5.

Revaluation and test trials. The revaluation procedure was conducted in the same way as in Experiment 5, with the exception that both contexts, A and C, were paired with shock. The rats received two days of testing. On the first of these two test days, half the rats in each of the groups (Positive and Negative patterning) received 2 test sessions in the element room and the remainder received 2 test sessions in the compound room. On the second of the test days, those rats that had received test sessions in the element room on the first day received test sessions in the compound room and those that had received test sessions in the compound room received test session in the element room. On each test day, half of the rats that were tested in the element room were placed in B and then D and the remainder received the reverse order. The order in which a rat received the element tests (e.g., B followed by D) matched the order in which these elements appeared in the compound test (AB followed by CD). The interval between tests of a given type (element or compound) was the same as in Experiment 5. All other details of Experiment 6 were the same as those of Experiment 5.

3.3.3. Results

Figure 18 shows the mean rates of magazine entries to elements (1st and 3rd panels) and compounds (2nd and 4th panels) in group Positive (top two panels) and Negative (bottom two panels). Inspection of the 1st and 3rd panels, reveals that both groups of rats came to respond at a higher rate to the reinforced elements (A and B in group Negative and C and D in group positive) than to the non-reinforced elements and that there was little difference in the levels of responding between the groups. Turning to the 2nd and 3rd panels, it is apparent that group Negative responded at a lower rate to the compounds (AB and CD) than did group Positive and that in group Positive, rats tended to respond at a higher rate to CD than to AB; an effect that was most marked in the final block. ANOVA with group (Negative or Positive), stimulus type (elements or compounds), training set (A/B and AB versus C/D and CD) and block (1-4) revealed an effect of group, $F(1,30)=927.16$, $p<.001$, stimulus type, $F(1,30)=16.93$, $p<.001$, training set, $F(1,30)=48.97$, $p<.001$, and block, $F(3,90)=22.79$, $p<.001$. This analysis also revealed an interaction between group and stimulus type, $F(1,30)=35.36$, $p<.001$, stimulus type and training set, $F(1,30)=29.62$, $p<.001$, stimulus type and block, $F(3,90)=11.84$, $p<.001$ and training set and block, $F(3,90)=6.77$, $p<.001$. There were also three-way interactions between group, stimulus type, and block, $F(3,90)=12.07$, $p<.01$ and between stimulus type, training set and block, $F(3,90)=2.77$, $p<.05$. There were no other significant interactions, largest $F(1,30)=1.69$, $p=.20$. In order to examine the source of the three-way interaction involving group, separate ANOVAs were conducted on the two groups.

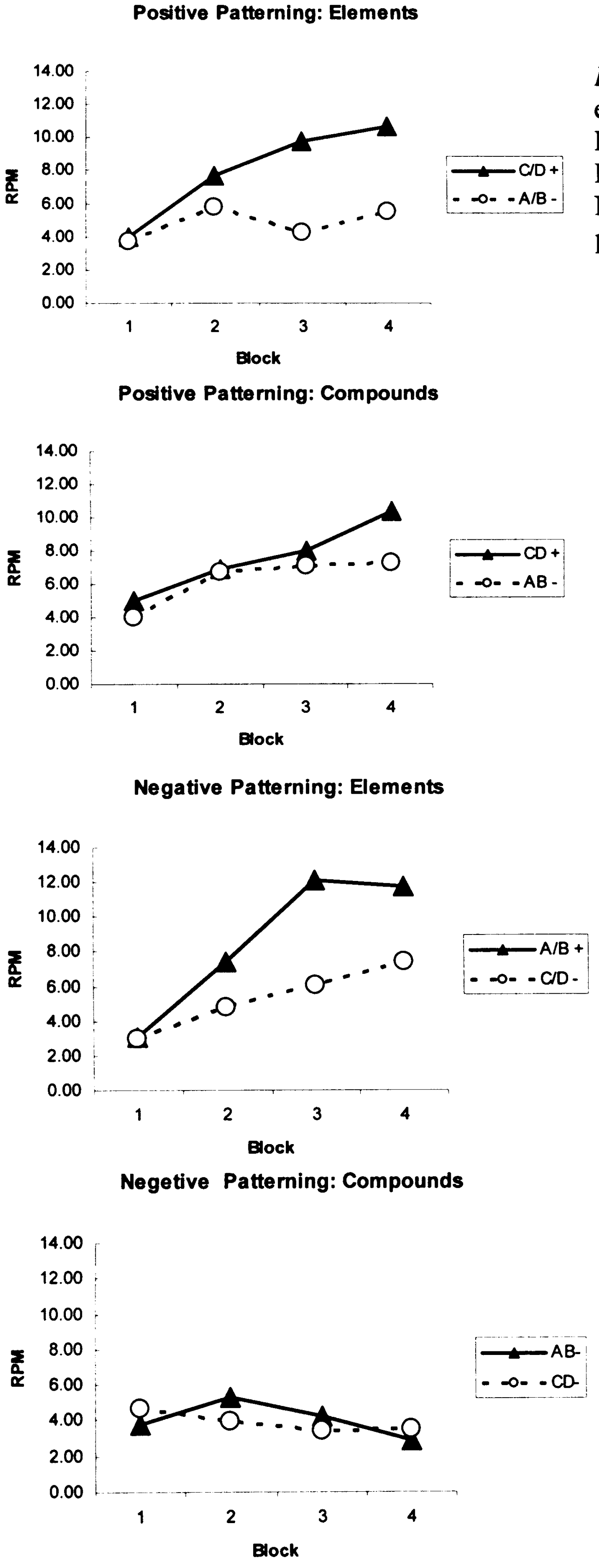


Figure 18. Experiment 6: Magazine entry scores (in responses per minute; RPM) during training for groups Positive patterning (top two panels) and Negative patterning (bottom two panels).

The separate ANOVA for group Negative revealed a main effect of stimulus type, $F(1,15)=75.38$, $p<.001$, training set, $F(1,15)=13.45$, $p<.005$, and block, $F(3,45)=7.83$, $p<.001$. This analysis also revealed interactions between stimulus type and training set, $F(1,15)=16.05$, $p<.005$, stimulus type and block, $F(3,45)=25.15$, $p<.001$, and between training set by block, $F(3,45)=4.13$, $p<.05$. There was no three-way interaction, $F(3,45)=2.13$, $p=.11$. Analysis of simple main effects (using the pooled error term) revealed that there was a difference between the reinforced and non-reinforced elements, $F(1,15)=23.78$, $p<.001$, no difference between the non-reinforced compounds, $F<1$ and that there were differences between both the reinforced elements (A and B) and the compound from that training set (AB), $F(1,15)=76.89$, $p<.001$ and a difference between the non-reinforced elements (C and D) and the compound from that training set (CD), $F(1,15)=7.25$, $p<.05$.

The separate ANOVA for group Positive revealed that there was no main effect of stimulus type, $F(1,15)=1.26$, $p=.28$, and that there were significant effects of training set, $F(1,15)=49.56$, $p<.001$ and block, $F(3,45)=16.34$, $p<.001$. This analysis also revealed interactions between stimulus type and training set, $F(1,15)=14.68$, $p<.005$, no interaction between stimulus type and block, $F<1$ and a significant interaction between training set by block, $F(3,45)=4.21$, $p<.05$. There was no three-way interaction, $F(3,45)=1.17$, $p=.33$. Analysis of simple main effects revealed that there was a difference between the reinforced and non-reinforced elements, $F(1,15)=108.42$, $p<.001$, a difference between the reinforced compounds, $F(1,15)=7.09$, $p<.05$ and that there were differences between both the reinforced elements (C and D) and the compound from that training set (CD), $F<1$, and a difference between the non-reinforced elements (A and B) and the compound from that training set (AB), $F(1,15)=14.80$, $p<.005$.

The results from the test are shown in Figure 19. Inspection of each of the panels reveals that irrespective of group (negative or positive patterning), rats showed similar levels of fear (i.e., activity) during B (involved in a patterning discrimination) and D (involved in a control treatment). Comparison of the top two and lower two panels reveals that there was more evidence of fear to B and D than to AB and CD. Inspection of all panels also reveals that the level of fear increased across the test. ANOVA confirmed that there was a main effect of stimulus type (elements versus compounds), *F*(1, 30)=28.83, *p*<.001, an effect of block, *F*(8, 240)=16.89, *p*<.001, and no other significant effects or interactions, largest *F*(1,30)=2.07, *p*=.16.

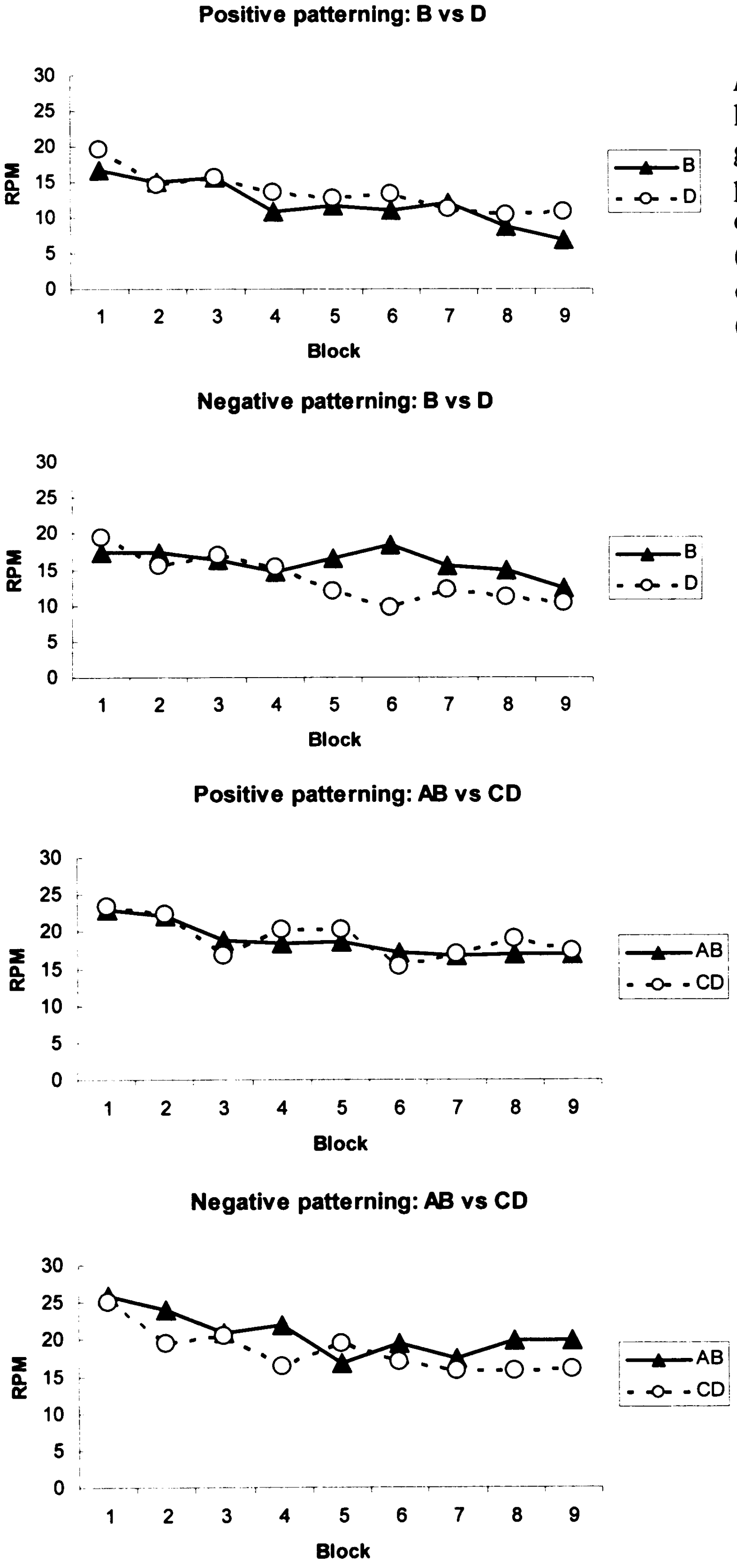


Figure 19. Experiment 6. Mean levels of activity when rats in groups Positive and Negative patterning were placed in element contexts B and D (upper two panels) and compound contexts AB and CD (lower two panels).

3.3.4. Discussion

As in Experiment 5, the acquisition of the discriminations proceeded uneventfully and successfully. Following revaluation training with A and C, as was anticipated on the basis of the results of Experiment 5, context B (a patterning element) elicited similar levels of fear to context D (a control element). This pattern of results held true for both kinds of patterning and control training. What is of more theoretical interest is the level of fear to the compound contexts. The levels of fear to AB (the patterning compound) and CD (the control compound) were statistically indistinguishable from one another, but significantly less than to B and D. This pattern of results is inconsistent with the predictions of the models developed in Section 3.3.1. To reiterate: the mediated conditioning view predicted that AB and CD would elicit greater fear than B or D and the modified configural also predicted this general pattern of results, with the additional prediction that CD would elicit more fear than AB. Paradoxically, revaluation of A and C resulted in greater fear to B and D than AB and CD.[1] Given the surprising nature of this pattern of results, I sought to assess further the merit of the elemental and configural accounts using a different experimental design.

---

[1] One potential problem in taking these results at face value reflects a procedural aspect of Experiment 6. In particular, the rats always received revaluation training in the room in which the elements had been trained and were later tested. This might have resulted in an overestimation fear to B and D relative to AB and CD. This issue is one of those dealt with in Chapter 4.

3.4. Experiment 7

3.4.1. Introduction

The design of Experiment 7 is depicted in Table 11. The initial training was formally identical to Experiment 6 with the notable exception that two of the contexts (warm and cool) were replaced with contexts (smell and object). This change allowed the use of grid floors throughout in both the element and compound contexts. This change in apparatus allowed me to pair the compounds (AB and CD) with shock during the second stage of the experiment. As in Experiment 6, after revaluation training all rats then received test trials with the elements (B and D) and the compounds (AB and CD). The effect that conditioning the compounds will have on test performance is of theoretical importance for the reasons given below.

Table 11: Design of Experiment 7.

Group	Training	Revaluation	Test
Positive	A-, B-, AB+	AB→shock	B versus D
	C+, D+, CD+	CD→shock	AB versus CD
Negative	A+, B+, AB-	AB→shock	B versus D
	C-, D-, CD-	CD→shock	AB versus CD

First, consider predictions that can be derived from the elemental analysis. According to this account, the training that group Negative receive will result in the pattern of associations depicted in the left-hand side of Figure 20; where A and B are linked to food, and C, D and the unique elements activated by both AB (U1) and CD (U2) are linked to no food. When AB and CD are subsequently paired with shock, this will result in A, B, C, D, U1, U2 and the no food unit becoming linked to shock. The food unit will not be linked to shock, because the AB trial activation of U1 will result in the no food being activated and thereby inhibiting the food unit. Now, when AB and CD are presented at test they will elicit similar levels of fear through the associations with shock involving A, B, U1, and the no food unit (for AB) and C, D, U2 and the no food unit (for CD). More interestingly, when B and D are presented at test, B will elicit fear to the extent that it was been directly linked to shock, whereas D will elicit fear for two reasons: because it is directly linked to shock and as a product of the fact that the representation of no food has become linked to shock. As in previous experiments, the same form of predictions can be derived for the positive patterning discrimination.

The modified configural analysis results in the connections depicted in the right-hand panel of Figure 20. During training in group Negative, configural units activated by A and B will become linked to food, whereas the unit activated by AB and that activated by C, D, and CD, will become linked to no food. When AB and CD are subsequently paired with shock, this will result in their configural/hidden units becoming linked to shock; when AB and CD are subsequently presented they will elicit similar levels of fear through these configural associations. Now, when B and D are presented at test, B will elicit fear to the extent that it weakly activates AB, whereas D will elicit fear because it activates CD. On this basis, one might also predict greater fear to D than to B. (Any influence of the reciprocal connections between output and hidden-layer units would be constrained by the ability of B and D to activate the configural units, AB and CD.)

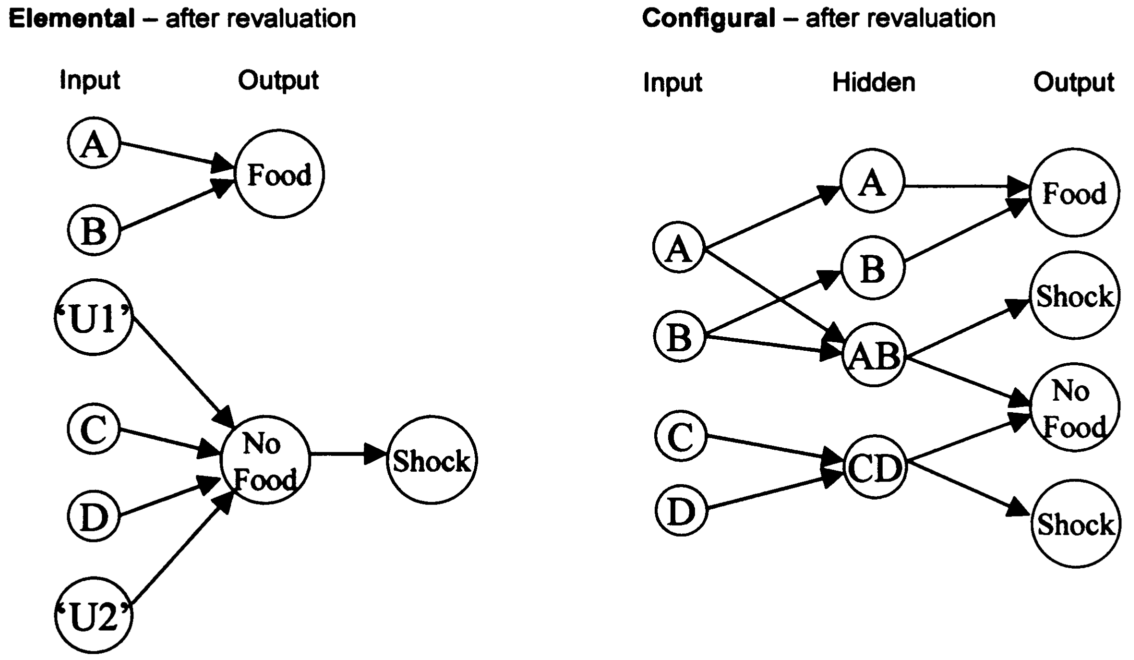


Figure 20. Application of elemental and configural models for Experiment 7. U1 and U2 refer to different unique elements. Direct links between the various input units and shock are not shown in the left-hand panel.

3.4.2. Method

Subjects and apparatus. 32 naïve Lister hooded rats from the same source as Experiment 5 (*ad lib* weight *M*=357g; range=321–375g). The rats were housed and maintained at 80% of their *ad lib* weight in the same way as Experiment 5. The experimental sessions took place at the same time of day as in Experiments 5 and 6. Two of the contexts were the same as those used in Experiments 5 and 6: the spotted and checked contexts. To enable presentation of the compound contexts to be paired shock, the warm context was replaced by the presence of a large (10x10x5 cm) glass object (a slide bath) placed on the floor in the right-hand corner of the test chamber, with the cavity facing the wall of the test chamber. The cool context was replaced by the presence of a red odour cube (supplied by Dale Air) with dimensions (5x5x5 cm), which dispensed a strong odour of vanilla. This cube was fixed, using Velcro, to the inside of the middle of the door of the test chamber, and was replaced every four days (these two types of contexts, object and odour, were used by

Allman, Ward-Robinson & Honey, (2004)). The solid aluminium floors of the warm and cool contexts were replaced with grid floors identical to those in the two visual contexts.

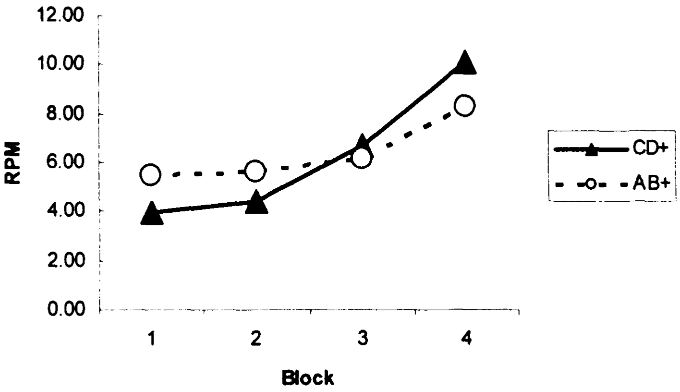
Appetitive training, revaluation and test. The rats were magazine trained in the same way as in previous experiments; details of appetitive training involving contexts A, B, C and D, and AB and CD were the same as in Experiment 6. The revaluation and test trials were conducted in an identical fashion to Experiment 6 with the notable exception that the compound contexts, AB and CD were paired with shock.

3.4.3. Results

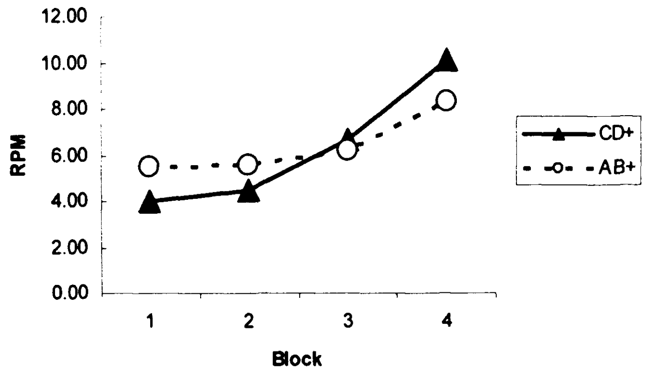
Training. Figure 21 shows the mean rates of magazine entries in groups Positive (top two panels) and Negative (bottom two panels). Inspection of this figure reveals a similar pattern to that seen in Experiment 6. Examination of the 1st and 3rd panels reveals that groups Positive and Negative came to respond at a higher rate to the reinforced elements (A and B for group Negative and C and D for group Positive) than to the non-reinforced elements and that there was little difference in the levels of responding between the groups. Turning to the 2nd and 4th panels, it is clear that group Positive responded at a higher rate to their reinforced compounds than group Negative responded to their nonreinforced compounds. ANOVA with group (Negative or Positive), stimulus type (elements or compounds), training set (A/B and AB versus C/D and CD) and block (1-4) revealed an effect of group, $F(1,30)=965.10, p<.001$, and block, $F(3,90)=7.06, p<.001$, but no effect of stimulus type, $F<1$, or training set, $F(1,30)=3.489, p>.07$. This analysis also revealed an interaction between group and stimulus type, $F(1,30)=43.65, p<.001$, group and block, $F(3,90)=12.56, p<.001$, group and training set, $F(1,30)=72.99, p<.001$, and between stimulus type and block, $F(3,90)=3.63, p<.05$. The remaining two-way interactions were not significant, $Fs<1$. Each of the four, three-way interactions were significant, smallest $F(3,90)=3.70, p<.05$, and there was also a four-way interaction between the factors,

$F(3,90)=7.99$, $p<.001$. In order to examine the source of the four-way interaction involving group, separate ANOVAs were conducted on the two groups.

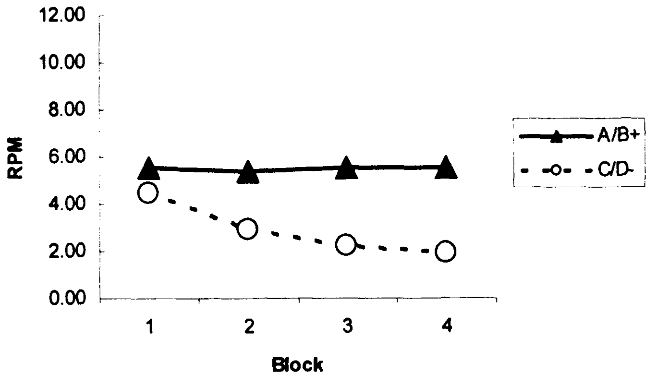
Positive Patterning: Compounds



Positive Patterning: Compounds



Negative Patterning: Elements



Negative Patterning: Compounds

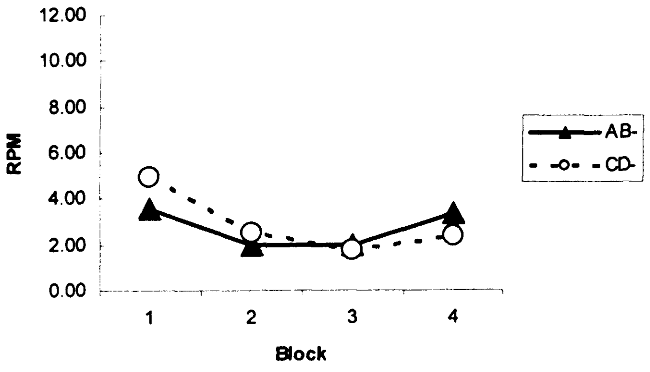
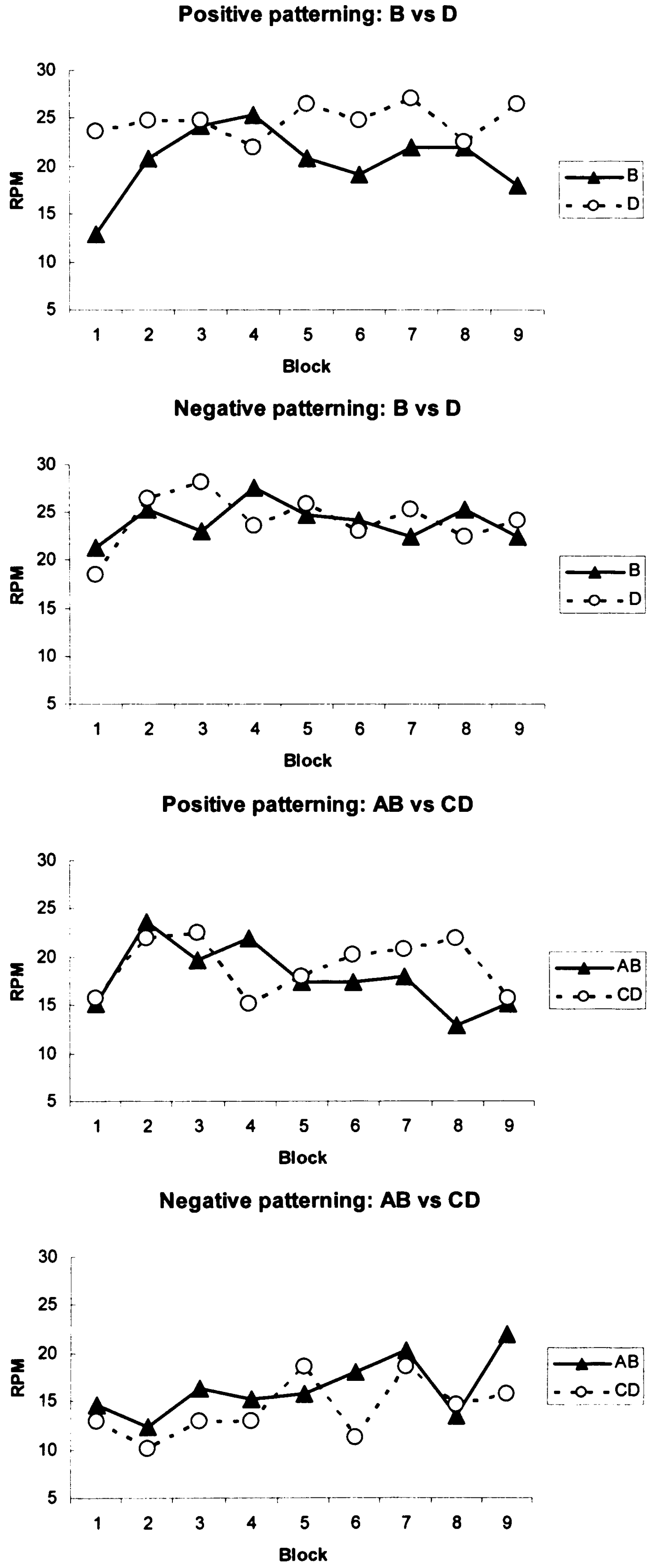


Figure 21. Experiment 7: Magazine entry scores (in responses per minute; RPM) during training for groups Positive patterning (Top two panels) and Negative patterning (bottom two panels).

The separate ANOVA for group Negative revealed a main effect of stimulus type, $F(1,15)=20.49$, $p<.001$, training set, $F(1,15)=19.66$, $p<.001$, and block, $F(3,45)=6.13$, $p<.005$. This analysis also revealed interactions between stimulus type and training set, $F(1,15)=25.73$, $p<.001$, and between training set and block, $F(3,45)=3.40$, $p<.05$. There was no interaction between stimulus type and block and no three-way interaction, largest $F(3,45)=1.29$, $p>.29$. Analysis of simple main effects conducted on the interaction between stimulus type and training set revealed an effect of stimulus type on the training set A/B/AB, $F(1,15)=45.07$, $p<.001$, but not on set C/D/CD, $F<1$, and an effect of training set on the elements, $F(1,15)=40.06$, $p<.001$, but not on the compounds, $F<1$. Analysis of simple main effects conducted on the interaction between training set and block revealed an effect of block on set C/D/CD, $F(3,45)=11.44$, $p<.001$, but not on set A/B/AB, $F<1$, and no differences between training sets on blocks 1 and 2, largest $F(1,15)=4.06$, $p>.06$, but differences between training sets on blocks 3 and 4, smallest $F(1,15)=31.09$, $p<.001$.

The separate ANOVA for group Positive revealed that there was an effect of stimulus type, $F(1,15)=17.26$, $p<.005$, and training set, $F(1,15)=28.14$, $p<.001$, and block, $F(3,45)=7.42$, $p<.001$. This analysis also revealed interactions between stimulus type and training set, $F(1,15)=31.19$, $p<.001$, stimulus type and block, $F(3,45)=14.27$, $p<.001$, and training set by block, $F(3,45)=8.34$, $p<.001$. There was no three-way interaction, $F(3,45)=1.01$, $p=.40$. Analysis of simple main effects conducted on the interaction between stimulus type and training set revealed an effect of stimulus type on training set A/B/AB, $F(1,15)=56.46$, $p<.001$, but not on training set C/D/CD, $F<1$, and an effect of training set on the elements, $F(1,15)=51.00$, $p<.001$, but not on the compounds, $F<1$. Analysis of simple main effects conducted on the interaction between training set and block revealed an effect of block on set C/D/CD, $F(3,45)=14.88$, $p<.001$, but not on set A/B/AB, $F(3,45)=123$, $p>.30$, and no differences between training sets on blocks 1 and 2, largest $F(1,15)=4.05$, $p>.06$, but differences between training sets on blocks 3 and 4, smallest *F*(1,15)=17.24, *p*<.005. Analysis of simple main effects conducted on the interaction between stimulus type and block revealed no effect of block on the elements, *F*<1, and effect of block on the compounds, *F*(3,45)=13.56, *p*<.001. There were no differences between elements and compounds on blocks 1 and 2, largest *F*(1,15)=1.55, *p*>.23, but there were differences between the elements and the compounds on blocks 3 and 4, smallest *F*(1,15)=4.96, *p*<.05.

Test. The results from the test are shown in Figure 22. Inspection of each of the panels reveals that irrespective of group (Positive or Negative patterning), rats showed less activity (i.e. more evidence of fear) during the compounds (AB and CD) than during the elements (B and D). Inspection of the figure also indicates that there are no consistent differences either between B and D or between AB and CD. ANOVA confirmed that there was a main effect of stimulus type, *F*(1, 30)=12.23, *p*<.005, an effect of block, *F*(8, 240)=1.99, *p*<.05, and no effect of group or training set, *F*s<1. There was an interaction between group, stimulus type, and block, *F*(8,240)=2.48, *p*<.05, but no other interactions largest *F*(1,30)=2.26, *p*>.14. A separate ANOVA conducted on the scores from group Negative revealed an effect of stimulus type, *F*(1,15)=10.13, *p*<.01, and no other main effects or interactions, largest *F*(1,15)=3.00, *p*>.10. A parallel ANOVA conducted on the scores from group Positive, revealed a marginally significant effect of training stimulus, *F*(1,15)=4.22, *p*=058, and an effect of block, *F*(8,120)=2.24, *p*<.05, but no effect of training set and no other significant interactions, largest *F*(1,15)=2.49, *p*>.13. The fact that the effect of training set is less marked in group Positive than in group Negative might merely reflect that compounds in group Positive had been paired with food whereas those in group Negative had not.



Positive patterning: B vs D

Negative patterning: B vs D

Positive patterning: AB vs CD

Negative patterning: AB vs CD

Figure 22. Experiment 7. Mean levels of activity when rats in group Positive and Negative patterning were placed in contexts B and D (upper two panels) and AB and CD (lower two panels).

3.4.4 Discussion

Experiment 7 was designed to assess predictions derived from two successful accounts for the results of Experiment 5: the mediated conditioning view and the modified configural account. According to the mediated conditioning account, revaluation of AB and CD should result in B eliciting less fear then D. The same prediction follows from the modified configural analysis. Both analyses are, therefore, undermined by the results of Experiment 7.

3.5. General Discussion

The purpose of Experiments 5–7 was to investigate whether a process of functional categorisation (the idea that stimuli that predict a common outcome will be grouped) still operates during patterning discriminations (A+, B+, AB-). This is an interesting theoretical question, because recent configural/connectionist models of learning that provide an account of functional categorisation and patterning discriminations (e.g., Honey & Ward-Robinson, 2002) do not predict that following patterning training, the elements will cohere into a common category (i.e., activate a common hidden configural or hidden unit). According to this form of account, following patterning training, presentation of the compound will activate a single hidden unit and each of the elements will activate different hidden-layer nodes. When one of the elements is presented and paired with shock (e.g., A), this will allow the hidden-layer unit activated by A to become linked to shock, but when B is presented at test it should activate its own hidden unit that is not directly linked to shock. However, Experiment 5 established that an acquired equivalence/distinctiveness effect is observed after patterning training and this effect is every bit as marked as the effect observed when the compound and its elements predict the same outcome. Experiments 6 and 7 revealed that the functional grouping effect evident after patterning training was not a product of (1) reciprocal links between the output and hidden layer links, (2) mediated conditioning (cf. Honey & Hall, 1989), or (3) mediated by a representation of the compound. The results of Experiment 5–7 thereby pose a serious problem for recent accounts of functional grouping using connectionist architectures and, more worrying generally, both standard accounts of associative learning (e.g., Wagner, 1981) and more recent connectionist accounts of conditioning (e.g., Pearce, 1994). The results of Experiments 5–7 therefore require an alternative analysis.

Existing connectionist models have hidden units within them that exhibit some of the properties of an "AND" gate, in the sense that these hidden units become disproportionately more active upon activation of two of their input units (e.g., A and B) than on one (A alone). Although models of this kind can explain complex instances of acquired equivalence/distinctiveness (e.g., Honey & Ward-Robinson, 2002) they provide no basis for the results of Experiments 5–7. One possible way in which the latter results might be explained is to postulate a system with two types of unit: standard AND-type units and XOR-type units. There are a number of ways of instantiating an XOR-type unit, but whichever method one chooses they have the same (desirable in this context) property: they become active upon presentation of a single input units (A alone), but when they receive greater (accumulated) activation than their internal threshold permits, they will become inactive (e.g., if they received input from two units, A and B). How the addition of this second type of unit can help to explain the results of Experiments 5–7 is shown in Figure 23.

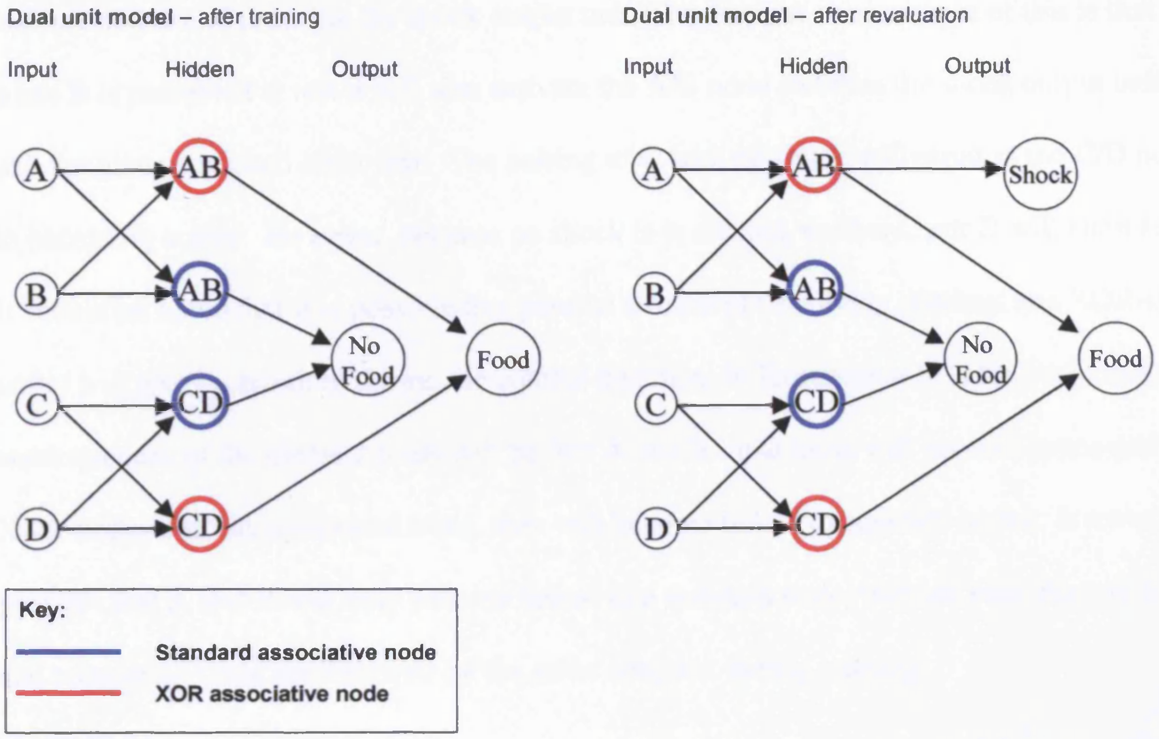


Figure 23. A dual unit model. A potential neural network model for functional grouping after patterning training with standard (sometimes AND-type) and XOR-type hidden-layer units.

Consider first the associative structures acquired during patterning training in Experiment 5. According to the model outlined in the left-hand panel of Figure 23, the input units activated by the AB compound become connected to a standard hidden-layer node (A+B node) and this node becomes connected to the no food output unit. Over the course of training (involving feedback from the output units, e.g., Honey & Ward-Robinson, 2002), the input units activated by A alone and B alone will become connected to an XOR-type node (A/B node) and this node becomes linked to the food output unit. This state of affairs, given certain assumptions about the strength of the links to the standard and XOR-type units will ensure that whenever A or B are presented the A/B node will become active and whenever A and B are presented the AB node will become active (see Section 5.3.2. in Chapter 5). The important consequence of the revaluation procedure is shown in the right-hand panel of Figure 23. Here, it is supposed that the pairing of A with shock will allow the A/B node to become active and linked to the shock output unit. The obvious consequence of this is that when B is presented at test it will also activate the A/B node and thus the shock output unit will become active and elicit fear. The pairing of C with no shock will result in the C/D node to becoming active. However, because no shock is presented, neither C nor D will elicit fear. It should be noted that it is possible that parallel structures (involving standard and XOR-type units) will also be acquired during the control treatment in Experiment 5, in the sense that the consequences of the element trials will be that A and B input units will become connected to XOR-nodes and, on compound trials, they will become linked to standard nodes. It remains possible that A and B will only become linked to a standard node, because both element trials and compound trials are followed by the same outcome during training.

The analysis outlined in the previous paragraph can also readily explain the results of Experiment 6, where revaluation of A and C resulted in greater fear to B and D than to AB and CD. This result reflects the fact that whereas the A/B and C/D XOR-type nodes are connected to shock, the A+B and C+D standard nodes are not (see Figure 24). In Experiment 7, of course, it was the compounds that were paired with shock and, again, there is no reason to anticipate fear to the elements (e.g., B and D). Although this is not a unique prediction of the model, it is at least consistent with the results.

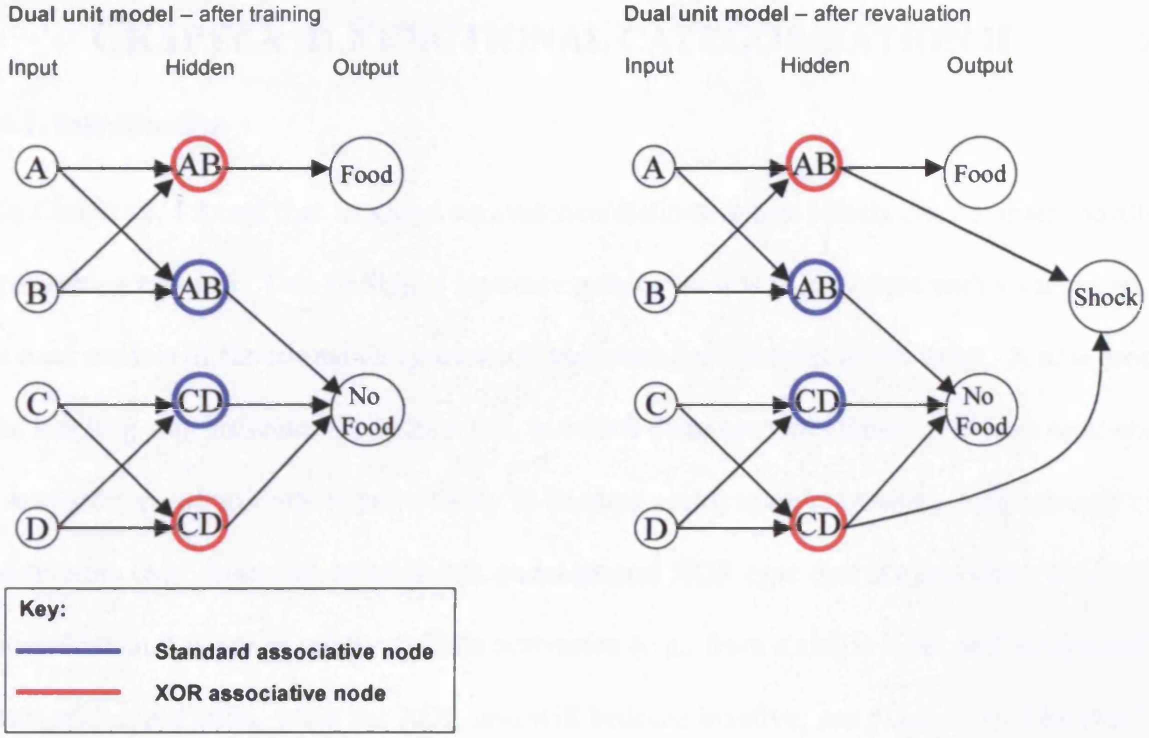


Figure 24. Application of dual unit model to Experiments 6 and 7.

The new dual unit model developed above provides an elegant account of the results of Experiments 5–7. This modification to extant connectionist models that are popular within the animal learning community is as simple as it is radical, but it should be acknowledged that aspects of it need to be specified and considered in greater detail. However, before doing so, and given the important nature of the departure that this model takes from existing models, it seems prudent to attempt to both replicate and extend the results upon which this new model is founded.

CHAPTER 4: FUNCTIONAL CATEGORISATION II

4.1. Introduction

In Chapter 3, I found that acquired equivalence/distinctiveness effects can be observed after patterning training. This finding is interesting, because it is inconsistent with a variety of extant models of functional categorisation and extended versions of the same. A new model of learning was presented in Section 3.5, in which there are two classes of hidden unit: one standard type of unit that is more likely to become active upon receiving a large amount of activation (e.g., from two input units), and a second XOR type that is more likely to become active when it receives relatively little activation (e.g., from a single input unit as opposed to two such input units, when the XOR unit will become inactive; see Figure 23). The two experiments reported in Chapter 4 attempt to replicate and extend the findings upon which this analysis was based.

4.2. Experiments 8 and 9

Experiment 8 represents an attempt to provide a within-subjects replication of the effects seen in Experiments 5 and 6: that is, to replicate the acquired equivalence/distinctiveness effects observed in Experiment 5 and to show that these differences in fear to B and D are not mirrored by differences in responding to the compounds, AB and CD. The design of Experiment 8 is summarised in Table 12. In this experiment there were two groups, Positive and Negative, that received the same training as Experiment 6 and 7. Following this training, rats received fear conditioning with either one of the contexts from the patterning half of training (e.g., A; see Table 12) or with one of the contexts from the control half of training. All rats then received test trials with B and D and with AB and CD. It was assumed, particularly on the basis of extrapolation from Experiments 5 and 6, that rats would show greater fear to B than to D, irrespective of whether

B was a context that had been part of a patterning discrimination and D was a context that was a part of a control treatment or vice versa. I also assumed that this difference in fear between B and D would not be reflected in behaviour to AB and CD.

Table 12: Design for Experiment 8.

Group	Training	Revaluation	Test
Positive	A-, B-, AB+ C+, D+, CD+	A→shock C→ No shock	B versus D AB versus CD
Negative	A+, B+, AB- C-, D-, CD-	A→shock C→ No shock	B versus D AB versus CD

The design used in Experiment 9 (see Table 13) is akin to Experiment 5. However, in Experiment 9, all rats received patterning training and then a context that was either involved in negative or positive patterning (e.g., A) was paired with shock prior to the test stage where fear to B, D, AB and CD was assessed. The predictions were the same as for Experiment 8.

Table 13: Design of Experiment 9.

Group	Training	Revaluation	Test
Patterning	A+, B+, AB- C-, D-, CD+	A→shock C→ No shock	B versus D AB versus CD

The final purpose of Experiments 8 and 9 was to show that the failure to observe fear to compound contexts following fear conditioning with the elements is not a consequence of the room in which fear conditioning was given. In order to do so, for half of the rats in each experiment, fear conditioning was given in the room in which the element contexts were housed and for the remainder it was given in the room in which the compound contexts were housed. I anticipated that irrespective of the room in which fear conditioning was given rats would be more likely to show fear to context B than to context D and that they would be more likely to show fear to context B than to context AB (or CD).

4.2.1. Method

Experiment 8. 32 naïve Lister hooded rats from the same source as Experiment 5 (*ad lib* weight *M*=382g; range=354–416g) and were housed and maintained at 80% of these *ad lib* weights. The apparatus and appetitive training procedure were identical to those described in Experiment 7. The revaluation trials were the same as in Experiment 5 in which A was paired with shock and C was not. Half of the rats in each of the groups, Positive and Negative patterning, received this revaluation treatment in the experimental room in which the elements had been presented during training. The remaining rats received this revaluation procedure, but in the experimental room in which the compounds had been presented. The chambers in this room had been altered so that contexts A and C were no longer accompanied by B and D, respectively. That is, the objects and odours were removed from the chambers. Other details of this experiment were identical to Experiment 7.

Experiment 9. 32 naïve Lister hooded rats from the same source as Experiment 5 (*ad lib* weight *M*=375g; range=341–398g) and were housed and maintained at 80% of these *ad lib* weights. The apparatus was identical to that described in Experiment 7. The appetitive training procedure was identical to that of Experiment 5, with the exception that all rats received both positive and negative patterning. The revaluation and test trials were identical to Experiment 8

4.3.1 Results of Experiment 8

Training. Figure 25 shows the mean rates of magazine entries in group Positive (top two panels) and Negative (bottom two panels). Inspection of the 1st and 3rd panels reveals that both groups of rats came to respond at a higher rate to the reinforced elements (A and B in group Negative and C and D in group positive) than to the non-reinforced elements and that there was little difference in the levels of responding between the groups. Turning to the 2nd and 4th panels, it is apparent that group Positive responded at higher rate to their reinforced compounds (AB and CD) than did group Negative respond to their nonreinforced compounds. ANOVA with group (Negative or Positive), stimulus type (elements or compounds), training set (A/B and AB versus C/D and CD) and block (1-4) revealed an effect of group, $F(1,30)=927.16, p<.001$, stimulus type, $F(1,30)=16.06, p<.001$, training set, $F(1,30)=74.60, p<.001$, and block, $F(3,90)=5.37, p<.05$. This analysis also revealed an interaction between group and stimulus type, $F(1,30)=47.7$, $p<.001$, stimulus type and training set, $F(1,30)=45.61, p<.001$, stimulus type and block, $F(3,90)=9.54, p<.001$ and training set and block, $F(3,90)=6.54, p<.001$. There were also three-way interactions between stimulus type, training set and block, $F(3,90)=12.10, p<.001$. There were no other significant interactions, largest $F(1,30)=4.42, p=.06$. In order to examine the source of the three-way interaction involving group, separate ANOVAs were conducted on the two groups.

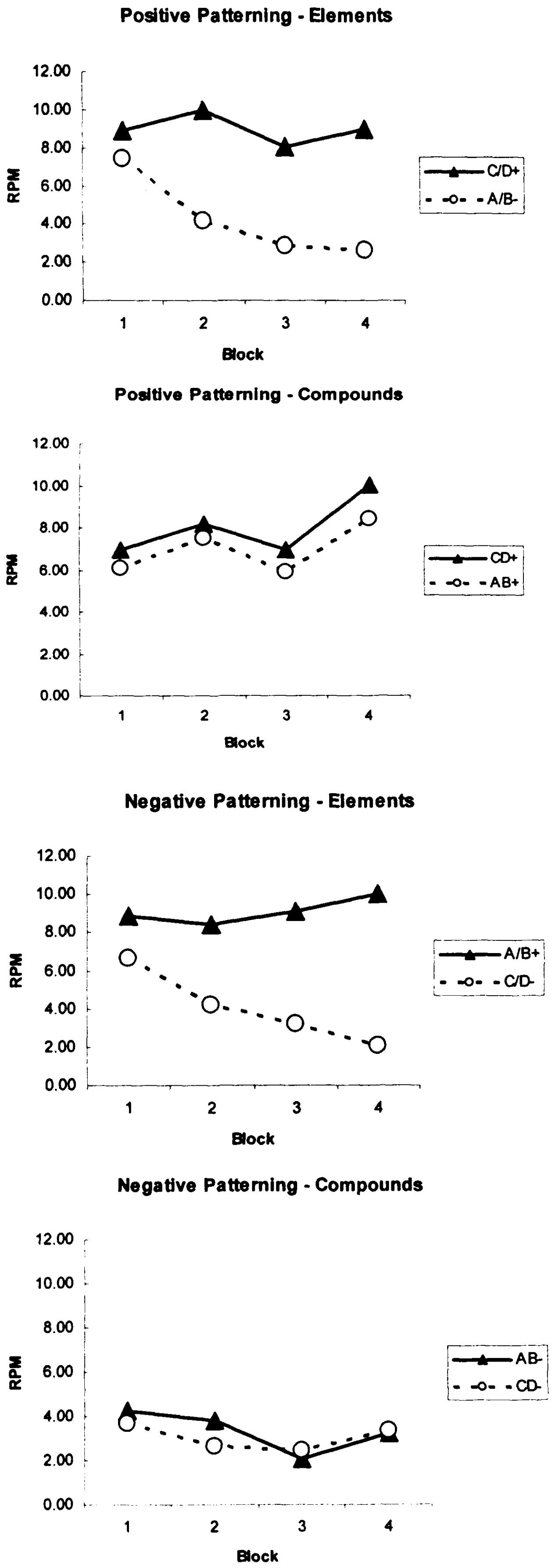
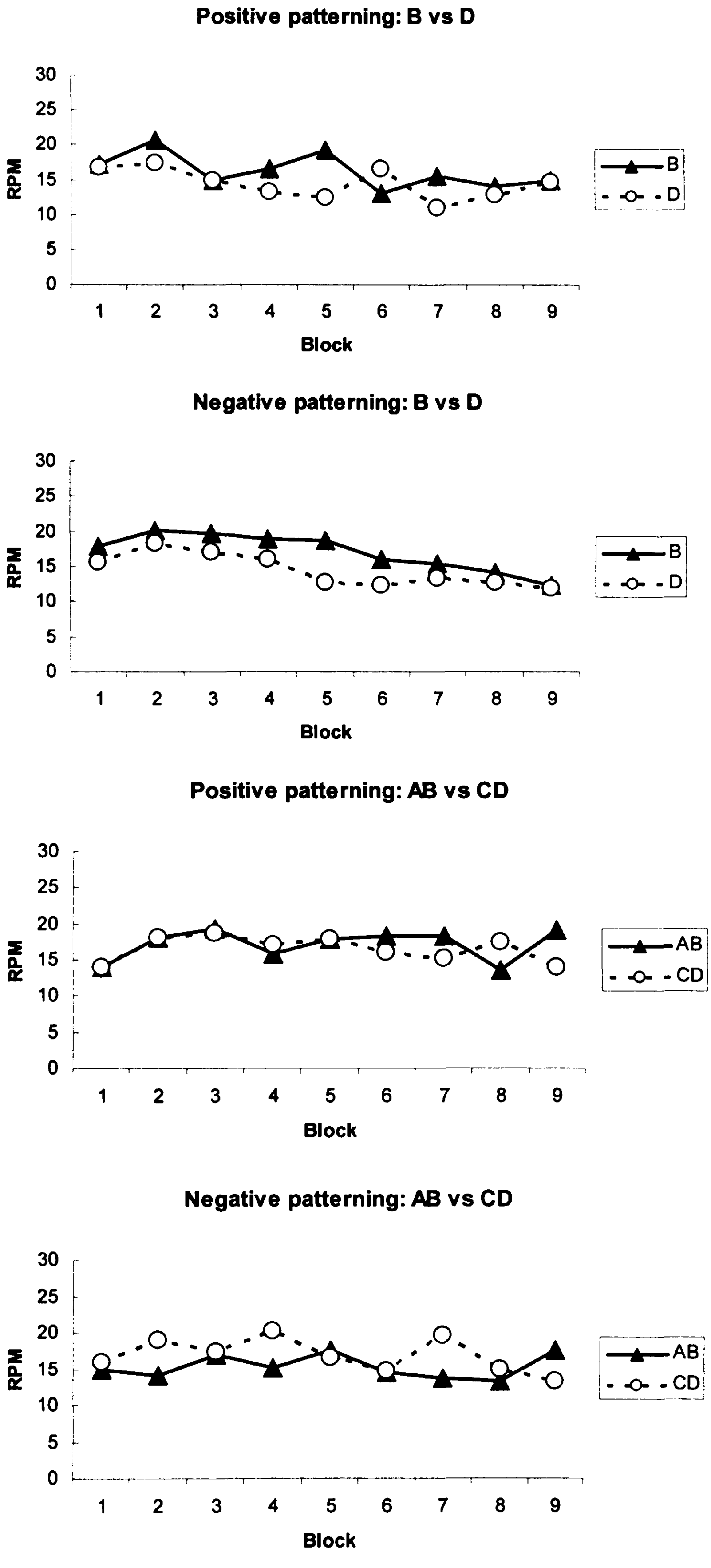


Figure 25. Experiment 8. Magazine entry scores (in responses per minute; RPM) during training for groups Positive (top two panels) and Negative (bottom two panels).

The separate ANOVA for group Negative revealed a main effect of stimulus type, $F(1,15)=46.90$, $p<.001$, training set, $F(1,15)=49.66$, $p<.001$. This analysis also revealed interactions between stimulus type and training set, $F(1,15)=36.79$, $p<.001$. There was a three-way interaction, $F(3,45)=10.06$, $p=.001$. The interaction between stimulus type and training set was analysed using analysis of simple main effects. This analysis revealed that there was a difference in responding to the elements between the two training sets, $F(1,15)=59.38$, $p<.001$, no difference between the compounds, $F<1$; there was also a difference between the elements and the compound for the patterning training set, $F(1,15)=63.47$, $p<001$, but no difference between the elements and the compound for the control set, $F(1,15)=3.72$, $p>.07$.

The separate ANOVA for group Positive revealed that there was no main effect of stimulus type, $F(1,15)=5.64$, $p<.05$, and that there were significant effects of training set, $F(1,15)=30.68$, $p<.001$ and block, $F(3,45)=3.02$, $p<.05$. This analysis also revealed interactions between stimulus type and training set, $F(1,15)=14.30$, $p<.005$, and an interaction between stimulus type and block, $F(3,45)=14.18$, $p<.001$, and no significant interaction between training set by block, $F(3,45)=3.77$, $p>.05$. There was a three-way interaction, $F(3,45)=5.16$, $p=.05$. The interaction between stimulus type and training set was analysed using analysis of simple main effects. This analysis revealed that there was a difference in responding to the elements between the two training sets, $F(1,15)=94.44$, $p<.001$, no difference between the compounds, $F(1,15)=1.38$, $p>.25$; there was also a difference between the elements and the compound for the patterning training set, $F(1,15)=27.84$, $p<001$, but no difference between the elements and the compound for the control set, $F(1,15)=1.92$, $p>.18$.


Positive patterning: B vs D

Negative patterning: B vs D

Positive patterning: AB vs CD

Negative patterning: AB vs CD

Figure 26. Experiment 8. Mean levels of test activity when rats in group Patterning and Control were placed in contexts B/D and AB/CD.

Test. The results from the test are shown in Figure 26. A preliminary analysis that included the room in which fear conditioning was conducted as a factor revealed no effect of this manipulation, *F*(1,30)=1.15, *p*>.30, and no interactions involving this factor, largest *F*(1,3)=1.93, *p*>.17. Accordingly, the results were combined across this factor for the purpose of presentation in Figure 26 and the analysis that follows. Inspection of each of the panels reveals that irrespective of group (Positive or Negative patterning), rats showed equal activity (i.e., equal evidence of fear) during B and D, and less activity during B and D than during the compounds (AB and CD). Also, the level of activity tended to decline across the test. ANOVA (with the same factors as the analysis of the training data) confirmed that there was an effect of stimulus type (elements or compound), *F*(1, 30)=7.93, *p*<.05, and block *F*(0,8)=3.72, *p*<.005, but no other significant effects or interactions, largest *F*(1,30)=1.99, *p*=.48.

4.3.2 Results of Experiment 9

Training. Figure 27 shows the mean rates of magazine entries in the elements (top panel) and compounds (bottom panel). Inspection of the top panels reveals that rats came to respond at a higher rate to the reinforced elements (A and B) than to the non-reinforced elements (C and D). Turning to the bottom panels, it is apparent that rats came to responds at a higher rate to the reinforced compound (CD) than to the non-reinforced compound (AB). ANOVA stimulus type (elements or compounds), patterning type (negative, A+B+AB-, versus positive, C-D-CD+) and block (1-4), revealed an effect of stimulus type, *F*(1,31)=9.37, *p*<.01, patterning type, *F*(1,31)=7.11, *p*<.05, and no effect of block, *F*<1. This analysis also revealed an interaction between stimulus type and patterning type, *F*(1,31)=23.94, *p*<.001, and also a three-way interaction, *F*(3,93)=7.22, *p*<.001. There were no other significant interactions, largest *F*(3,93)=1.59, *p*>.19. An analysis of simple main effects, conducted to explore the interaction between stimulus and patterning type, revealed that there is an effect of patterning type for elements and compound and an effect of stimulus type for both types of patterning, smallest $F(1,31)=4.42$, $p<.05$.

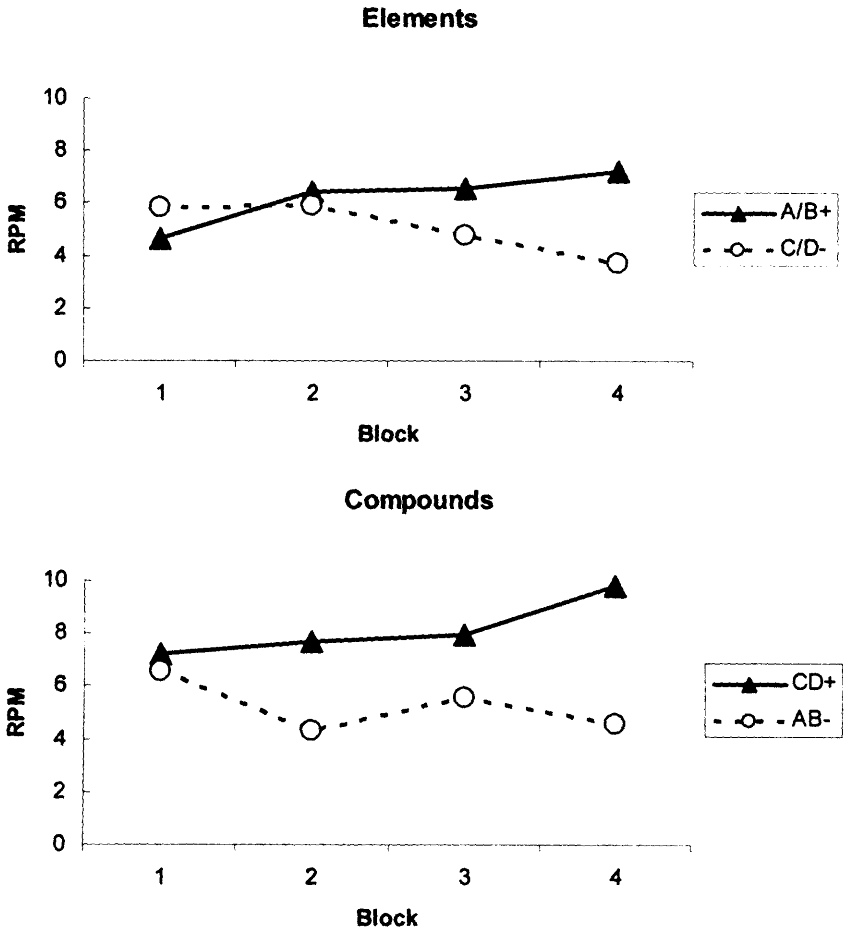


Figure 27. Experiment 9. Magazine entry scores (in responses per minute; RPM) during training.

Test. A preliminary analysis that included the room in which fear conditioning was conducted as a factor revealed no effect of this manipulation, $F<1$ and no interactions involving this factor, largest $F(8,248)=2.12$, $p>.05$. Accordingly, the results were collapsed across this factor for the purpose of presentation in Figure 28 and the analysis that follows. Inspection of the left-hand panel reveals that there is less activity (i.e., more evidence of fear) to B than to D.

Inspection of the right-hand panel reveals that both compounds elicited similar levels of fear. Comparison of the left- and right-hand panels indicates that B elicited more fear than AB and that D and CD elicited similar levels of fear. ANOVA confirmed that there was a main effect of stimulus type (element versus compound), $F(1, 31)=7.27, p<.05$, an effect of block, $F(8, 248)=3.95, p<.001$, and no main effect of revaluation set (B/AB versus D/CD), $F(1,31)=2.36$, $p>.13$. There was an interaction between stimulus type and revaluation set, $F(1,31)=8.75, p<.01$, but no other interactions, largest $F(8,248)=1.35, p>.21$. Analysis of simple main effects revealed that there was a difference between B and D, $F(1,31)=14.81, p<.005$, and that B differed from AB, $F(1,31)=8.39, p<.01$.

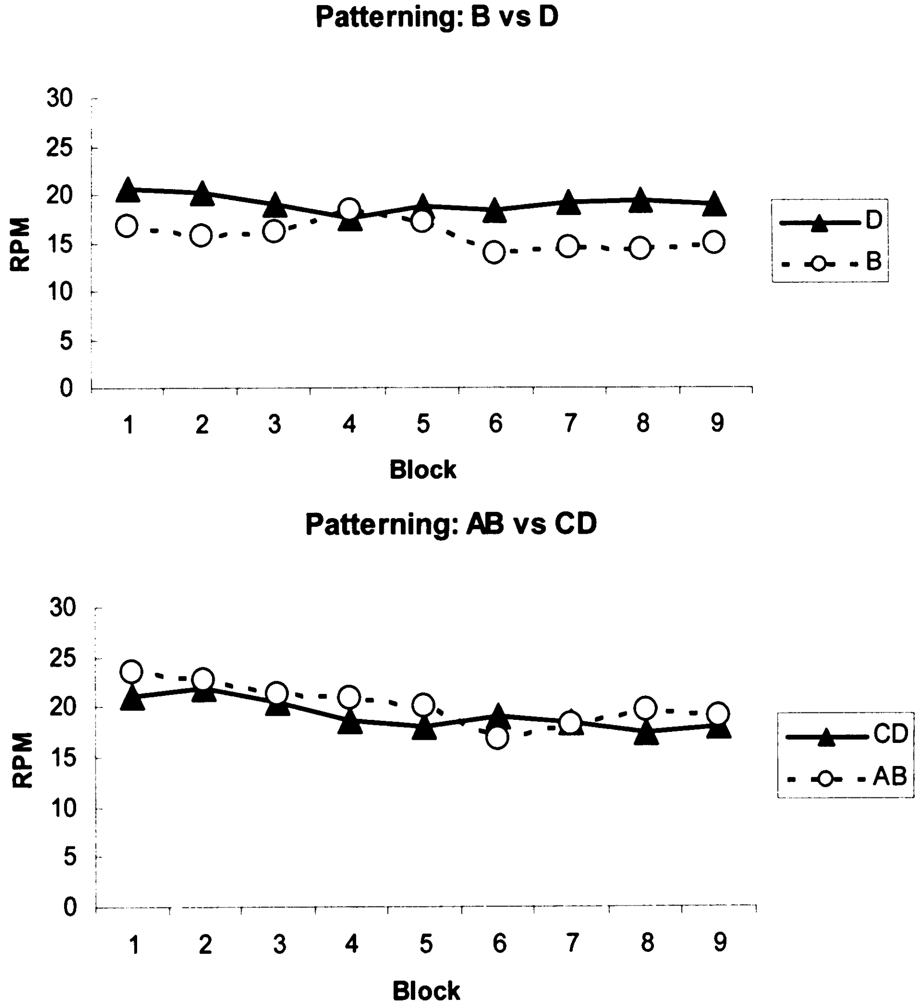


Figure 28. Experiment 9. Mean levels of activity when rats were placed in contexts B/D and AB/CD.

4.4. Discussion

Experiments 8 and 9 were conducted to confirm the results presented in Chapter 3. The theoretically important results from Chapter 3 were that acquired equivalence/distinctiveness effects can be observed after patterning training and that these effects were not reflected in fear to the compound contexts. Thus, after patterning training (A+, B+, AB-) rats showed greater generalisation between A and B than between A and AB. These results were confirmed in Experiment 9, which also controlled for the room in which revaluation occurred. These results represent a compelling confirmation of the results presented in Experiment 5 and important aspects of Experiment 6 and add further weight to the dual unit model. What is more, these results are inconsistent with a variety of other models. For example, they are inconsistent with the suggestion that generalisation between A and B is either mediated (1) by a representation of the presentation of the compound, or (2) by the fact that A and B activate a representation of a common outcome (e.g., food). If the effect was mediated by a representation of the compound then one would predict that compound AB would, at the very least, elicit greater fear than CD, and similar levels to B. This was not the case. Also, if the effect was mediated by the shared ability of A and B to activate a representation of the outcome that was associated with fear during revaluation (cf. Honey & Hall, 1989) then one should expect to see greater fear to the compound that was also paired with that outcome (CD in this instance). This was also not the case. Of course, one might argue that the effect of direct generalisation from A to AB (due to the presence of A) exactly matches the extent of mediated generalisation to CD. However, this analysis provides no account for the fact that B elicits greater fear than does AB (or CD).

If the results of Experiment 9 were entirely foreseeable on the basis of the interpretation of those presented in Chapter 3 (in particular those of Experiment 6) then those of Experiment 8 were in equal measure unforeseen. In Experiment 8 there was no evidence that the revaluation of A (following either patterning or a control treatment) resulted in greater fear to B than to D; there was, however, greater fear to both of these elements than to the compounds. The simplest way of reconciling the two sets of results was than in both Experiment 6 and 8 there is, counter to my original assumption, no acquired equivalence/distinctiveness effect. The sole difference between Experiments 6 and 8 and the successful Experiments 5 and 9 is that whereas in the former experiments the compound contexts were both paired with the same outcome during training, in the latter experiments they were not. Why this difference should have been critical is one question that will be discussed in Chapter 5.

While the results of Experiment 8 require further analysis, those of Experiments 5 and 9 provide clear support for the dual unit model and are inconsistent with other models. Any complete analysis of functional categorisation will need to explain the results of the experiments presented in Chapters 3 and 4.

CHAPTER 5: GENERAL DISCUSSION

5.1. Summary of rationale for the new research presented in this thesis

In my introductory chapter, two forms of categorisation were distinguished: perceptual and functional. I used the term "perceptual categorisation" to refer to the process of forming an integrated representation for patterns of stimulation that are often similar to one another (e.g., different views of a banana). In the context of perceptual categorisation, elemental and configural analyses were outlined and the need for further empirical work that examined the influence of similarity on integrating patterns of stimulation was developed (cf. Rescorla & Gillan, 1980). In particular, it was undoubtedly possible that the results of the experiments of the form reported by Rescorla and his colleagues might have limited generality and the theoretical analyses offered for them might not apply to other experimental preparations and species. I, therefore, conducted a series of four experiments with human participants (Experiments 1-4) to examine the role of similarity in associative learning. This set of experiments necessitated the development of novel appropriate stimuli and procedures.

I used the term "functional categorisation" to refer to the process of integrating otherwise arbitrarily related patterns of stimulation according to their uses or consequences. Again, elemental and configural analyses were presented and I identified a prediction that followed from elemental theorising (supplemented by mediated conditioning), but not configural analyses. In particular, it was argued that indices of functional categorisation (i.e., acquired equivalence/distinctiveness effects) should be observed following patterning training (e.g., A+B+AB-) according to elemental, but not configural accounts. Five experiments with rats were conducted to assess this prediction (Experiments 5-9). As it transpired, neither account provided a complete analysis for the new results.

5.2. Perceptual categorisation

5.2.1. A summary of the new results

Experiments 1–3 used a novel procedure in which stimuli were portrayed as alien bugs and the participants' task was to eliminate these bugs. The bugs were distinguished by their patterns. Bugs with feature A preceded similar patterns with feature B (AX→BX, AY→BY, AZ→BZ) and bugs with feature C preceded dissimilar patterns with feature C (CX→DY, CY→DZ, CZ→DX). Participants learnt, using feedback, that bugs with feature B (i.e., BX, BY, BZ) could be eliminated with one insecticide spray (e.g., red) and those with feature D (i.e., DX, DY, DZ) could be eliminated with a second spray (e.g., blue). Either spray was effective in eliminating bugs with features A and C. In Experiments 1–3, the participants learned which spray to use during patterns with features B and D and this learning was mirrored in their choice of which spray to use during patterns with features A and C. Critically, participants' spray choice for bugs with feature A came to match that used to eliminate bugs with feature B, but there was no such associative transfer between C and D (Experiments 2 and 3; but see Experiment 1). As noted above, both the elemental analysis of the effect of similarity on association formation offered by Rescorla and Gillan (1980) and the configural analysis introduced in Section 1.2.2. of Chapter 1 have no difficulty explaining the results of Experiments 2 and 3 (where the elements used to manipulate similarity, X, Y and Z, were irrelevant). However, in Experiment 4, where X, Y and Z were now relevant to the solution of a configural discrimination, transfer of 'correct' responding between the contiguous patterns was greater on dissimilar than similar trials.

5.2.2. Preliminary conclusion

The results of Experiments 1–4 are inconsistent with the configural analysis developed in Section 1.2.2. of Chapter 1 and, what is more, are also inconsistent with there being a general principle that similarity promotes association. However, the analysis provided by Rescorla and Gillan (1980), which gives no special role to similarity *per se* in the process of association, can be readily applied not only to the results of Experiments 2 and 3, but also to those of Experiment 4. This analysis simply supposes that the processing of a given pattern (BX) will be influenced by whether any of the elements of that pattern have recently been presented and that this short-term change in processing will produce a bias towards associating the pattern that precedes BX with B rather than B and X. It is, of course, possible that under other conditions similarity might have an influence on perceptual categorisation, but that is beyond the scope of this elemental analysis. In the section that follows, I describe a number of ways in which this elemental model could be subject to further experimental analysis.

5.2.3. Further research on perceptual categorisation

Comparative analysis. The results of Experiment 4 are counter-intuitive and it would be interesting to examine whether they can be observed in an analogous setting with pigeons. In particular, one could use the patterns developed in humans and a touch-sensitive screen with four locations in which the pigeons could respond (to map on to the four response options). If the results of Experiment 4 are replicable in pigeons then this would establish that the effect observed in humans is not restricted to organisms that can use explicit, verbally-based strategies.

Direct assessment of the role of short-term changes in stimulus processing. The results of Experiments 2–4 and the parallel experiments by Rescorla and Gillan (1980) assume that the influences of similarity reflect short-term changes in the processing of the elements of the second patterns (e.g., short-term habituation of X within the sequence AX→BX). However, as far as I am aware, there is no direct support for this assumption and it should be possible to assess its validity directly. In particular, the state of habituation of the elements within the second compound (e.g., BX and DY; see Table 14 for a sample of the trial types) could be influenced by interposing Y between the first and the second patterns. According to the elemental analysis of Experiments 2 and 3, this manipulation should reverse the pattern of results because it is now the second pattern within the dissimilar sequences (rather than the similar sequences) that will be more influenced by a process of short-term habituation. In short, the prediction under these circumstances is that transfer between the patterns in the dissimilar sequences should be greater than between those in the similar sequences. A parallel reversal in the pattern of results would be predicted in an analogous version of Experiment 4; now, there should be greater transfer between the patterns in the similar sequences than between the patterns in the dissimilar sequences.

Table 14: Design of proposed habituation experiment.

Condition	**Trial types**
Similar	AX→Y→BX→red
Dissimilar	CX→Y→DY→blue

Are the effects of similarity restricted to the first patterns? In Experiments 2–4 the influence of similarity was restricted to the first patterns in the sequences. One explanation for this pattern of results was presented in Section 2.4.3. of Chapter 2. Briefly, there it was supposed that the influence of differences in the processing of the second patterns had a greater effect on learning (about the adjacent patterns) than on performance to the second patterns. Of course, performance to the second patterns is being influenced (to an unknown extent and direction) by performance to the first patterns (which differs between conditions). In this context, it would be of some interest to compare the levels of performance to the second patterns as a function, for example, of whether or not they were preceded by the first patterns (see Table 15 for a sample of the trial types).

Table 15: Assessment of performance to second patterns.

Condition	**Training**	**Test**
Similar	AX→BX→red	BX?
Dissimilar	CX→DY→blue	DY?

5.3. Functional categorisation

5.3.1. A summary of the new results

The results of Experiments 5 and 9 indicate that the process of functional categorisation (grouping stimuli with the same outcome) occurs after patterning discriminations (e.g., A+B+AB-). Configural models (e.g., Honey & Ward-Robinson, 2002; Pearce, 1994) predict that acquired equivalence (one index of functional categorisation) should not occur following patterning training. In order to solve a patterning discrimination, A and B would come to activate different hidden layer units; when A is paired with shock, its hidden unit will be connected to shock, but when B then tested there is no reason to expect that it will elicit fear because it will activate its original hidden unit. In fact, Experiments 5 and 9 show that following patterning training there was generalisation between A and B. Furthermore, the results of Experiments 5 and 9, and those of Experiments 6 and 7, do not appear to be based on various alternative explanations (e.g., those based on reciprocal links between the hidden and output layers, mediated conditioning or mediation by a representation of the compound).

5.3.2 Preliminary conclusions and speculations

On the basis of the results presented in Chapter 3, I presented a development of connectionist/configural theorising that was able to account for these results. This dual unit model provides a simple account for the results of Experiments 5, 6, 7 and 9. Briefly, this model incorporates two different kinds of unit that come to code for either A or B, or A and B (see repeated Figure 23 below).

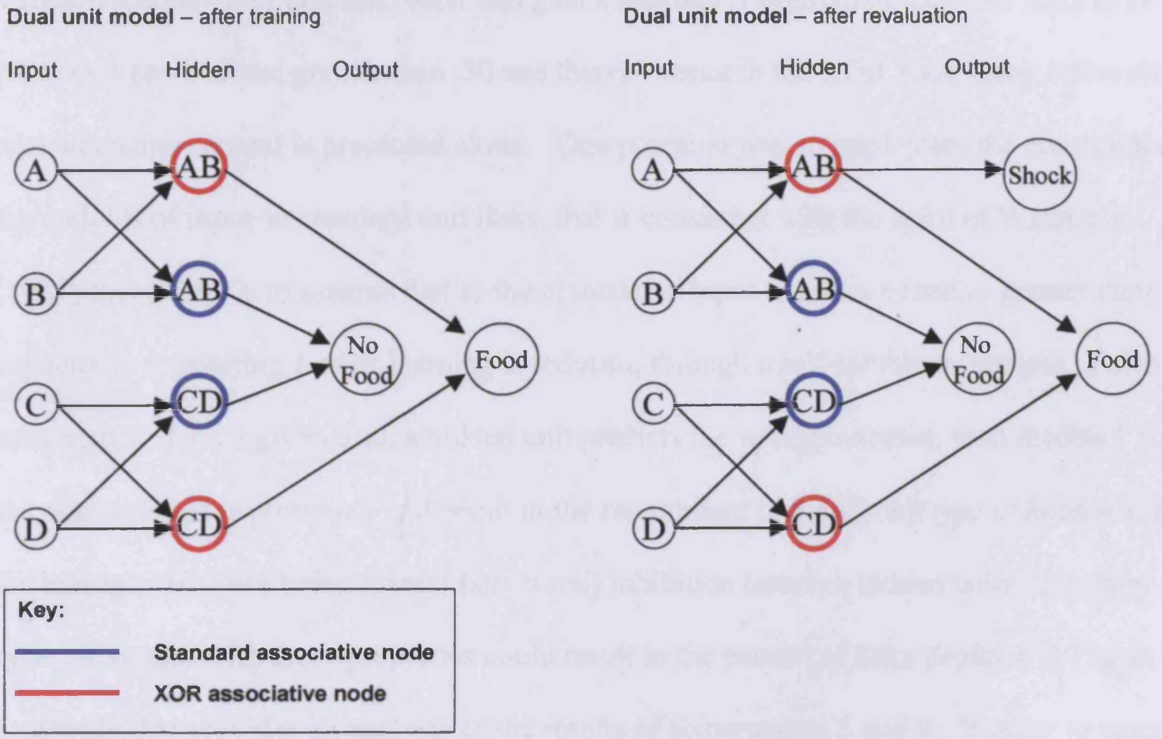


Figure 23. A dual unit model. A potential neural network model for functional grouping after patterning training in with standard (sometimes AND-type) and XOR-type hidden-layer units.

Two issues remain to be considered: how is the associative structure depicted in Figure 23 acquired and how could this model account for the results of Experiment 8. Briefly, in Experiment 8, rats received the following trial types: A+, B+, AB- and C-, D-, CD-. After this form of training, there was no evidence that pairing A with shock resulted in generalisation of fear to B. The sole difference between the design of Experiment 8 and that used in Experiments 5 and 9 is that in Experiment 8 both compounds were followed by the same outcome while in Experiments 5 and 9 they were not. How could this difference have altered the pattern of results? I will address these issues in turn.

In keeping with many other neural network accounts I assume that the link weights change according to Hebbian principles and, in this instance, I assume that there is an upper boundary to the total connection strengths to the standard units of 1, but no similar restriction for XOR units. For example, when a compound of two stimuli becomes linked to a given standard hidden-layer unit then each will gain a maximum strength of .50. The links to an XOR unit can become greater than .50 and thereby result in the XOR node being active on a trial when one context is presented alone. One potential way to implement the constraint on the weights of input-to-standard unit links, that is consistent with the spirit of Wagner's (1981) theorising, is to assume that as the associative input to nodes becomes greater their capacity to supporting further learning is reduced, through a self-inhibitory process. I also assume that if, on a given trial, a hidden unit predicts the wrong outcome, then feedback from the outcome that is presented will result in the recruitment of a different *type of hidden unit*; for example, by there being greater (structural) inhibition between hidden units of a given type. A system with these properties could result in the pattern of links depicted in Figure 23: a network that provides an analysis of the results of Experiments 5 and 9. In order to assess, in more detail, the predictions of that one could derive from this set of ideas, and the dual unit model more generally, it would be necessary to formally implement these properties within a computational model. Without such a model it is difficult to assess the necessity of the various assumptions and the manner in which they interact.

The results of Experiment 8 suggest that when the compounds of AB and CD are followed by the same outcome, then the associative structures depicted in Figure 23 are not acquired. I have considered a number of alternative associative structures (both allied to Figure 23 and other more conventional analyses) that might be generated when the compounds are followed by the same outcome. It is possible to generate a network within the framework summarised in Figure 23 that will no longer result in an acquired equivalence effect (i.e., a difference between B and D). However, such a network will not result in the difference in fear between the elements and compounds that was observed in Experiment 8. Clearly, further research is needed to pinpoint the reason that this small difference in procedure between Experiment 8 and 5 and 9 generated rather different results. It would be worth, as a first step, replicating the results of Experiment 8.

5.3.3. Further work on functional categorisation

Further assessment of the new model: in Experiment 5, the control and patterning groups produced similar a similar difference between the levels of fear generated to the element contexts (more to context B than D). There are two possible reasons for the fact that these two groups generate similar levels of acquired equivalence: first, the training given to the control group might have resulted in the same associative structures as that given to the patterning group (i.e., with A or B activating an XOR-type unit and A and B activating a standard unit). Second, the training given to the control group might have resulted in a different associative structure, within which A or B (and AB) are connected to a single hidden unit. One way to assess which of these two possibilities is the case would be to examine the levels of fear to the compounds. If the first possibility is the case, then there should be no difference in fear between the compounds AB and CD in the control group, whereas if the second possibility is the case, then there should be greater fear to AB than to CD.

I should also like to examine whether after patterning training, and following fear conditioning to A, the level of fear to A and B is the same. If it is, then this would represent to strong confirmation of the model, where it is assumed that the XOR unit alone has become connected to shock. If it is not, then I would have to assume that the fear conditioning stage also results in further associative structures in which, for example, the input unit for A alone becomes linked to another hidden unit that is itself linked to the shock output unit. Finally, following the form of patterning training given in Experiments 5 and 9, I should like to establish whether there is *any* mediation by a common outcome. This could be assessed by pairing AB with shock and examining the level of fear to B and D. It appears that there is relatively little generalisation between AB and B (following patterning training): the question of interest is whether there is greater generalisation between AB and D because they share a common outcome. This result would suggest that there are direct links between units in the output layer (specifically those activated by food and shock).

Assessment of generality: Humans. It would be of interest to assess whether the critical results described in Chapters 3 and 4 (i.e., Experiments 5 and 9) are also evident in other species and, in particular, in humans. One could use the general procedures outlined in Chapter 2 to do so. One potential problem with directly translating the procedure from rats to humans is the possibility that humans might use (non-associative) verbal strategies (see Shanks & Darby, 1998). To circumvent this possibility, participants could be instructed to respond rapidly and while carrying out some form of task that would reduce the likelihood of them using verbal strategies.

5.4. Concluding comments

The rationale for this thesis was to investigate the processes that underlie perceptual and functional categorisation, both of which were assumed to confer important advantages to animals. The results of my investigation of perceptual categorisation (in this instance meaning forming an integrated representation for different patterns of stimulation) revealed that an elemental analysis proved to be entirely adequate. In contrast, the results of my investigation of functional categorisation revealed that both existing elemental and configural analyses are inadequate. I presented an alternative configural model for these results. Clearly, an adequate complete model of these forms of categorisation will need to incorporate aspects of both types of model. One way in which the configural model that I have developed could be extended would be to allow elemental processes to be implemented within the input layer. In many senses it is surprising how far that one can extend elemental and configural connectionist models in order to explain new and counter-intuitive results. It is worth questioning, at this juncture, whether there is any empirical result that would render inadequate such models and require a more radical departure from them. If such a result was to be found, then the form that the departure would take is not obvious; apart, that is, from it becoming more closely allied to the complex and interactive nature of real neural systems.

REFERENCES

Allman, M., & Honey, R.C. (2005). Associative change in connectionist networks: An addendum. *Journal of Experimental Psychology: Animal Behavior Processes*, *31*, 363-367.

Allman, M., Ward-Robinson, J., & Honey, R.C. (2004). Associative change in the representations acquired during conditional discriminations: Further analysis of the nature of conditional learning. *Journal of Experimental Psychology: Animal Behavior Processes, 30,* 118-128

Allman, M.J. (2004). *Analysis of the nature of the representations that mediate acquired equivalence and distinctiveness* (unpublished doctoral thesis).

Brogden, W.J. (1939). Sensory pre-conditioning. *Journal of Experimental Psychology, 25*, 223-232.

Forbes, D.T., & Holland, P.C. (1980). Positive and negative patterning after CS pre-exposure in flavour aversion conditioning. *Animal Learning and Behaviour, 8.* 595-600

Gibson, E.J., & Walk, R.D. (1956). The effect of prolonged exposure to visually presented patterns on learning to discriminate them. *Journal of Comparative and Physiological Psychology, 49*, 239-242.

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

Guttman, N. & Kalish, H.I. (1956). Discriminability and stimulus generalisation. *Journal of Experimental Psychology, 51,* 79-88.

Hall, G. (1991). *Perceptual and associative learning.* Oxford, England: Clarendon.

Hall, G. (1994). Pavlovian conditioning: Laws of association. In N. J. Mackintosh (Ed.) *Handbook of perception and cognition. Vol 9: Animal learning and cognition* (pp. 13-43). San Diego, CA: Academic Press.

Hall, G. (2003). Learned changes in the sensitivity of stimulus representations: Associative and non-associative mechanisms. *Quarterly Journal of Experimental Psychology, 56B*, 43-55.

Hall, G., & Honey, R.C. (1989). Contextual effects in conditioning, latent inhibition, and habituation: Associative and retrieval functions of contextual cues. *Journal of Experimental Psychology: Animal Behavior Processes, 15*, 232-241.

Holland, P.C., & Ross, R.T. (1981). Within-compound associations in serial compound conditioning. *Journal of Experimental Psychology: Animal Behavior Processes, 7*, 228-241.

Holland, P.C. (1981). Acquisition of representation mediated conditioned food aversions. *Learning and Motivation, 12*, 1–18.

Honey, R.C. (2000). The Experimental Psychology Society Prize Lecture: Associative priming in Pavlovian conditioning. *Quarterly Journal of Experimental Psychology, 53B*, 1-23.

Honey, R.C., & Hall, G. (1989). Acquired equivalence and distinctiveness of cues. *Journal of Experimental Psychology: Animal Behavior Processes, 15*, 338-346.

Honey, R.C., & Hall, G. (1991). Acquired equivalence and distinctiveness of cues using a sensory preconditioning procedure. *Quarterly Journal of Experimental Psychology, 43B*, 121-135.

Honey, R.C., & Ward-Robinson, J. (2001). Transfer between contextual conditional discriminations: An examination of how stimulus conjunctions are represented. *Journal of Experimental Psychology: Animal Behavior Processes, 27*, 196-205.

Honey, R.C., & Ward-Robinson, J. (2002). Acquired equivalence and distinctiveness of cues: I. Exploring a neural network approach. *Journal of Experimental Psychology: Animal Behavior Processes, 28*, 378-387.

Honey, R.C., & Watt, A. (1998). Acquired relational equivalence: Implications for the nature of associative structures. *Journal of Experimental Psychology: Animal Behavior Processes, 24*, 325-334.

Honey, R.C., & Watt, A. (1999). Acquired relational equivalence between contexts and features. *Journal of Experimental Psychology: Animal Behavior Processes, 25*, 324-333.

Hull, C. (1943). *Principles of Behavior.* New York: Appleton-Century-Crofts.

James, W. (1890). *The Principles of Psychology.* London: McMillan and Co.

Kehoe, E.J. (1988). A layered network model of associative learning: Learning-to-learn and configuration. *Psychological Review, 95*, 411-433.

Mackintosh, N.J. (1974). *The Psychology of Animal Learning.* New York: Academic Press.

McLaren, I.P.L., Kaye, H., & Mackintosh, N. J. (1989). An associative theory of the representation of stimuli: Applications to perceptual learning and latent inhibition. In R.G.M. Morris (Ed.), *Parallel distributed processing: Implications for psychology and neurobiology* (pp. 102-120). Oxford, UK: Clarendon Press.

McLaren, I.P.L., & Mackintosh, N.J. (2002). Associative learning and elemental representation: II. Generalisation and discrimination. *Animal Learning & Behavior 30*, 177-200.

McLaren, I.P.L., & Mackintosh, N.J. (2000). An elemental model of associative learning: I. Latent inhibition and perceptual learning. *Animal Learning & Behavior, 28*, 211, 246.

McLaren, I.P.L., & Mackintosh, N.J. (2002) Associative learning and elemental representation: II. Generalisation and discrimination. *Animal Learning & Behavior, 30*,177-200.

Minsky, M., & Papert, S. (1969). *Perceptrons: an introduction to conceptual geometry.* Harvard: MIT Press.

Pavlov, I.P. (1927). *Conditioned Reflexes* (translated by G. V. Anrep). London: Oxford University Press.

Pearce, J.M. (1994). Similarity and discrimination: A selective review and a connectionist model. *Psychological Review, 101*, 587-607.

Redhead, E.S., & Pearce, J.M. (1995). Stimulus salience and negative patterning. *Quarterly Journal of Experimental Psychology, 48B*, 67-83

Rescorla, R.A., & Cunningham, C. (1978). Within-compound flavor associations. *Journal of Experimental Psychology: Animal Behavior Processes, 4*, 267-75.

Rescorla, R.A. (1974). Effect of inflation of the unconditioned stimulus value following conditioning. *Journal of Comparative Physiology and Psychology, 86*, 101-106.

Rescorla, R.A., & Wagner, A.R. (1972). *A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement.* In A. H. Black & W. A.

Prokasy (Eds), *Classical conditioning: II. Current research and theory* (pp. 64-99). New York: Appleton-Century-Crofts.

Rescorla, R.A. (1980). *Pavlovian second-order conditioning: Studies in associative learning.* Hillsdale, NJ: Erlbaum.

Rescorla RA. (1982) Effect of a stimulus intervening between CS and US in autoshaping. *Journal of Experimental Psychology: Animal Behaviour Process.* 1982 Apr;**8**(2):131–141

Rescorla, R.A., & Furrow, D.R. (1977). Stimulus similarity as a determinant of Pavlovian conditioning. *Journal of Experimental Psychology: Animal Behavior Processes, 3,* 203-215.

Rescorla, R.A., & Furrow, D.R. (1977). Stimulus similarity as a determinant of Pavlovian conditioning. *Journal of Experimental Psychology: Animal Behavior Processes, 3,* 203-215.

Rescorla, R.A., & Gillan, D.J. (1980). An analysis of the facilitative effect of similarity on second-order conditioning. *Journal of Experimental Psychology: Animal Behavior Processes, 6,* 339-351.

Shanks, D.R., & Darby, R.J. (1998). Feature- and rule-based generalisation in human associative learning. *Journal of Experimental Psychology: Animal Behavior Processes, 24,* 405-415.

Spence, K.W. (1936). The nature of discrimination learning in animals. *Psychological Review, 43,* 427-449.

Testa T.J., & Ternes, J.W. (1977). Specificity of conditioning mechanisms in the modification of food preferences. In L. M. Barker, M. R. Best, & M. Domjan (Eds), *Learning mechanisms in food selection* (pp. 229-253). Waco, TX: Baylor University Press.

Thomas, G.V., Robertson, D., & Liebermann, D.A., (1990). The effects of relative intensity of cue and marker on marked trace conditioning in pigeons. *Quarterly Journal of Experimental Psychology, 42B,* 267-287.

Wagner, A.R. (1981). SOP: A model of automatic memory processing in animal behavior. In N.E. Spears and R.R. Miller (Eds), *Information processing in animals: Memory mechanisms* (pp. 5-47). Hillsdale, NJ: Erlbaum.

Wagner, A.R. (2003). Context-sensitive elemental theory. *Quarterly Journal of Experimental Psychology*, *56B*, 7-29.

Ward-Robinson, J., & Hall, G. (1999). The role of mediated conditioning in acquired equivalence. *Quarterly Journal of Experimental Psychology, 52B*, 335-350.

Warren, H.C. (1921). *A history of the association psychology*. London: Constable and Company Ltd.

Williams, D.A., Mehta, R. Poworoznyk, M., Orihel, J.S., George, D.N., Pearce, J.M (2002) Acquisition of superexcitatory properties by and irrelevant background stimulus. *Journal of Experimental psychology: Animal behaviour Processes, 28, 284-297*

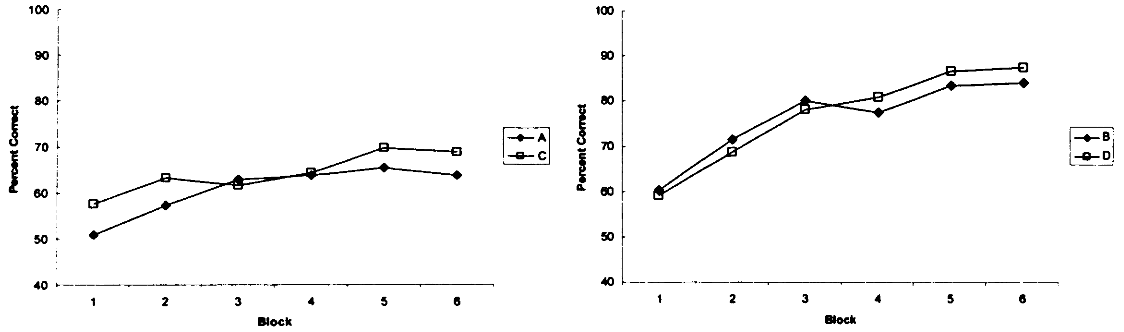
Williams, D.A., Dumont J.L., & Mehta, R (2004) Conditions favouring superconditioning of irrelevant conditioned stimuli. *Journal of Experimental psychology: Animal behaviour Processes, 30, 148-159*

Woodbury, C.B. (1943). The learning of stimulus patterns by dogs. *Journal of Comparative Psychology, 35,* 29-40.

Zentall, T.R., Steirn, J.N., Sherburne, L.M., & Urcuioli, P.J. (1991). Common coding in pigeons assessed through partial versus total reversals of many-to-one conditional and simple discriminations. *Journal of Experimental Psychology: Animal Behavior Processes*, *17*, *194-201*.

APPENDIX 1

Experiment 1 was run in an identical fashion to Experiment 2, with the exceptions that have already been noted. A summary of the results is shown in the figure below. Inspection of the right-hand panel indicates what statistical analysis confirms: participants scores increased as training progressed, $F(5,155)=18.28, p<.001$, and there was no difference between similar and dissimilar trials and no interaction with training block, *F*s<1. Inspection of the left-hand panel reveals that although participants scores increased across training, $F(5,155)=3.77, p<.005$, there was no difference between similar and dissimilar trials and no interaction between similarity and training block, *F*s<1.



Mean percentage correct for the first patterns (A and C; left panel) and second patterns (B and D; right panel).

same outcome while in Experiments 5 and 9 they were not. How could this difference have altered the pattern of results? I will address these issues in turn.

In keeping with many other neural network accounts I assume that the link weights change according to Hebbian principles and, in this instance, I assume that there is an upper boundary to the total connection strengths to the standard units of 1, but no similar restriction for XOR units. For example, when a compound of two stimuli becomes linked to a given standard hidden-layer unit then each will gain a maximum strength of .50. The links to an XOR unit can become greater than .50 and thereby result in the XOR node being active on a trial when one context is presented alone. One potential way to implement the constraint on the weights of input-to-standard unit links, that is consistent with the spirit of Wagner's (1981) theorising, is to assume that as the associative input to nodes becomes greater their capacity to supporting further learning is reduced, through a self-inhibitory process. I also assume that if, on a given trial, a hidden unit predicts the wrong outcome, then feedback from the outcome that is presented will result in the recruitment of a different *type of hidden unit*; for example, by there being greater (structural) inhibition between hidden units of a given type. A system with these properties could result in the pattern of links depicted in Figure 23: a network that provides an analysis of the results of Experiments 5 and 9. In order to assess, in more detail, the predictions of that one could derive from this set of ideas, and the dual unit model more generally, it would be necessary to formally implement these properties within a computational model. Without such a model it is difficult to assess the necessity of the various assumptions and the manner in which they interact.

The results of Experiment 8 suggest that when the compounds of AB and CD are followed by the same outcome, then the associative structures depicted in Figure 23 are not acquired. I have considered a number of alternative associative structures (both allied to Figure 23 and other more conventional analyses) that might be generated when the compounds are followed by the same outcome. It is possible to generate a network within the framework summarised in Figure 23 that will no longer result in an acquired equivalence effect (i.e., a difference between B and D). However, such a network will not result in the difference in fear between the elements and compounds that was observed in Experiment 8. Clearly, further research is needed to pinpoint the reason that this small difference in procedure between Experiment 8 and 5 and 9 generated rather different results. It would be worth, as a first step, replicating the results of Experiment 8.

5.3.3. Further work on functional categorisation

Further assessment of the new model: in Experiment 5, the control and patterning groups produced similar a similar difference between the levels of fear generated to the element contexts (more to context B than D). There are two possible reasons for the fact that these two groups generate similar levels of acquired equivalence: first, the training given to the control group might have resulted in the same associative structures as that given to the patterning group (i.e., with A or B activating an XOR-type unit and A and B activating a standard unit). Second, the training given to the control group might have resulted in a different associative structure, within which A or B (and AB) are connected to a single hidden unit. One way to assess which of these two possibilities is the case would be to examine the levels of fear to the compounds. If the first possibility is the case, then there should be no difference in fear between the compounds AB and CD in the control group, whereas if the second possibility is the case, then there should be greater fear to AB than to CD.

I should also like to examine whether after patterning training, and following fear conditioning to A, the level of fear to A and B is the same. If it is, then this would represent to strong confirmation of the model, where it is assumed that the XOR unit alone has become connected to shock. If it is not, then I would have to assume that the fear conditioning stage also results in further associative structures in which, for example, the input unit for A alone becomes linked to another hidden unit that is itself linked to the shock output unit. Finally, following the form of patterning training given in Experiments 5 and 9, I should like to establish whether there is *any* mediation by a common outcome. This could be assessed by pairing AB with shock and examining the level of fear to B and D. It appears that there is relatively little generalisation between AB and B (following patterning training): the question of interest is whether there is greater generalisation between AB and D because they share a common outcome. This result would suggest that there are direct links between units in the output layer (specifically those activated by food and shock).

Assessment of generality: Humans. It would be of interest to assess whether the critical results described in Chapters 3 and 4 (i.e., Experiments 5 and 9) are also evident in other species and, in particular, in humans. One could use the general procedures outlined in Chapter 2 to do so. One potential problem with directly translating the procedure from rats to humans is the possibility that humans might use (non-associative) verbal strategies (see Shanks & Darby, 1998). To circumvent this possibility, participants could be instructed to respond rapidly and while carrying out some form of task that would reduce the likelihood of them using verbal strategies.

5.4. Concluding comments

The rationale for this thesis was to investigate the processes that underlie perceptual and functional categorisation, both of which were assumed to confer important advantages to animals. The results of my investigation of perceptual categorisation (in this instance meaning forming an integrated representation for different patterns of stimulation) revealed that an elemental analysis proved to be entirely adequate. In contrast, the results of my investigation of functional categorisation revealed that both existing elemental and configural analyses are inadequate. I presented an alternative configural model for these results. Clearly, an adequate complete model of these forms of categorisation will need to incorporate aspects of both types of model. One way in which the configural model that I have developed could be extended would be to allow elemental processes to be implemented within the input layer. In many senses it is surprising how far that one can extend elemental and configural connectionist models in order to explain new and counter-intuitive results. It is worth questioning, at this juncture, whether there is any empirical result that would render inadequate such models and require a more radical departure from them. If such a result was to be found, then the form that the departure would take is not obvious; apart, that is, from it becoming more closely allied to the complex and interactive nature of real neural systems.

REFERENCES

Allman, M., & Honey, R.C. (2005). Associative change in connectionist networks: An addendum. *Journal of Experimental Psychology: Animal Behavior Processes, 31,* 363-367.

Allman, M., Ward-Robinson, J., & Honey, R.C. (2004). Associative change in the representations acquired during conditional discriminations: Further analysis of the nature of conditional learning. *Journal of Experimental Psychology: Animal Behavior Processes, 30,* 118-128

Allman, M.J. (2004). *Analysis of the nature of the representations that mediate acquired equivalence and distinctiveness* (unpublished doctoral thesis).

Brogden, W.J. (1939). Sensory pre-conditioning. *Journal of Experimental Psychology, 25,* 223-232.

Forbes, D.T., & Holland, P.C. (1980). Positive and negative patterning after CS pre-exposure in flavour aversion conditioning. *Animal Learning and Behaviour, 8.* 595-600

Gibson, E.J., & Walk, R.D. (1956). The effect of prolonged exposure to visually presented patterns on learning to discriminate them. *Journal of Comparative and Physiological Psychology, 49,* 239-242.

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

Guttman, N. & Kalish, H.I. (1956). Discriminability and stimulus generalisation. *Journal of Experimental Psychology, 51,* 79-88.

Hall, G. (1991). *Perceptual and associative learning.* Oxford, England: Clarendon.

Hall, G. (1994). Pavlovian conditioning: Laws of association. In N. J. Mackintosh (Ed.) *Handbook of perception and cognition. Vol 9: Animal learning and cognition* (pp. 13-43). San Diego, CA: Academic Press.

Hall, G. (2003). Learned changes in the sensitivity of stimulus representations: Associative and non-associative mechanisms. *Quarterly Journal of Experimental Psychology, 56B,* 43-55.

Hall, G., & Honey, R.C. (1989). Contextual effects in conditioning, latent inhibition, and habituation: Associative and retrieval functions of contextual cues. *Journal of Experimental Psychology: Animal Behavior Processes, 15*, 232-241.

Holland, P.C., & Ross, R.T. (1981). Within-compound associations in serial compound conditioning. *Journal of Experimental Psychology: Animal Behavior Processes, 7*, 228-241.

Holland, P.C. (1981). Acquisition of representation mediated conditioned food aversions. *Learning and Motivation, 12*, 1–18.

Honey, R.C. (2000). The Experimental Psychology Society Prize Lecture: Associative priming in Pavlovian conditioning. *Quarterly Journal of Experimental Psychology, 53B*, 1-23.

Honey, R.C., & Hall, G. (1989). Acquired equivalence and distinctiveness of cues. *Journal of Experimental Psychology: Animal Behavior Processes, 15*, 338-346.

Honey, R.C., & Hall, G. (1991). Acquired equivalence and distinctiveness of cues using a sensory preconditioning procedure. *Quarterly Journal of Experimental Psychology, 43B*, 121-135.

Honey, R.C., & Ward-Robinson, J. (2001). Transfer between contextual conditional discriminations: An examination of how stimulus conjunctions are represented. *Journal of Experimental Psychology: Animal Behavior Processes, 27*, 196-205.

Honey, R.C., & Ward-Robinson, J. (2002). Acquired equivalence and distinctiveness of cues: I. Exploring a neural network approach. *Journal of Experimental Psychology: Animal Behavior Processes, 28*, 378-387.

Honey, R.C., & Watt, A. (1998). Acquired relational equivalence: Implications for the nature of associative structures. *Journal of Experimental Psychology: Animal Behavior Processes, 24*, 325-334.

Honey, R.C., & Watt, A. (1999). Acquired relational equivalence between contexts and features. *Journal of Experimental Psychology: Animal Behavior Processes, 25*, 324-333.

Hull, C. (1943). *Principles of Behavior*. New York: Appleton-Century-Crofts.

James, W. (1890). *The Principles of Psychology*. London: McMillan and Co.

Kehoe, E.J. (1988). A layered network model of associative learning: Learning-to-learn and configuration. *Psychological Review, 95*, 411-433.

Mackintosh, N.J. (1974). *The Psychology of Animal Learning.* New York: Academic Press.

McLaren, I.P.L., Kaye, H., & Mackintosh, N. J. (1989). An associative theory of the representation of stimuli: Applications to perceptual learning and latent inhibition. In R.G.M. Morris (Ed.), *Parallel distributed processing: Implications for psychology and neurobiology* (pp. 102-120). Oxford, UK: Clarendon Press.

McLaren, I.P.L., & Mackintosh, N.J. (2002). Associative learning and elemental representation: II. Generalisation and discrimination. *Animal Learning & Behavior 30*, 177-200.

McLaren, I.P.L., & Mackintosh, N.J. (2000). An elemental model of associative learning: I. Latent inhibition and perceptual learning. *Animal Learning & Behavior, 28*, 211, 246.

McLaren, I.P.L., & Mackintosh, N.J. (2002) Associative learning and elemental representation: II. Generalisation and discrimination. *Animal Learning & Behavior, 30*,177-200.

Minsky, M., & Papert, S. (1969). *Perceptrons: an introduction to conceptual geometry.* Harvard: MIT Press.

Pavlov, I.P. (1927). *Conditioned Reflexes* (translated by G. V. Anrep). London: Oxford University Press.

Pearce, J.M. (1994). Similarity and discrimination: A selective review and a connectionist model. *Psychological Review, 101*, 587-607.

Redhead, E.S., & Pearce, J.M. (1995). Stimulus salience and negative patterning. *Quarterly Journal of Experimental Psychology, 48B*, 67-83

Rescorla, R.A., & Cunningham, C. (1978). Within-compound flavor associations. *Journal of Experimental Psychology: Animal Behavior Processes, 4*, 267-75.

Rescorla, R.A. (1974). Effect of inflation of the unconditioned stimulus value following conditioning. *Journal of Comparative Physiology and Psychology, 86*, 101-106.

Rescorla, R.A., & Wagner, A.R. (1972). *A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement.* In A. H. Black & W. A.

Prokasy (Eds), *Classical conditioning: II. Current research and theory* (pp. 64-99). New York: Appleton-Century-Crofts.

Rescorla, R.A. (1980). *Pavlovian second-order conditioning: Studies in associative learning.* Hillsdale, NJ: Erlbaum.

Rescorla RA. (1982) Effect of a stimulus intervening between CS and US in autoshaping. *Journal of Experimental Psychology: Animal Behaviour Process.* 1982 Apr;**8**(2):131–141

Rescorla, R.A., & Furrow, D.R. (1977). Stimulus similarity as a determinant of Pavlovian conditioning. *Journal of Experimental Psychology: Animal Behavior Processes, 3,* 203-215.

Rescorla, R.A., & Furrow, D.R. (1977). Stimulus similarity as a determinant of Pavlovian conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, *3*, 203-215.

Rescorla, R.A., & Gillan, D.J. (1980). An analysis of the facilitative effect of similarity on second-order conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, *6*, 339-351.

Shanks, D.R., & Darby, R.J. (1998). Feature- and rule-based generalisation in human associative learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *24*, 405-415.

Spence, K.W. (1936). The nature of discrimination learning in animals. *Psychological Review*, *43*, 427-449.

Testa T.J., & Ternes, J.W. (1977). Specificity of conditioning mechanisms in the modification of food preferences. In L. M. Barker, M. R. Best, & M. Domjan (Eds), *Learning mechanisms in food selection* (pp. 229-253). Waco, TX: Baylor University Press.

Thomas, G.V., Robertson, D., & Liebermann, D.A., (1990). The effects of relative intensity of cue and marker on marked trace conditioning in pigeons. *Quarterly Journal of Experimental Psychology*, *42B*, 267-287.

Wagner, A.R. (1981). SOP: A model of automatic memory processing in animal behavior. In N.E. Spears and R.R. Miller (Eds), *Information processing in animals: Memory mechanisms* (pp. 5-47). Hillsdale, NJ: Erlbaum.

Wagner, A.R. (2003). Context-sensitive elemental theory. *Quarterly Journal of Experimental Psychology, 56B*, 7-29.

Ward-Robinson, J., & Hall, G. (1999). The role of mediated conditioning in acquired equivalence. *Quarterly Journal of Experimental Psychology, 52B*, 335-350.

Warren, H.C. (1921). *A history of the association psychology*. London: Constable and Company Ltd.

Williams, D.A., Mehta, R. Poworoznyk, M., Orihel, J.S., George, D.N., Pearce, J.M (2002) Acquisition of superexcitatory properties by and irrelevant background stimulus. *Journal of Experimental psychology: Animal behaviour Processes, 28, 284-297*

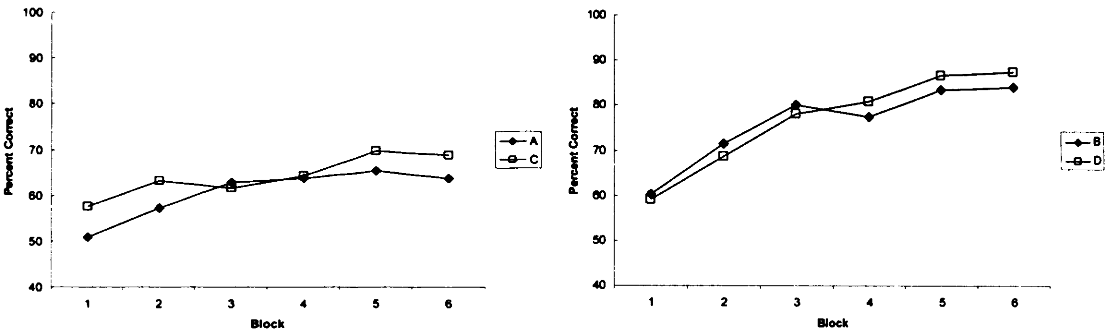
Williams, D.A., Dumont J.L., & Mehta, R (2004) Conditions favouring superconditioning of irrelevant conditioned stimuli. *Journal of Experimental psychology: Animal behaviour Processes, 30, 148-159*

Woodbury, C.B. (1943). The learning of stimulus patterns by dogs. *Journal of Comparative Psychology, 35,* 29-40.

Zentall, T.R., Steirn, J.N., Sherburne, L.M., & Urcuioli, P.J. (1991). Common coding in pigeons assessed through partial versus total reversals of many-to-one conditional and simple discriminations. *Journal of Experimental Psychology: Animal Behavior Processes*, *17, 194-201.*

APPENDIX 1

Experiment 1 was run in an identical fashion to Experiment 2, with the exceptions that have already been noted. A summary of the results is shown in the figure below. Inspection of the right-hand panel indicates what statistical analysis confirms: participants scores increased as training progressed, *F*(5,155)=18.28, *p*<.001, and there was no difference between similar and dissimilar trials and no interaction with training block, *F*s<1. Inspection of the left-hand panel reveals that although participants scores increased across training, *F*(5,155)=3.77, *p*<.005, there was no difference between similar and dissimilar trials and no interaction between similarity and training block, *F*s<1.



Mean percentage correct for the first patterns (A and C; left panel) and second patterns (B and D; right panel).

APPENDIX 2

Flow chart describing trial sequence and events for experiment 1 to 4.