

CARDIFF UNIVERSITY

Mechanisms Responsible for Cue-Competition Effects

Guillermo Octavio Ramos Esber

B.Sc. (Barcelona)

A thesis submitted for the degree of

Doctor of Philosophy

School of Psychology

Cardiff University, Wales

Cardiff CF10 3AT

UMI Number: U585021

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U585021

Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.



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

<u>CONTENTS</u>	<u>Page</u>
<u>DECLARATION</u>	v
<u>ACKNOWLEDGEMENTS</u>	vi
<u>SUMMARY</u>	viii
<u>CHAPTER 1. General Introduction</u>	1
1.1. Continuity and noncontinuity accounts of the fate of irrelevant stimuli in discrimination learning	5
1.1.1. Lashley-Krechevsky's (1929, 1932, 1938, 1942) noncontinuity theory of discrimination learning	6
1.1.2. Spence-Hull's (1936, 1943) continuity theory: Conditioning-Extinction theory	8
1.2. Evidence bearing on the early continuity-noncontinuity controversy	12
1.2.1. First experiments on the fate of irrelevant stimuli	12
1.2.1.1. Presolution-period experiments	12
1.2.1.2. Lashley's (1938) experiment on an added incidental discrimination	14
1.2.2. Evidence for differential attention in experiments on discrimination transfer	16

CONTENTS

Page

1.2.2.1. Changing response requirements in discrimination transfer: Lawrence's (1949) experiments on the acquired distinctiveness of cues	17
1.2.2.2. Changing the values of the discriminanda across transfer stages: intradimensional-extradimensional shift experiments (IDS-EDS)	19
1.3. Alternatives of Hull-Spence (1936, 1943) theory to attentional changes	21
1.3.1. Discriminanda transfer as the product of overt mediating responses	21
1.3.2. Hull's (1950, 1952) mechanism of asymptotic neutralisation by irrelevant stimuli as an account of discrimination-transfer effects	22
1.4. The impact of the discovery of blocking (Kamin, 1968) and relative validity (Wagner et al., 1968) on the definitive abandonment of Spence-Hull (1936, 1943) theory	24
1.5. Modern theories of associative learning	28
1.5.1. Modified continuity theories	28
1.5.1.1. Variations in US processing as the source of cue-competition phenomena	28
1.5.1.1.1. Kamin's hunch: surprise as a condition for learning	28
1.5.1.1.2. The Rescorla-Wagner (1972) model	29
1.5.1.2. The Comparator Hypothesis of R. R. Miller and associates (Miller & Schachtman, 1985; Miller & Matzel, 1988, Denniston et al., 2001): cue-competition effects as a retrieval/performance deficit	31
1.5.1.2.1. The original Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988)	31

CONTENTS

Page

1.5.1.2.2. The Extended Comparator Hypothesis (Denniston et al., 2001; Savastano et al., 2003)	37
1.5.2. Modified noncontinuity theories	42
1.5.2.1. The attentional theory of Sutherland & Mackintosh (1971)	42
1.5.2.2. The attentional theory of Mackintosh (1975)	44
1.6. Summary	49

CHAPTER 2. Analysis of the Comparator Hypothesis (Miller &

Schachtman, 1985; Miller & Matzel, 1988) 50

2.1. Introduction	50
2.2. Experiment 1	54
2.3. Experiment 2	64
2.4. General Discussion	75

CHAPTER 3. Further analysis of the Comparator Hypothesis

**(Miller & Schachtman, 1985; Miller & Matzel,
1988)** 81

3.1. Introduction	81
3.2. Experiment 3	87
3.3. Experiment 4	93
3.4. General Discussion	103

CONTENTS

Page

CHAPTER 4. Assessment of a novel technique for the study of

associability changes

111

4.1. Introduction

111

4.2. Experiment 5

116

4.3. Experiment 6

128

4.4. Experiment 7

139

4.5. General Discussion

151

CHAPTER 5. Analysis of the attentional theory of Mackintosh (1975)

163

5.1. Experiment 8

163

CHAPTER 6. General Discussion

175

6.1. Analysis of the Comparator Hypothesis (Miller & Matzel, 1988; Denniston et al., 2001)

175

6.2. Future directions in the analysis of the Comparator Hypothesis (Miller & Matzel, 1988; Denniston et al., 2001)

178

6.3. Analysis of the attentional theory of Mackintosh (1975)

182

6.4. Future directions in the analysis of Mackintosh's (1975) theory

183

6.5. General conclusions

192

ACKNOWLEDGEMENTS

I am deeply grateful to my supervisor Prof. John M. Pearce. Throughout the course of my Ph.D. his door always stayed open, and his guidance and enthusiasm were vital whenever the going got tough. During the composition of this thesis, his stylistic counselling and truly endless patience alleviated my painful progress through the syntactic thicket of the English language.

Special thanks go to Dr. Mark Haselgrove for his step-by-step introduction to the theoretical and practical aspects of research in the field of associative learning. Discussing with him the ins-and-outs of experimental designs soon became my favourite sport. I am also indebted to him for his constant first-line assistance whenever the computer in control of experimental events refused to obey my orders.

I would like to express my deepest thanks to all the support staff at Cardiff University, in particular Dennis Simmonds, Dennis Price and Rick Nash for promptly repairing laboratory equipment. I thank the IT group, especially David Griffiths, Hilmar Jay and Kevin Hotson, for all matters computer related, as well as Lorraine Woods for printing the colourful stimuli in Chapter 4. I gratefully acknowledge the support of Lynn Hill and Kerry Housler for providing animal care.

I would also like to thank all my colleagues in the behavioural neuroscience laboratory for countless stimulating discussions over the past four years. I am particularly grateful to Dr. David George for his scholarly advice on issues including theoretical analysis, literature reviewing and computer programming. I am also indebted to Prof. Rob Honey for drawing my attention to multiple interesting points.

Last but not least, I would like to thank my parents for the encouragement along the journey, as well as my partner Serena who, like a true Penelope, awaited my return amongst the living.

To The Beloved Memory Of
Martina Fernandez Chamorro “Nina”

ABSTRACT

The mechanisms responsible for cue competition were investigated. In Chapter 1, an overview of the literature that led to and originated from the discovery of cue competition effects (Kamin, 1969; Wagner, Logan, Haberlandt & Price, 1968) attested the diversity of theoretical accounts available to explain these phenomena. The subsequent empirical chapters focused on the predictions made by two rather distinct classes of theory: the Comparator Hypothesis (Miller & Matzel, 1988; Denniston, Savastano & Miller, 2001) and the attentional theory of Mackintosh (1975). Throughout the thesis, their predictions were contrasted to those derived from Standard Associative Theory [e.g. Rescorla-Wagner (1972) model].

The experiments contained in Chapters 2 and 3 used a Pavlovian appetitive procedure with rats to examine a number of predictions made by the Comparator Hypothesis. In Chapter 2, Experiment 1 tested the prediction that a conditioned inhibitor should have no influence on the excitatory status of the CS in which presence it is trained. Experiment 2 examined whether single-phase blocking disappears with asymptotic training.

Further analysis of the Comparator Hypothesis was provided in the two experiments contained in Chapter 3. Experiments 3 and 4 assessed the prediction that adding a stimulus to a continuously trained CS should deteriorate conditioned responding to the latter.

The experiments in Chapters 4 and 5, which used an autoshaping procedure in pigeons, were concerned with the attentional theory of Mackintosh (1975). In Chapter 4, Experiments 5 and 6 tested a novel behavioural technique intended to measure associability changes. Evidence of associability changes was found when visual patterns, but not colours, were compared. Experiment 7 explored the locus —central or peripheral— of the mechanism responsible for these changes.

Drawing from the results in Chapter 4, Experiment 8 (Chapter 5) examined whether associability changes can provide a complete account of the relative validity effect in pigeons.

Overall, the results challenge the accounts of cue competition advanced by both the Comparator Hypothesis (Miller & Matzel, 1988, Denniston et al., 2001) and the attentional theory of Mackintosh (1975). Without necessarily validating it, the results are mostly compatible with the analysis provided by the Rescorla-Wagner (1972) model.

“O this learning, what a thing it is!”

William Shakespeare

Chapter 1

1. General Introduction

One commonly held view within animal learning theory is that associative learning is the process that allows an animal to represent the causal structure of its environment (Tolman & Brunswick, 1935; Dickinson, 1980; Mackintosh, 1983). As causal detectors, animals face the challenge of identifying those elements in the environment signalling relevant causal relations out of all the stimulation impinging on their receptors. It has been a long-held intuition (e.g. Lashley, 1929; Hull, 1950) that the success of such enterprise is proportional to the organism's ability to attenuate the background *noise* posed by irrelevant stimuli.

Instances of a reduction in the processing of irrelevant stimuli have indeed been found in the laboratory with the discovery of stimulus-selection or cue-competition phenomena such as overshadowing (Pavlov, 1927), blocking (Kamin, 1968) and relative validity (Wagner, Logan, Haberlandt & Price, 1968). The term overshadowing was originally used by Pavlov (1927) to denote the loss of control over behaviour that a —typically weak— conditioned stimulus (CS) undergoes when it is trained in compound with another cue. The standard designs of the blocking and relative validity effects are shown in Table 1.1. In a typical blocking experiment (e.g. Kamin, 1968, 1969), a continuously reinforced stimulus X is trained in compound with a cue A in two groups. In the blocking group, the presence of conditioning trials with A prior to compound training detracts from X's ability to evoke a conditioned response (CR) at test, as revealed by a comparison with an overshadowing, control group. This effect has been successfully replicated with a number of procedures and species (e.g. Mackintosh & Honig, 1970; Mackintosh & Turner, 1971; Shanks, 1985; Arcediano, Matute & Miller, 1997).

Table 1.1. Classical experimental designs of the blocking (Kamin, 1968) and relative-validity (Wagner et al., 1968) effects.

Effect	Group	Acquisition		Test
Blocking	Blocking	A+	AX+	X?
	Overshadowing control		AX+	X?
Relative validity	True discrimination	AX+ BX-	AX+ BX-	X?
	Pseudodiscrimination	AX+ BX-	AX- BX+	X?

Described by Kamin himself as an attention-like phenomenon, blocking appeared at the time it came to light to be most aptly couched in attentional terms. Yet, instead, Kamin opted for an interpretation founded on the notion of surprise as a necessary condition for learning. In the blocking group, his argument runs, the occurrence of the unconditioned stimulus (US) is well predicted by A at the beginning of compound training with AX. As a consequence, the US will no longer be able to engage the learning machinery required to establish an association between X and the US. Kamin's ground-breaking experiments revealed that little learning may accrue to a stimulus perfectly correlated with reward if it is rendered redundant by further cues.

A related phenomenon is the relative-validity effect, in which all subjects receive training with two compound stimuli, AX and BX, with the common element X being partially reinforced. In a true-discrimination group, X is accompanied by cue A on all reinforced trials, whereas B accompanies X on all nonreinforced trials. In a pseudodiscrimination group, the compounds AX and BX are themselves partially reinforced, and therefore A and B are no better correlated with the delivery of the US.

At test, X is found to evoke a weaker CR following training in the true discrimination than in the pseudodiscrimination despite receiving equivalent individual treatment in both groups. This result was originally taken to indicate that it is the *relative* predictive validity of a cue with regards to other CSs, rather than its individual conditioning parameters, that determines how effective a CS will be (Wagner et al., 1968; Wagner, 1969a, b). A number of experiments using different species in a variety of procedures have confirmed Wagner and associates' finding (Wasserman, 1974; Udell & Rescorla, 1979; Luongo, 1976; Shanks & Dickinson, 1987; Cole, Barnet & Miller, 1995; Cole, Denniston & Miller, 1996; Cole, Gunther & Miller, 1997; Baker, Murphy & Vallee-Tourangeau, 1996; Van Hamme & Wasserman, 1994).

As laboratory miniatures of real-life stimulus-selection processes, cue-competition phenomena have been granted a central place in the study of selective learning over the past four decades. It was in fact the invitation posed by them to account for the fate of irrelevant stimuli that motivated the ascent of modern theories of associative learning (e.g. Rescorla & Wagner, 1972; Mackintosh, 1975; Miller & Matzel, 1988). The mechanisms underlying cue competition propounded by these theories range from sheer competition for a limited amount of conditionable associative strength (Rescorla & Wagner, 1972), through variations in the distribution of attentional resources (Mackintosh, 1975; Pearce & Hall, 1980), to competition between activated memory traces at the moment of retrieval (Miller & Matzel, 1988; Denniston, Savastano & Miller, 2001). Notwithstanding the evidence collected in support of some of these processes (e.g. Mackintosh & Turner, 1971; Kruschke & Blair, 2000; Cole, Barnet & Miller, 1995; Balaz, Gutsin, Chacheiro & Miller, 1982), it still holds true that no systematic attempt has been made to elucidate which of these

mechanisms provides the most satisfactory account of cue-competition phenomena, or, should several coexist, what is their relative contribution in each case. To be sure, an answer to these questions can only come from the thorough empirical investigation of the various theoretical interpretations at hand.

It is the purpose of this thesis to initiate such endeavour by focusing on the theoretical accounts of blocking and relative validity provided by 1) the Comparator Hypothesis of Miller and collaborators (Miller & Schachtman, 1985; Miller & Matzel, 1988; Denniston et al., 2001) and 2) the attentional theory of Mackintosh (1975). To this end, a number of predictions made by these theories concerning the fate of the irrelevant cue in each of these phenomena are evaluated against the background of what some regard (Roitblatt, 1985; Hall, 1991) as Standard Associative Theory, i.e. the Rescorla-Wagner (1972) model (see also Wagner & Rescorla, 1972). In this chapter, I begin by retracing the theoretical treatment that irrelevant stimuli received by continuity and noncontinuity theories of discrimination learning prior to the discovery of stimulus selection phenomena, accompanied by a succinct review of the empirical data on which their assumptions were grounded. The advantage of a historical account is twofold. First, it enables the reader to appreciate the argumentative counterpoint that led to the discovery of the phenomena under study. Second, it justifies the demise of strict continuity theory as advanced by Spence and Hull (Spence, 1936, 1937; Hull, 1943, 1952), and therewith the quest for mechanisms that took account of one basic fact: that the extent to which a CS will condition does not only depend on its individual correlation with reinforcement, but also on that of other CSs present during training.

Following this historical account, I introduce in some detail the Rescorla-Wagner (1972) model, as well as the Comparator Hypothesis (Miller & Schachtman,

1985; Miller & Matzel, 1988; Denniston et al., 2001) and the attentional theory of Mackintosh (1975). For each of the last two accounts, the evidence concerning blocking and relative validity is then selectively reviewed. As will be noted, a critical analysis of the extant data suggests a number of deficiencies in our knowledge of the topic.

1.1 Continuity and noncontinuity accounts of the fate of irrelevant stimuli in discrimination learning

It has been suggested on more than one occasion (e.g. Spence, 1951; Estes, 1969), that any comprehensive theory of associative learning should specify the contents of learning, the necessary and sufficient conditions for learning to occur, and the mechanisms whereby learning is expressed in the form of a response. The question of the fate of irrelevant stimuli in discrimination learning bears heavily upon the quest for the necessary and sufficient conditions for learning. It was Pavlov's (1927) contention that the same conditions necessary and sufficient for simple conditioning to occur could be readily transferred to conditioning with multiple cues. The conditions that influence the course of simple Pavlovian conditioning have been amply investigated. It has been known for some time that parametric variations in stimulus-reinforcer contiguity, correlation, and intensity have a direct impact on the acquisition function of the particular CR under observation across a wide range of procedures (Perkins, 1953; Logan, 1954; Kamin, 1965; Ross & Ross, 1971; Coleman & Gormezano, 1971; Rescorla, 1968). Following Sutherland and Mackintosh (1971), these parameters will be hereafter referred to as *static variables*.

The Pavlovian view that all stimuli present in a discrimination simultaneously acquire as much conditioned excitation (or inhibition) as their individual static variables can foster was later developed in the West in the so-called continuity theories. One influential account of this sort is the Conditioning-extinction theory of Spence (e.g. 1936, 1937) and Hull (e.g. 1943, 1952), described shortly in detail. Pavlov's assumptions on discrimination learning, however, were not without its critics. Two of the most notable discordant voices on the noncontinuity side of the controversy that ensued were those of Lashley (e.g. 1929, 1942) and Krechevsky (e.g. 1932, 1938), to be dealt with next.

1.1.1. Lashley-Krechevsky's (1929, 1932, 1938, 1942) noncontinuity theory of discrimination learning

Although the continuity-noncontinuity controversy included issues as dissimilar as the gradual vs. one-trial associations formation in conditioning (e.g. Guthrie, 1935; Estes, 1950) and the absolute vs. relative encoding of the discriminanda, (see Mackintosh, 1965; Kimble, 1961), it is the postulation of attention as a necessary condition for learning by noncontinuity authors that concerns us here. Indeed, the common assumption of noncontinuity or two-stage theories is that animals must choose which stimuli to attend to before associative learning can take place. Lashley (1929, 1942) and Krechevsky (1932, 1938), in particular, rejected the notion that the processes underlying discrimination learning may be reduced to the independent progress of simple conditioning to all cues present. Lashley (1929, 1942) first reported that when rats are required to solve a simultaneous visual discrimination in the jumping stand they adopt a position habit or a spatial strategy, either choosing

always left or right or alternating between the two, before showing any sign of discriminative responding to the relevant stimuli (e.g. black v white) arranged by the experimenter (e.g. Lashley, 1929). This observation, later replicated by Krechevsky (1932), led the two authors to propose that animals can only attend to one particular stimulus dimension at a time, from which it logically follows that learning about one dimension totally interferes with learning about another. This assumption deeply influenced a number of theorists, which incorporated it in later attentional models of discrimination learning (e.g. Zeaman & House, 1963, Lovejoy, 1968).

On this account, when rats are presented with a simultaneous visual discrimination there is obviously nothing in the situation itself to inform them of the relevant features to its solution. On the first trial that the subjects make a reinforced choice there will be inevitably more than one stimulus followed by reinforcement (say, black card and left stand). According to Lashley (1929, 1942) and Krechevsky (1932, 1938), rats will then generate a hypothesis about which stimulus best predicts the delivery of reinforcement which is founded on the stimulus dimension with the highest salience (e.g. left v right), and will subsequently respond to it. While so doing they will be unable to attend to —and therefore learn about— any other dimension including that arranged by the experimenter to carry the solution of the problem. Responding to the most salient but uncorrelated dimension is, however, bound to cease as no consistent prediction can be derived from it. At that time a new hypothesis will be generated and the animals' attention will be correspondingly redirected towards a second stimulus dimension. If this dimension happens to carry the solution of the discrimination, then the likelihood of switching attention away from it will be small, since attention is posited to reside wherever reinforcement can be consistently predicted. Once the discrimination is solved, no interference is then expected from

such hypotheses as happened to be previously generated, and thus whatever was learned about those irrelevant stimuli will bear no weight on subsequent discriminative behaviour (Kimble, 1961).

As a reaction against the cognitive overtones of Lashley (1929, 1942) and Krechevsky (1932, 1938) noncontinuity account, Spence (1936) questioned the need for an attentional construct to explain the rat's behaviour in the jumping stand by means of an elegant, albeit incomplete, account of discrimination learning.

1.1.2. Spence-Hull's (1936, 1943) continuity theory: Conditioning-Extinction theory

In a series of influential articles, Spence (1936, 1937a, 1937b) proposed an extension of Pavlov's (1927) theory capable of accounting for stimulus-selection phenomena of the sort reported by Lashley (1929, 1942) and Krechevsky (1932, 1938) without recourse to an attentional construct. Spence's theory, later espoused by Hull (1943, 1952), is formulated in terms of stimulus-response associations (S-R) as opposed to the Pavlovian stimulus-reinforcer connections, but the particular elements of association have no bearing upon the processes it posits at the heart of discrimination learning. A number of assumptions are made in order to explain discrimination learning as emerging from the interaction of independently conditioned elements.

Spence (1936, 1937) and Hull (1943, 1952) follow Pavlov (1927) in the assumption that reinforcement leads to an increase of conditioned excitation, whereas nonreinforcement leads to an increase of conditioned inhibition. The rate at which

conditioned excitation develops is a function of the static variables scheduled for each particular stimulus. Conditioned excitation and inhibition may co-exist for every single stimulus as a consequence of a history of reinforced and nonreinforced presentations, and their values of opposite sign will sum algebraically to cancel one another out. No distinction is made between the experimenter-manipulated discriminanda (S+/S-) and those stimuli that are incidental to the discrimination (e.g. context cues) as far as acquisition is concerned. In addition, the excitatory and inhibitory values acquired by each feature conditioned in the environment will generalise to other features in proportion to the amount of common elements they share.

By assuming that the amount of conditioned excitation accrued on a reinforced trial exceeds the amount of conditioned inhibition developed on an equivalent nonreinforced trial, the theory can accommodate the abundant evidence that 50% partial reinforcement results in a fair amount of responding (e.g. Humphreys, 1939). Conditioned responding in such circumstances is assumed to be driven by the surplus of excitation left uncanceled by the relatively slow-growing inhibition. This prediction can be readily extended to the uncorrelated irrelevant stimuli involved in any discrimination. Consider for instance a simple discrimination of the form S+/S-, in which S+ accordingly becomes a conditioned excitor and S- an inhibitor. The features held in common by both discriminanda, together with those contextual features present all along the learning episode (i.e. all irrelevant stimuli X) in the situation will experience a partial-reinforcement schedule. In keeping with the assumption above, X will directly acquire some net measure of conditioned excitation that will enable it to contribute its share to the net associative value on each trial that results from the algebraic interaction of all present excitatory and inhibitory

tendencies. What will be then the contribution of irrelevant stimuli to discriminative performance?

The answer to this question is well illustrated by Spence (1936, 1937) and Hull's (1943, 1952) account of Lashley (1929, 1942) and Krechevsky's (1932, 1938) purported evidence of stimulus selection in the jumping stand. In their experiments, the position (left or right) dimension, of higher intrinsic salience than the visual discriminanda (black or white), fosters the adoption of a spatial strategy. The partial schedule of reinforcement incidental to the spatial dimension guarantees that it will acquire a sizeable amount of excitatory strength. A systematic choice of position, however, exposes the rat to the relevant contingences between the second, more highly correlated dimension (e.g. brightness) and reinforcement, and ensures the undisputed development of the discrimination even while the subject is still responding to the position dimension. Only when the S+'s associative strength surpasses that of the excitatory position cue may behaviour come under the control of the experimenter-arranged discriminanda.

In general, therefore, the effect of irrelevant stimuli in the acquisition of a discrimination is twofold. First, that of translating the summed associative value of all cues present on each trial upwards along the inhibition-excitation scale. Paradoxically, Hull (1952) points out that such contribution might benefit the emergence of an incipient discrimination by boosting the discriminanda's gradients across a hypothetical performance threshold. Second, as irrelevant stimuli acquire further net excitation they will tend to flatten the generalisation gradients around the discriminanda, thereby obscuring the real extent to which the discrimination is being mastered. With no further assumptions, the model seems doomed to predict that

discriminative performance should, if anything, deteriorate over training as irrelevant stimuli continue to gain excitatory tendencies.

In order for the theory to avoid such an unreasonable prediction, Hull (1943, 1952) postulated a neutralisation mechanism that keeps irrelevant stimuli in check. This assumption also enables the theory to account elegantly for the evidence suggesting the steepening of generalisation gradients around discriminanda over the course of discrimination training, which evidently overcomes the flattening effect that irrelevant stimuli may initially produce. Such a neutralisation mechanism avails itself of the notion of asymptotic acquisition of conditioned excitation and inhibition. During discrimination training, the argument runs, irrelevant stimuli initially acquire a net excitatory tendency thanks to a higher rate of acquisition for excitation than inhibition. With sufficient training, the faster-growing excitation will reach an asymptote beyond which inhibition alone will continue to grow. With the further assumption that excitatory and inhibitory asymptotes are equal in magnitude and opposite in sign, it becomes possible to predict that the inhibition accrued by a stimulus on nonreinforced trials will gradually overtake its excitation until it virtually cancels out any remaining surplus of excitatory strength. In Hull's own words, "the effective reaction potential under the control of the incidental stimuli will for most purposes gradually become relatively neutral and unimportant in the determining of overt action" (Hull, 1952, p. 67-8). He elaborates on the importance of such processes for the organism's survival as it serves the function of eliminating "unadaptive" responses resulting from "the indiscriminate action of the law of reinforcement" (Hull, 1943, p. 268).

Despite its apparent success, Hull's mechanism of neutralisation for uncorrelated irrelevant stimuli gave rise to the one critical inconsistency in

Conditioning-extinction theory that led to the discovery of the relative-validity effect, and which, along with blocking, brought about the account's demise. Discussion of this point will nonetheless be postponed until the major developments involving the continuity-noncontinuity controversy have been presented. Accordingly, a brief summary follows of the evidences that settled the debate between Spence (1936, 1937) and Hull's (1943, 1952) "neo-Pavlovian" theory and Lashley (1929, 1942) and Krechevsky's (1932, 1938) noncontinuity account of the fate of irrelevant stimuli.

1.2. Evidence bearing on the early continuity-noncontinuity controversy

1.2.1. First experiments on the fate of irrelevant stimuli

A number of experimental strategies were employed in the early days of the continuity-noncontinuity controversy in an attempt to understand the mechanisms underlying discrimination learning (Kimble, 1961; Mackintosh, 1965; Mackintosh, 1974). According to Mackintosh (1974, p. 572), one popular strategy on both sides of the controversy was the so-called "presolution reversal", introduced next.

1.2.1.1. Presolution-period experiments

In presolution experiments, the predictive value of the discriminanda (S+/S-) is reversed while the subject is still responding to position and no hint of the discrimination is yet apparent. Subsequent retardation in the acquisition of the discrimination's reversal is held as evidence that *some* measure of learning about the original discrimination, albeit silent, must have occurred while behaviour was under

the control of the spatial strategy. Although a few reports have found no such retardation, in keeping with noncontinuity theory, (e.g. Krechevsky, 1938 Group II; Bower and Trabasso, 1963; Bradley, 1961), overall these studies have revealed deleterious effect for the emergence of the reversed discrimination, including one by Krechevsky himself (Krechevsky, 1938, Group III; McCulloch & Pratt, 1934; Spence, 1945; Gatling, 1951). Occasional failure to observe retardation has been attributed to insufficient training prior to the reversal (Mackintosh, 1965).

Two additional problems for noncontinuity theory associated with the presolution period derive from the following observations. Firstly, it has been reported that the latency of responding to S+ declines relative to that of S- even before a preference for S+ has surfaced, suggesting the presence of a gradual though otherwise behaviourally silent learning process (e.g. Mahut, 1954). Secondly, adding support to the latter suggestion Sutherland and Mackintosh (1971, p. 96) have documented that on the very first trial a rat abandons its position habit in the jumping stand, it tends to do so to respond to S+ rather than S-. Evidences from presolution period studies are not alone in posing difficulties for Lashley (1929, 1942) and Krechevsky's (1932, 1938) strict version of noncontinuity theory. The hypothesis that any learning about a stimulus will inevitably be reflected in behaviour has been challenged by a number of experiments directly measuring discriminative performance on a superimposed, irrelevant discrimination. We turn now to the pioneer experiment of this kind (Lashley, 1938).

1.2.1.2. Lashley's (1938) experiment on an added incidental discrimination

In Lashley's (1938) experiment, animals received training on a simultaneous visual discrimination involving one dimension S1+ S2- (e.g. size) before a redundant discrimination involving a second dimension S3+ S4- (e.g. shape) was superimposed at a later stage. Evidence of solving the added discrimination, S3+ S4-, was assessed by subsequently presenting it by itself during a transfer test. Assuming that this initial treatment succeeds in pre-setting attention to the original discrimination, noncontinuity theory anticipates the complete absence of discriminative responding at test. Such was the outcome reported by Lashley (1938).

Before considering the implications of Lashley's (1938) results, the parallel may be noted between his procedure and that used in Kamin's (1968) blocking experiments, provided Lashley's discriminations are equated to Kamin's discrete cues. In both procedures pretraining with an event —be it a discrimination or a cue— is found to prevent learning about a second event superimposed on it. Lashley's (1938) results can indeed be readily explained away by suggesting that substantial blocking of S3+ in the added discrimination should leave S3+ and S4-'s associative strength close to floor. Indiscriminate responding to S3+ and S4- at test is then justified through the generalisation of excitation from the pretrained S1+. Lashley's and Kamin's procedures differ, however, in the critical comparisons to be drawn at test, which are determined in either case by each experimenter's set of intentions. Whereas Kamin's intention was to test whether pretraining with an overshadowing cue prevents a target stimulus from gaining *as much associative strength as* in the absence of such pretraining, Lashley could only find support for his theory in the complete absence of discriminative performance to the redundant discrimination. The

absence in Lashley's design of an equivalent, no-pretraining control group accords with the all-or-nothing spirit in which strict noncontinuity theory was couched.

In hindsight, it is only natural to expect that whether any measure of learning is indeed observed should depend, just as in blocking, on procedural details such as the lengths of the phases of training or the relative salience of the stimulus dimensions used. Rather unsurprisingly, several subsequent replications of Lashley's design managed to reveal *some* degree of learning about the incidental discrimination (Blum & Blum, 1949; Bitterman & Coate, 1950; Hughes & North, 1959).

As Mackintosh (1965) has pointed out, the mixed pattern of results that emerged from the aforementioned studies did just enough to falsify noncontinuity theory in Lashley and Krechevsky's extreme form without necessarily lending support to its continuity adversary. Indeed, as suggested by Wagner (1969a), the fundamental disagreement over the necessary and sufficient conditions for learning was left unresolved by these experiments. However, the barren search by noncontinuity researchers for the complete absence of learning about unattended cues was conducive to the more moderate assumption that identifies most modified noncontinuity theories; namely, that the response-evoking properties of irrelevant stimuli should fall below the level generated by their specific static variables (Mackintosh, 1965). Direct testing of this prediction, much at variance with Conditioning-extinction theory as far as preasymptotic learning is concerned, had to await the discovery of the blocking and relative-validity effects. Prior to Kamin and Wagner's decisive contribution to the matter, consideration was given to the legitimacy of the attentional process at the heart of the noncontinuity account of discrimination learning. The first rigorous attempt to address this issue was Lawrence's (1949, 1950) work on the acquired distinctiveness of cues.

1.2.2. Evidence for differential attention in experiments on discrimination transfer

As noted by Hall (1991), it was Miller and Dollard (Miller & Dollard, 1941; Miller, 1948; Dollard & Miller, 1950) who first stated the view according to which discrimination training endows stimuli not only with a certain amount of associative strength, but also with a degree of attention-getting power or distinctiveness consistent with their relevance. On this account, a decline of attention to irrelevant stimuli is not only assumed to influence the course of the discrimination in progress, but also to limit their prospect of guiding future discriminative behaviour. Conversely, relevant stimuli that come to command high measures of attention are seen as likely to receive further attention in subsequent discriminations.

Three experimental techniques have been devised to test this notion. All of them involve some sort of transfer task in which the amount of attention allocated to a particular dimension is measured by its readiness to govern behaviour in a subsequent discrimination (i.e. its associability). Differences in associability may be detected for the various stimulus dimensions present during a discrimination so long as any direct associative or response transfer to the new discrimination is neutralised (Hall, 1991). The three strategies comprise: 1) changing the response requirements between discriminations while keeping the stimuli constant, as in experiments on the acquired distinctiveness of cues (e.g. Lawrence, 1949; Mackintosh & Holgate, 1967), 2) assessing the effects of extensive training on discrimination reversal, as in the overtraining extinction effect (e.g. Reid, 1953; Pubols, 1956) and 3) changing the particular values within the dimensions used across two subsequent discriminations

while keeping the response requirements constant. The last of these strategies is represented by two procedures known as transfer along a continuum (e.g. Pavlov, 1927, p. 121; Lawrence, 1952) and intradimensional-extradimensional shift learning (IDS-EDS, e.g. Shepp & Eimas, 1964; Mackintosh & Little, 1969). The reliability of each of these procedures in demonstrating associability changes has been amply discussed elsewhere (e.g. Mackintosh, 1974; Hall, 1991). General agreement exists that experiments on transfer along a continuum and discrimination reversal present less of a challenge to any continuity theory as alternative mechanisms have been put forward which obviate an account in terms of associability changes. More resilient to dismissal have proved those experiments on the acquired distinctiveness of cues and on intradimensional-extradimensional shift learning. The impetus they gave to the development of formal attentional theories of discrimination learning justifies a deeper exposition of their rationales.

1.2.2.1. Changing response requirements in discrimination transfer: Lawrence's (1949) experiments on the acquired distinctiveness of cues

One of the first rigorous instances of the use of transfer tasks to reveal associability changes can be found in the seminal work of Lawrence (1949, 1950, 1952) on the acquired distinctiveness of cues. Lawrence's original designs are highly complex, and therefore, for the sake of clarity, only the basic rationale will be considered. In the first stage of training, experimental subjects received a simultaneous discrimination in which two stimuli, black and white, reliably signalled the presence and absence of reinforcement, whereas a third stimulus, intermediate floor texture, was uncorrelated with reinforcement. For the control subjects, the

relevance of the brightness and texture dimensions was reversed, so that reinforcement and its absence were perfectly correlated with smooth and rough floor textures, whereas the third, irrelevant stimulus was a grey shade of brightness. Following training in this stage, all subjects were then transferred to a successive, conditional discrimination in which either black or white were presented at any given time and the texture of the floor was always of an intermediate value. The location of reinforcement was signalled for all subjects by specific brightness-position conjunctions (e.g. black & left reinforced, white & right nonreinforced). Experimental subjects, for whom brightness had been relevant in Stage 1, exhibited better conditional use of that dimension in conjunction with position than did control subjects, who had received irrelevance training with the brightness dimension.

In Lawrence's (1949) Experiment 1, the particular associative strengths initially acquired during the acquisition of a discrimination were made irrelevant to the solution of the transfer discrimination by changing the response requirements across stages. Lawrence assumed that a simultaneous-to-successive discrimination transfer (or vice versa) should suffice to ensure no direct transfer of the responses acquired. Although this assumption has not gone uncontested (Siegel, 1967; Pullen and Turney, 1977), there is reason to believe that this type of effect —and at least some replications of it (Mackintosh & Holgate, 1967; Mumma & Warren, 1968)—cannot be explained entirely by appealing to any transfer of response tendencies (Sutherland & Mackintosh, 1971, p. 178-180). Therefore, insofar as the discrimination given in Stage 1 results in attentional changes of opposite sign to the dimensions involved, positive transfer is predicted provided that the solution of the new discrimination depends on the subject coming under the control of such stimuli as were previously relevant. Alternatively, using the information supplied by

previously irrelevant stimuli might prove rather exacting, and negative transfer should likewise ensue. Indeed, either attending to the relevant stimuli or ignoring the irrelevant ones (or both) have been identified as potential sources of transfer that may contribute to the effect, and as admitted by Lawrence (1949) himself, his design does not shed any light on the matter. Whatever the case, acquired distinctiveness has traditionally been placed outside the scope of Conditioning-extinction theory as developed by Spence (1936, 1937 a, b) and Hull (1943, 1952).

1.2.2.2. Changing the values of the discriminanda across transfer stages: intradimensional-extradimensional shift experiments (IDS-EDS)

Further support for the case of attentional changes in discrimination learning has been supplied by experiments on intradimensional-extradimensional shift (IDS-EDS; e.g. Shepp & Eimas, 1964, Mackintosh & Little, 1969, George & Pearce, 1999). The design used by Mackintosh and Little (1969) will serve to illustrate their rationale, in which pigeons were trained with compounds composed of lines on a coloured background. An experimental, IDS group received training in which stimuli from the colour dimension (red vs. green) reliably signalled the occurrence of reinforcement, whereas stimuli from the dimension line orientation (vertical vs. horizontal) indicated no change in the probability of reinforcement. For the control, EDS group, the relevance of the dimensions was exchanged. Both groups were then transferred to a common discrimination problem involving other values along the same dimensions. For all subjects, the transfer discrimination had colour as the relevant dimension (blue vs. yellow) and line orientation as the irrelevant dimension (45° to left or right of the vertical plane). Acquisition of the second discrimination

was reported to proceed more rapidly in the IDS group, for which relevance of the dimensions was preserved across stages, than in the EDS group, for which the relevance of dimensions was shifted. In this procedure, it should be remarked, the response requirements are kept constant across stages, and great care is taken in the selection of stimulus values to minimise direct generalisation of associative strengths to the new stimuli in Stage 2. Although not a robust effect in non-human animals (Hall, 1991, p. 193; Hall & Channell, 1985; Couvillon, Tennant & Bitterman, 1976; Tennant & Bitterman, 1973, 1975; Sutherland & Andelman, 1969), IDS-EDS experiments have been identified by Mackintosh (1974) as “perhaps the best evidence that transfer between discrimination problems may be based partly on increases in attention to relevant dimensions and decreases in attention to irrelevant dimensions”. The accumulated evidence provided by this and the aforementioned class of studies had a profound effect on the field, stimulating the development of formal attentional theories of discrimination learning (e.g. Restle, 1955; Lawrence, 1963; Zeaman and House, 1963; Lovejoy, 1965, Sutherland, 1959; Sutherland & Mackintosh, 1971; Mackintosh, 1975). At the dawn of the 1960s, for instance, one general textbook on conditioning and learning sustained that “the phenomenon of attention no longer seems one which can be omitted in the description of even the simplest behaviour” (Kimble, 1961). Before formally introducing one of the most influential of these theories, the attentional theory of Mackintosh (1975), it is worth considering how strict continuity theorists attempted to accommodate these findings.

1.3. Alternatives of Hull-Spence (1936, 1943) theory to attentional changes

1.3.1. Discrimination transfer as the product of overt mediating responses

One of the earliest reactions to the proposal that attention, as a central, internal factor in the nervous system, mediates discrimination learning was the attempt to reduce it to an observable variable. Indeed, by attention some continuity authors effectively came to mean receptor-adjustment acts (Spence, 1936, 1937, 1960), or observing responses (Wyckoff, 1952; Miller & Dollard, 1941). If it is accepted that stimuli more highly correlated with reinforcement may induce the subject to adjust its sensory organs so as to maximise their reception, then it is conceivable that other stimuli present in the environment may fail to impinge on the subject's sensorium and, thereby, to influence its behaviour. An important implication of this premise is that the subject may come under schedules of reinforcement that are not necessarily those arranged by the experimenter.

Such a possibility was indeed recognised by some noncontinuity theorists (Mackintosh, 1965), without necessarily accepting that all purported demonstrations of attentional changes are to be reduced to such gross peripheral mechanisms (e.g. Mackintosh & Little, 1969). To qualify as evidence of *true* attention, therefore, an experiment must meet the requirement that all stimuli under investigation be equally well perceived so as to secure that the contingencies the subject is exposed to are those specified by the experimenter. The extended use of simultaneous visual discriminations involving instrumental procedures well into the second half of the past century made such possibility extremely difficult to reject (Wagner, 1969a).

Later replications of discrimination-transfer effects employing Pavlovian conditioning procedures in which extreme care has been taken in the selection of appropriate stimuli have rendered this alternative less plausible.

1.3.2. Hull's (1950, 1952) mechanism of asymptotic neutralisation by irrelevant stimuli as an account of discrimination-transfer effects

The sharpening of generalisation gradients over the course of discrimination training was not the sole kind of evidence encouraging Hull to propose a mechanism for the neutralisation of irrelevant stimuli. In a brief footnote to his article of 1950, he noted in passing that the “the previous neutralisation of irrelevant stimuli may play an important role in the transfer of training in discrimination learning” (Hull, 1950, p. 307). He declined, however, to provide a detailed account of how such a mechanism could apply to the different instances of discrimination transfer known at the time. It is straightforward to envisage how learning about a second discrimination could benefit from the prior neutralisation of such irrelevant stimuli (e.g. contextual) as are available during the acquisition of a previous discrimination, even when both discriminations involve dimensions sharing little in common. The elimination of “unadaptive responses” (Hull, 1943) through the neutralisation of the constant irrelevant stimuli should indeed carry some general benefit with it by facilitating the acquisition of “task-appropriate response tendencies” (Rodgers & Thomas, 1982; Hall, 1991). In particular, Hull’s neutralisation process may be applied with some measure of success to experiments on transfer along a continuum and the overtraining reversal effect, as Hall (1991) argued in his review of alternative explanations to

attentional changes in discrimination transfer. In these cases, the treatment given in the experimental and control groups differ, albeit for different reasons, in the extent to which irrelevant stimuli have been allowed to neutralise.

With slightly more difficulty, a similar account can be brought to bear on Lawrence's (1949) experiments on the acquired distinctiveness of cues. Hull (1943, 1952) might indeed expect the experimental group to perform better upon transfer on the grounds that, unlike the control group, it encounters exactly the same stimulus (e.g. intermediate floor texture) as irrelevant across stages. Some measure of neutralisation is also expected, however, in the control group via the generalisation of both excitatory and inhibitory tendencies from the previously relevant stimuli (e.g. rough and smooth textures) towards the subsequently irrelevant ones (e.g. intermediate floor texture). It is not clear, therefore, whether the operation of neutralisation in this case will result in better performance by the experimental than the control group during the transfer discrimination. The application of Hull's neutralisation process to discrimination transfer, however, is most implausible in IDS-EDS learning experiments. It is true, on the one hand, that the irrelevant stimuli in the first stage of group IDS (e.g. vertical and horizontal line orientation) will transfer some measure of neutralisation to the same-dimension, irrelevant stimuli in the second stage (e.g. 45° to left or right of the vertical plane), but it is just as likely that no less neutralisation will ensue in the EDS group as a result of the independent generalisation of excitation and inhibition from the previously relevant stimuli (e.g. vertical and horizontal line orientation) to their equidistant, subsequently irrelevant stimuli dimension (e.g. 45° to left or right of the vertical plane). Even conceding that some IDS-EDS demonstrations may include some degree of unbalanced, within-

dimension generalisation across stages cannot spare the fact that Hull's approach is not exempt from some serious problems, as we shall now see.

1.4. The impact of the discovery of blocking (Kamin, 1968) and relative validity (Wagner et al., 1968) on the definitive abandonment of Spence-Hull (1936, 1943) theory

The problem with the neutralisation mechanism proposed by Hull (1950, 1952), it has been pointed out, emerges from the apparent neglect to distinguish between those situations in which a partially-reinforced cue is trained in isolation from those in which the same cue is embedded in a discrimination, as null terminal associative strength is predicted in either case (Wagner, 1969a; Mackintosh, 1974, 1983). There is little evidence, however, to suggest that responding to a single partially-reinforced cue may come to show any sign of abatement despite extensive training. Hull incidentally dismissed this apparent contradiction when addressing the partial-reinforcement extinction effect, on the grounds that simple conditioning experiments featuring partially-reinforced cues tend to use short intertrial intervals (ITI). In the course of training, he argued, the presentation of events leaves certain "perseverative stimulus traces" upon their termination in the manner of decaying representations. The stimulation associated with the occurrence of both reinforcement and its absence is assumed to leave traces which persist long enough to condition the particular conditioned response evoked in their presence. With a short ITI (e.g. 30s), the quick succession of reinforcement and nonreinforcement is supposed to bring the nonreinforcement traces into sufficient vicinity with the oncoming reinforcements as

to endow them with some second-order reinforcing properties. As a result, in the course of partial reinforcement “this will naturally oppose the extinction effects of the non-reinforcements interposed among the genuine reinforcements” (Hull, 1952; p.121).

Hull (1950, 1952) went on to suggest that discrimination training is not impervious to this phenomenon, as it shares the intermixed presentation of reinforced and nonreinforced trials with partial-reinforcement schedules. In fact, a sufficiently long ITI is required on his account if any discrimination is to be solved at all, in order that the nonreinforcement traces may have time to decay before the next presentation of reinforcement. With the aid of this subsidiary process, Hull temporarily armed himself to explain why partial reinforcement results in the neutralisation of partially-reinforced cues on some occasions but not others.

Testing this questionable set of assumptions could appear therefore simply a matter of comparing the behaviour elicited by a partially-reinforced cue after receiving identical individual training in two groups either by itself or as part of a discrimination. One might indeed be tempted to predict on Hull’s behalf that a common ITI should secure in such a case an equivalent level of responding across the groups; whether it will be low through neutralisation or sizeable through the mediation of stimulus traces should be left of course as an empirical question. This prediction would notwithstanding be unfounded, for it is still possible for Hull to justify greater neutralisation to a partially-reinforced cue when forming part of a discrimination simply on the grounds of differential overshadowing.

Hull’s (1943, 1952) account of overshadowing as the product of generalisation decrement follows naturally from his notion of “afferent neural interaction”. According to this view, stimuli presented in compound interfere with each other at the

perceptual level so that they do not come to activate the same representations as they do when presented separately. As stated by Hull himself, “the afferent impulses produced by the components of a dynamic stimulus compound are to some extent different when the component is acting “alone”, i.e., in a relatively static combination, than when it is acting with the remainder of the dynamic compound” (Hull, 1943, p. 377). Besides overshadowing, the concept of afferent neural interaction allows Hull to explain patterning phenomena in discrimination learning, as well as guard the elemental tenets of Conditioning-extinction theory (1936, 1943) against cross-experimental indications that feature-positive (AX+ X-) and feature-negative (X+ AX-) discriminations do not result in identical levels of responding to X.

It is in the light of these intricacies that one can best appreciate the contribution of Wagner, Kamin and associates (Wagner et al., 1968; Kamin, 1968, 1969): a proper assessment of Hull’s (1950, 1952) mechanism for the neutralisation of irrelevant stimuli requires matching the degree of overshadowing across treatments as well as the static variables specific to the cue under investigation. This is precisely what the relative-validity and blocking designs achieved.

In Wagner et al.’s (1968) studies on relative validity, X is a partially-reinforced stimulus always presented in compound with A and B for either a true discrimination (AX+ BX-) or a pseudodiscrimination (AX+/- BX+/-). There are, on the other hand, no *a priori* grounds to expect a net difference in the generalisation of excitation and inhibition from A and B to X across the two groups. Having matched X’s static variables, degree of overshadowing and net generalisation from A and B’s excitatory and inhibitory tendencies, there is little scope for Hull to predict that responding to X will be stronger after a pseudodiscrimination than a true discrimination.

An even more compelling challenge for conditioning-extinction theory comes from Kamin's (1968, 1969) experiments on blocking, in which the degree of overshadowing by A and schedule of reinforcement of the target cue X are equated. According to the Spence-Hull account, pretraining A in the blocking group should, if anything, enhance X's excitatory tendencies by increasing the generalised excitation coming from A toward X in that group. The opposite is in fact most commonly observed.

The 16 years of empirical research separating Hull's last publication in 1952 from the discovery of blocking and relative validity beg the question of why it took so long for the controversy over the necessary and sufficient conditions for learning to settle. According to Wagner (1969a), a pivotal circumstance that paved the way towards the level of experimental control required to reveal these conditions was the switch from instrumental to Pavlovian conditioning procedures. The joint evidence for stimulus-selection phenomena arising from discrimination-transfer and cue-competition experiments was in any event instrumental in the definitive abandonment of Hull and Spence's (1936, 1943) Conditioning-extinction theory. The revision imposed on the necessary and sufficient conditions for learning seemed initially to commend the attentional perspective common in modified noncontinuity accounts (e.g. Lovejoy, 1968; Zeaman & House, 1963; Sutherland & Mackintosh, 1971; Mackintosh, 1975). Over the following years, however, a number of theorists combined their efforts to advance mechanisms that took care of cue competition without leaving the continuity theoretical framework (e.g. Kamin, 1968; Revusky, 1971; Rescorla & Wagner, 1972; Wagner, 1981; Pearce, 1987, 1994; Miller & Matzel, 1988).

1.5. Modern theories of associative learning

1.5.1. Modified continuity theories

This section introduces two formal modified continuity theories which were specifically designed to provide a non-attentional account for cue-competition phenomena: the Rescorla-Wagner (1972) model and the Comparator Hypothesis of Ralph R. Miller and associates (Miller & Schachtman, 1985; Miller & Matzel, 1988; Denniston et al., 2001).

1.5.1.1. Variations in US processing as the source of cue-competition phenomena

1.5.1.1.1. Kamin's hunch: surprise as a condition for learning

An alternative to the noncontinuity interpretation of the “attention-like” processes suggested by cue-competition phenomena was advanced by Kamin (1968, 1969). The core statement of this account is that the occurrence of reinforcement in a conditioning situation needs to be surprising for it to be able to support learning about stimuli that precede its occurrence. Accordingly, Kamin attributes the deficient responding evoked to the added cue X in the blocking group to its being paired with a US whose occurrence is perfectly predicted by the presence of the formerly-conditioned A. In the hands of Rescorla and Wagner (1972), the account became what has been hailed by some (e.g. Hall, 1991) as the single most influential theory of associative learning.

1.5.1.1.2. The Rescorla-Wagner (1972) model

The Rescorla-Wagner (1972) model assumes that when two or more stimuli—including the context—are conditioned in compound, their elemental representations enter into separate associations with the US representation. During training, stimuli will compete for a common, limited supply of associative strength (λ) until it is exhausted by the total associative strength of the compound (ΣV). The amount of associative strength that a CS can gain on a compound trial (ΔV) is a function of the discrepancy between the available associative strength (λ) and the summed associative strengths (ΣV) of all concurrent stimuli. This relation is captured by the global error term ($\lambda - \Sigma V$) in the equation with which the model estimates the size of increments in associative strength, ΔV , for a CS following each trial:

$$\Delta V_X = \alpha_X \beta (\lambda - \Sigma V), \quad 1$$

In equation 1, α and β are learning-rate parameters determined by the properties of the CS and the US (e.g. salience), respectively, which remain fixed throughout the learning experience. The total amount of associative strength accumulated by each element during training is updated on a trial by trial basis by:

$$V_{X(n)} = V_{X(n-1)} + \Delta V_X \quad 2$$

By way of a global error term, the model renders conditioning sensitive to the presence of other cues in training. On nonreinforced trials, on the other hand, the value of λ is assumed to be null, and provided that the error term has negative sign,

the resulting value of ΔV will correspondingly be negative. In this manner, the theory is able to model losses in associative strength following extinction trials, as well as mathematically instantiate the phenomenon of conditioned inhibition as the acquisition of negative associative strength produced by nonreinforcement in the presence of excitatory CSs.

A detailed review of the Rescorla-Wagner (1972) model has been provided elsewhere (Miller, Barnet & Grahame, 1995). For the present, it should be sufficient to say that part of the reason for the success of this theory can be ascribed to the elegant way in which it handles cue-competition effects. These effects emerge naturally from the assumption that at asymptotic conditioning ($\lambda = \Sigma V$), the US will be jointly predicted by all stimuli of a compound. Further gains in associative strength will be thus denied to the stimuli involved (overshadowing), or to new stimuli added to the compound (blocking).

Similarly, the model is able to predict relative validity. To succeed, however, it needs to assume that the value of the parameter β is higher on reinforced trials than on non-reinforced trials, an idea that has received empirical support (Rescorla, 2002). Conceding such β differences, it follows from Equation 1 that the difference in responding to the common element X is the product of A and B's differential correlations with reinforcement across the two groups. During training, the three elements A, B and X accumulate different amounts of associative strengths in their race to asymptotic learning. In the true-discrimination group, stimulus A, better correlated with reinforcement than X, is anticipated to acquire most of the associative strength that X will lose on non-reinforced trials with B. By contrast, X will be in a better position to become a conditioned excitor in the pseudodiscrimination, where it

has the same probability of being associated with reinforcement as A or B, and is paired with reinforcement twice as either A or B.

As far as the learning-performance relationship goes, Rescorla and Wagner assume that the behaviour displayed by an animal as a result of Pavlovian conditioning is a monotonic function of the absolute associative strength that a stimulus acquires. Robust CR differences across treatments must be thus necessarily attributed to disparate amounts of associative strength, an assumption rejected by Miller and his associates in the Comparator Hypothesis (Miller & Matzel, 1988; Denniston et al., 2001).

1.5.1.2. The Comparator Hypothesis of R. R. Miller and associates (Miller & Schachtman, 1985; Miller & Matzel, 1988; Denniston et al., 2001): cue-competition effects as a retrieval/performance deficit

1.5.1.2.1. The original Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988)

Advocates of the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) reject that the mechanism responsible for cue competition necessarily operates at the encoding level, and suggests displacing its locus down in the information-processing chain to the time of association retrieval, or even further, to the point in which the retrieved information is translated into performance. As far as acquisition is concerned, the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) assumes a strict continuity position, allowing for conditioning to be influenced only by the static variables specific to each CS. Since the context

may gain associative strength just as any punctate CS, it is assumed that even simple conditioning actually entails some form of compound training, a notion that is compatible with the Rescorla-Wagner (1972) model. Acquisition by the elements of a compound is nonetheless posited to take an independent course from one another. Instead, these elements will interact (i.e. compete) at the time of retrieval of their respective associations with the US.

In its simplest form, the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) states that the magnitude of the CR to a target cue X at test will be directly related to its own associative strength, but inversely related to the associative strength of other cues, A, concomitantly trained. This relationship is the consequence of the following comparison process postulated in the theory. When X is presented by itself at test, it will evoke a primary representation of the US –by virtue of its association with it—, as well as a secondary representation of the US –by means of its within-compound association with A, its comparator stimulus. To the extent that the comparator-mediated US-representation is activated, it will detract the directly activated US-representation from being expressed in performance. The diagram in Figure 1 represents this comparison process.

Thus, one important respect in which the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) deviates from Conditioning-extinction theory (Spence, 1936; Hull, 1943) is in the proposal that all associative strength is positive. Extinction trials are indeed expected to reduce the strength of a given connexion in the manner envisaged by theories such as Bush and Mosteller's (1951), but the notion of inhibition in the sense of negative associative strength is rejected.

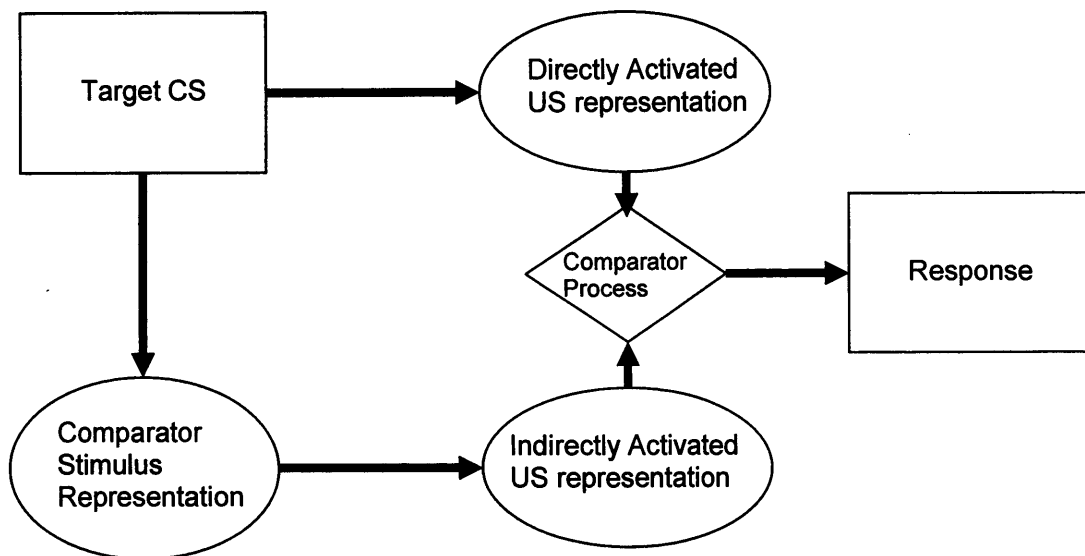


Figure 1. The comparator process.

Instead, the behaviour indicative of inhibition results from an extremely disruptive comparison process, which typically occurs when the associative strength of the target cue is much weaker than that of its comparator.

Cue competition naturally follows on this account from the presence of a more valid comparator leading to a more deleterious comparison process for the target cue X. In a blocking experiment, for example, the theory assumes that the X-US association will be equally strong in both groups, but not equally expressed in performance. This prediction follows because in the blocking group, A will serve as a more disruptive comparator for X as a consequence of its pretraining. Similarly, in a relative validity experiment, the superior correlation of A with the US in the true discrimination relative to that of X will lead to the characteristic response deficit in the presence of the latter cue. In the pseudodiscrimination, responding to X will be relatively high because, although its associative strength is the same as in the true discrimination, that of its comparator A is weaker.

From the dynamics of the comparison process, the prediction can be derived that any posttraining manipulation that weakens the activation of the mediated US-representation while leaving the directly activated US-representation intact will restore the response-evoking properties of X. This, according to the theory, can be achieved by weakening either the X-A within-compound association or the A-US association. Posttraining extinction of A has indeed been reported to abolish the response deficit to X in blocking (Blaisdell, Gunther & Miller, 1999; Arcediano, Escobar & Matute, 2001) and relative validity (Cole, Gunther & Miller, 1997).

Posttraining revaluation effects have also been found after delayed testing (i.e. spontaneous recovery) and “reminder” treatments (i.e. a few presentations with X or the US in isolation) in blocking (Miller, Jagielo & Spear, 1993; Batsell, 1997; Pineño, Urushihara & Miller, 2005; Balaz, Gutsin, Chacheiro & Miller, 1982; Schachtman, Gee, Kasproh & Miller, 1983) and relative validity (Cole, Barnet & Miller, 1995; Cole, Denniston & Miller, 1996). An obvious criticism against these last two techniques, which does not apply to that of extinguishing the comparator, is that they do not so readily follow from the principles of the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988). For instance, it is unclear, without further assumptions, why a few isolated presentations with the US following blocking treatment should selectively restore the memory of the X-US association rather than that of the A-US association, thereby eliminating blocking.

A further source of evidence —i.e., one not featuring posttraining manipulations— adding support to the Comparator Hypothesis comes from a series of experiments by Murphy, Baker and Fouquet (2001*a, b*). In Experiment 1 (Table 1.2), three groups of rats received a modification of the relative validity design using an appetitive conditioning technique. Three compounds, AX, BX and CX were present in

each treatment. In a true-discrimination group 1 (TD1), nonreinforced compounds AX and BX were intermixed with presentations of the reinforced compound CX. Another true-discrimination group (TD2) consisted of a reversal of this reinforcement schedule: compounds AX and BX were followed by reinforcement whereas CX was nonreinforced. Finally, in a pseudodiscrimination group (PD) all compounds were partially reinforced. In all three groups the compound CX was presented twice as many times as AX or BX in order for X to be partially reinforced on half the trials.

Table 1.2. Design of Experiment 1 by Murphy, Baker and Fouquet (2001a).

Group	Training			Test
True discrimination 1 (TD1)	AX-	BX-	2CX+	X?
True discrimination 2 (TD2)	AX+	BX+	2CX-	X?
Pseudodiscrimination (PD)	AX+/-	BX+/-	2CX+/-	X?

Following this training, responding to stimulus X was measured in extinction. Of special relevance was the comparison of the CR evoked by X in the two true-discrimination groups. Once conditioning has reached asymptote, the Rescorla-Wagner (1972) model predicts that stimulus C should have greater associative strength in the TD1 group than A or B in the TD2 group, and as a result C will be in a better position to overshadow X in the former group than A or B in the latter. This is because C is reinforced twice as many times as A or B. In contrast, Murphy et al. (2001a) derived from the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) the prediction that X should be equally overshadowed in the

TD1 and TD2 groups, because at asymptote C will be as strong a comparator for X in the TD1 group as A or B in the TD2 group. Both the Rescorla-Wagner (1972) model and the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) predict responding to X in the PD group to be higher than that in either the TD1 or TD2 groups (i.e. relative validity).

The results were consistent with the predictions they derived from the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988): responding to X was equivalent in the TD1 and TD2 groups, and weaker than that in PD. This is not the place for a detailed review of Murphy et al. (2001a), but some brief remarks in passing appear necessary. The simulations that these writers conducted of the Rescorla-Wagner (1972) model showed that, for all values of the parameter β on reinforced and nonreinforced trials, the difference in the strength of the CR in the presence of X predicted by this theory was always greater between the TD1 and TD2 groups, than between the TD2 and PD groups. This, as they recognise (p. 61), helps to ground their assessment of the Rescorla-Wagner (1972) model on what would otherwise be a questionable null result (TD1 = TD2). That such a null result comes to support the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) can, however, be called into question. One aspect in which the TD1 and TD2 groups are presumably not matched is in the comparator-target within-compound associations. If the C-X association is assumed to be stronger in the TD1 group than either the A-X or B-X associations in the TD2 group, as their contingencies would seem to suggest, then it follows that even at asymptote C should serve as a more forceful comparator for X in the TD2 group than A or B in the TD1 group. Thus, this prediction agrees with that derived from the Rescorla-Wagner (1972) model. Interestingly, the opposite prediction can also be obtained from the

Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988). To do so, however, one would need to make an assumption not yet contemplated in the theory; namely, that the deleterious effects of A and B may combine in the TD1 group to reduce responding to X further than C will by itself in the TD2 group.

Whatever the merit of the preceding analysis, it should be added that the occasional difficulties in deriving clear predictions from the original Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) in its original form can be ascribed to its lacking a mathematical formulation. An attempt to overcome this shortcoming has been made recently in the extended version of the theory (Denniston et al., 2001; Savastano, Arcediano et al., 2003).

1.5.1.2.2. The Extended Comparator Hypothesis (Denniston et al., 2001; Savastano et al., 2003)

Since the publication of the original Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) Miller and his colleagues have developed the theory further. The so-called Extended Comparator Hypothesis (e.g. Denniston et al., 2001; Savastano et al., 2003) comprises the same basic assumptions and structures as the original Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988). In addition to extending the range of circumstances in which comparator process may operate, this version of the theory has the advantage of being mathematically instantiated. The rules for acquisition and extinction are thus expressed, respectively, in the equations:

$$\Delta V_{ij} = \rho_1 (i \cdot j) (1.0 - V_{i,j})$$

3

$$\Delta V_{ij} = -\rho_2 \cdot i \cdot V_{ij},$$

4

where i and j are two arbitrary stimuli and ρ_1 and ρ_2 are positive learning rates parameters ranging from 0 to 1. As is the case with the original Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988), conditioned responding is determined by a comparison of the strength of (1) the target CS-US association with the product of the strengths of (2) the target CS-comparator association and (3) comparator-US association. Where the Extended Comparator Hypothesis (Denniston et al., 2001; Savastano et al., 2003) differs from the original theory is in the proposal that associations (2) and (3) are themselves modulated by a comparator process involving higher-order comparator stimuli. Since it is assumed that stimuli will compete with one another to become first- and second-order comparator stimuli, it is possible that the same stimulus will contribute to being both a first- and a second-order stimulus. However, Denniston et al. (2001, p. 86) note that, typically, an individual stimulus will serve either as the first- or as the second-order CS. Thus, taking X to represent the target stimulus, A to represent its first order comparator, and C to represent the context in the role of second-order comparator, the strength of the CR in the presence of X will be given by the equation:

$$R = V_{X,US} - (\rho_3 |V_{X,A} - \rho_3 V_{X,C} V_{C,A}| \cdot |V_{A,US} - \rho_3 V_{A,C} V_{C,US}| + \rho_3 |V_{X,C} - \rho_3 V_{X,A} V_{A,C}| \cdot |V_{C,US} - \rho_3 V_{C,A} V_{A,US}|),$$

5

in which R indicates the response potential of the target stimulus. Also in the above expression, ρ_3 is a parameter ranging from 0 to 1 which determines the degree to which each comparator term reduces the response potential R of the target stimulus X .

The | symbol indicates that only the modulus (i.e. positive value) of the expression enclosed should be taken.

A few recent experiments have found results compatible with the predictions of the Extended Comparator Hypothesis (Blaisdell & Miller, 2001; Urushihara, Wheeler, Pineño & Miller, 2005; Urcelay & Miller, 2006b). Of special interest here, given the topic of this thesis, are those reported by Blaisdell and Miller (2001). In two studies, rats received training on a true discrimination AX+ BX- using a lick-suppression procedure. After this phase of the experiment, which resulted in weak responding to the common cue X, subjects received a copious amount of extinction trials with A. As a result of this treatment, cue B was found to pass both retardation (Experiment 1) and summation tests (Experiment 2) for inhibition. These results exemplify the kind of second-order revaluation effect uniquely predicted by the Extended Comparator Hypothesis (Denniston et al., 2001; Savastano et al., 2003). According to this account, in a true discrimination A will serve as a comparator stimulus for X, which in turn will serve as a comparator for B. Because the capability of X to attenuate the expression of the weak B-US association is itself reduced by X's comparator A, at test B will only weakly elicit the behaviour typical of an inhibitor. By extinguishing A, the power of X as a comparator for B is effectively unleashed, turning the latter cue into an inhibitor sufficiently strong to pass retardation and summation tests.

The challenge posed to any researcher attempting to test the Extended Comparator Hypothesis (Denniston et al., 2001; Savastano et al., 2003) stems from the complex interactions it envisages between stimulus representations in discrimination learning. First-order posttraining revaluation effects are, as recognised above, of elusive replication in non-human animals (e.g. Holland, 1999), making the

investigation of second-order ones appear a rather unpromising strategy for theoretical assessment. For different reasons, more simplistic tests are equally bedevilled by the presence of extraneous variables. Consider, for instance, a situation where a cue X is conditioned in compound with two other CSs, A and B (ABX+). Compare this case with that in which the same cue is trained in compound with, say, A alone (AX+). According to the Rescorla-Wagner (1972) model, X should suffer greater overshadowing in the ABX+ group because A and B will more efficiently compete with X for the available amount of associative strength than A in the AX+ group. Without further assumptions, the Extended Comparator Hypothesis (Denniston et al., 2001; Savastano et al., 2003) is bound to make the opposite prediction; namely, that X will be overshadowed to a lesser degree in the ABX+ than the AX+ group. This is because A and B would compete with each other in the ABX+ group to serve as comparator stimuli for X, thereby mutually neutralising their deleterious influence on the expression of the X-US association. Despite the simplicity of this prediction, a closer inspection of the design reveals that interpreting a test with X will be hindered by nonassociative variables such as the unequal degree of generalisation decrement between training and testing conditions in the groups.

In any event, it should be recognised that experiments on posttraining revaluation, while providing the strongest support for the Comparator Hypothesis' (original or extended) account of cue competition, do not warrant acceptance of the particular associative structures postulated in the theory. Indeed, a number of accounts within the context of Standard Associative Theory have been devised to accommodate them (Van Hamme & Wasserman, 1994; Dickinson & Burke, 1996; Tassoni, 1995; Markman, 1989; Larkin, Aitken & Dickinson, 1998). Admittedly, to date these theories can only account for first-order posttraining revaluation. To my knowledge,

no theory other than the Extended Comparator Hypothesis (Denniston et al., 2001; Savastano et al., 2003) is able to predict second-order revaluation effects as those reported by Blaisdell and Miller (2001).

The ambiguity in the theoretical implications of posttraining revaluation effects, added to the difficulties attending their replication in non-human animals (e.g. Holland, 1999; Miller, Schachtman & Matzel, 1988), discouraged their use in this thesis as a method for assessing the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988; Denniston et al., 2001). In their place, some simple, novel predictions were derived in order to test some of the core tenets of the comparison mechanism proposed in the original version of the theory (Miller & Schachtman, 1985; Miller & Matzel, 1988). Care was taken that these predictions were in agreement with those made by the Extended Comparator Hypothesis (Denniston et al., 2001; Savastano et al., 2003). Thus, Chapter 2 compares 1) the role assigned by the theory to the inhibitor B in a true discrimination AX+ BX- with that assigned by the Rescorla-Wagner (1972) model and 2) the asymptotic predictions concerning blocking made by these two types of account. Chapter 3 analyses the influence that, according to the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988), a cue compounded with a previously trained CS will exert on the CR evoked by the latter. The remaining two experimental chapters of the thesis are devoted to the account of cue competition afforded by the attentional theory of Mackintosh (1975), introduced next.

1.5.2. Modified noncontinuity theories

That a number of accounts exist for blocking and relative validity which do not rely on attentional changes does not of course exclude the possibility that some attentional mechanism is indeed involved in these phenomena. Attentional theories, after all, most readily accommodate the evidence suggestive of associability changes in discrimination learning. One of the most influential theories of attention in the noncontinuity tradition is the theory of Mackintosh (1975), itself an extension of the theory of Sutherland & Mackintosh (1971).

1.5.2.1. The attentional theory of Sutherland & Mackintosh (1971)

According to Sutherland & Mackintosh (1971), discrimination learning is a two-stage process —perceptual (where associability changes take place) and associative. The associative component of the theory is couched in S-R terms, so that discrimination learning is said to result in the formation of response attachments. Each response is attached to the output of a particular stimulus-analyser. The function of any stimulus-analyser is that of processing the stimulus input along a certain dimension (Sutherland & Mackintosh, 1971, p. 33). In keeping with Thorndike's law of effect, the strength of such response attachments is posited to increase with reward and decrease with non-reward.

Of greater relevance for the associability changes the theory is designed to account for are the rules governing the changes in the strength of analysers. The strength of an analyser is equated with the amount of attention received by the stimuli along a certain dimension. It is assumed that the strength of a certain analyser

increases when its outputs consistently predict either the presence or the absence of reinforcement. The strengths of all analysers sum up to a fixed value, from which it is derived that the strengthening of one analyser directly leads to the weakening of all others. In this way, the theory avails itself of the notion of limited attentional capacity to propose a mechanism for the redistribution of attentional resources during discrimination learning. Other theories of attention prior to Sutherland and Mackintosh's (1971) share this notion (Lovejoy, 1968; Zeaman & House, 1963), which has come to be known by the name of *inverse hypothesis* (Thomas, 1970).

However, while it might be possible to salvage some aspects of Sutherland and Mackintosh (1971)'s account of discrimination learning, the phenomenon of unblocking (Kamin, 1969), as noted by Mackintosh (1975), clearly belies any account of cue competition derived from the inverse hypothesis. In this view, blocking results from an increase in attention to the pretrained cue A that involves a corresponding decrease in attention to the subsequently added cue X. If the inverse hypothesis is correct, then attention to X should be impaired by the prior increase in attention to A whatever X's correlation with reinforcement may be. Kamin (1969), however, demonstrated that rats do learn about the added cue (i.e. "unblocking") if the compound trial is followed by an increase in the magnitude of the US or the omission of the US altogether.

This result prompted Mackintosh (Mackintosh & Turner, 1971; Mackintosh, 1975) to reject the inverse hypothesis. He deduced that animals in a standard blocking experiment do not fail, as assumed in the inverse hypothesis, to attend to the blocked cue because they are already attending to a good predictor of reinforcement. Instead, they learn to ignore the added cue because it signals no change in the probability of reinforcement. This notion of learned inattention requires the postulation of a

mechanism underpinning associability losses that is independent from that promoting associability gains. Such is the task that Mackintosh (1975) set himself.

1.5.2.2. The attentional theory of Mackintosh (1975)

Like previous theories of discrimination learning, the theory of Mackintosh retains the cardinal assumption that animals pay more attention to reliable than unreliable predictors of important events. With respect to Sutherland & Mackintosh's (1971) theory, Mackintosh introduces a number of amendments (Mackintosh, 1975; Hall, 1991). For instance, associability changes are represented as variations in the value of a stimulus-specific parameter, α , ranging from 0 to 1. For a given stimulus, X, the value of α_X determines the size of the change in associative strength accrued on a subsequent trial (ΔV_X), given by the equation:

$$\Delta V_X = \alpha_X \Theta (\lambda - V_X), \quad 6$$

where the parameter Θ represents the properties of the particular US (e.g. intensity), λ represents the asymptotic associative strength supported by the US, and V_X the amount of associative strength acquired by X before the current trial. Following each learning episode, a comparison is made after each trial between the associative strength accumulated by X (V_X) and that acquired by other simultaneously present cues A (V_A). As a result of this comparison, the value of α_X is updated according to the rules:

$$\Delta \alpha_X > 0 \text{ if } |\lambda - V_X| < |\lambda - V_A| \quad 7$$

and

$$\Delta\alpha_X < 0 \text{ if } |\lambda - V_X| \geq |\lambda - V_A|, \quad 8$$

which capture the notion that the associability of a stimulus X will rise when it uniquely signals the occurrence of reinforcement, but it will drop when it signals no change in the prevailing expectation of reinforcement. Although Mackintosh (1975) does not give us a mathematical rule to calculate the magnitude of changes in α_X following a given trial, he proposes that it will be proportional to the difference between the error terms representing the predictive power of X ($\lambda - V_X$) and that of the other concurrent stimuli A ($\lambda - V_A$). As indicated above, this proposal of a rule for associability losses that is independent of the rule for associability gains is another distinctive feature of Mackintosh's (1975) theory.

A further novel element in the theory is the way it deals with stimulus salience in discrimination learning. Following Lovejoy (1968), Mackintosh (1975) recognises the inadequacy of adopting the simplistic view that stimuli arrive at the training situation with a certain level of associability, partly determined by their intrinsic salience, which can subsequently be freely modified according to their experience with reinforcement. Instead, he explains the influence of stimulus salience by referring to stimulus generalisation (Perkins, 1953; Logan, 1954). Thus, a stimulus of high salience is, on this account, one whose associative strength generalises little to other stimuli available, including the background cues. The degree of generalisation between two stimuli, X and A, is represented by the parameter $S_{X,A}$ (similarity). When a given cue X is paired with reinforcement on a given trial, the value of $S_{X,A}$

will determine the fraction of its increment in associative strength that will generalise to other CSs (A), in proportion to their degree of similarity:

$$\Delta V_A = S_{X,A} \alpha_X \Theta (\lambda - V_X) \quad 9$$

In this manner, the theory is able to predict that for two discriminanda signalling opposite outcomes, such as the presence and absence of reinforcement, a sufficiently high value of S will ensure that the discrimination is never solved completely (e.g. Hara & Warren, 1961). By applying a similar reasoning to simple conditioning, the theory can account for the fact that training with a weak cue can result in a relatively poor level of conditioning even at asymptote (Kamin, 1965).

Armed with this set of assumptions, the theory is well equipped to explain the results of discrimination-transfer experiments. For example, in an IDS-EDS experiment, the model successfully predicts the transfer discrimination to be difficult in the EDS group, as it rests on novel cues starting off with low associability generalised from the previously irrelevant stimuli belonging to the same dimension. By contrast, the cues on which the transfer discrimination depends in the IDS group will inherit high generalised associability by virtue of their resemblance with the prior relevant stimuli, and accordingly the discrimination should be relatively straightforward. Other phenomena accommodated by the theory are latent inhibition (Lubow & Moore, 1959; Lubow, 1973) and learned irrelevance (Mackintosh, 1973).

More important, the outcome of cue-competition experiments is correctly predicted by the theory. Relative validity is interpreted as resulting from a drop in associability by the target cue as its schedule of reinforcement, though the same in both groups, is comparatively poorer relative to that of the other concomitant stimuli

in the true- than the pseudodiscrimination group. Similarly, a decline in associability directly follows from the blocking condition, as a greater error term for X than for the pretrained A reduces X's associability and, therefore, thwarts its chances of acquiring some sizeable amount of associative strength.

The extant evidence specifically bearing on the theory's account of cue competition is rather scant. Mackintosh and Turner (1971) and Kruschke and Blair (2000) have succeeded in showing an attentional decrement to a blocked cue using the technique of "blocking of unblocking". As mentioned above, in the original study of "unblocking" Kamin (1969) demonstrated that blocking was abolished in rats if the compound trial signals an increase in the magnitude of the US, the delivery of a second US or the omission of the US altogether. Using the same species, Mackintosh and Turner (1971) were able to show that interposing a few regular blocking trials in between pretraining with A and the unblocking trials eliminated unblocking. This blocking of unblocking effect is readily explained by Mackintosh's (1975) theory, by assuming that the interpolated, blocking trials reduced the associability of the blocked cue X, subsequently impairing learning about this cue during unblocking trials. Kruschke and Blair (2000) confirmed the impact of blocking on the associability of the added cue using a disease diagnosis procedure, which also showed that this effect is not simply a consequence of the greater familiarity with the added cue for the group receiving blocking trials.

On the other hand, the evidence for attentional changes as a mechanism for relative validity is rather tangential. It is based on those demonstrations of associability changes in a true discrimination, AX+ BX-, using the techniques of the acquired distinctiveness of cues (Lawrence, 1949) and the IDS-EDS learning shift (e.g. Shepp & Eimas, 1964; Mackintosh & Little, 1969). As noted earlier, these

studies remain silent as to whether associability increases to the relevant cues A and B or decreases to the irrelevant cue X. In the absence of this basic knowledge, tackling the question of whether the associability of X is depressed in the true- with respect to the pseudodiscrimination in a relative validity experiment seems out of reach.

Put together, therefore, these results press for the need to devise behavioural techniques that enable us to measure associability changes in a more direct manner, so that the relevant questions can be probed. Such is the aim of Chapter 4 in this thesis, where a novel method designed to measure associability changes in true-discrimination learning is assessed. The success of such a method and its eventual application to the study of cue competition in general would not, in any case, offer proof that attentional changes are responsible for this class of phenomena. Indeed, the possibility remains that changes in attention play no causal role in the generation of the response deficits observed in these phenomena, but are rather a secondary effect of some other associative process. This point was acknowledged by Mackintosh (1975), who also suggested a strategy to test this is the case for blocking. According to the mechanism proposed in his theory, attentional changes should not begin, at the very least, until after the first conditioning trial with the compound. If this is correct, then provided that attentional changes are a necessary condition for blocking, it should not be possible to find one-trial blocking. Mackintosh (1976) himself confirmed this prediction. As he noted, however, this method has the disadvantage of being empirically confined to those experimental situations in which one-trial conditioning is sufficiently strong to make comparisons meaningful. Moreover, it is difficult to see how this method could be applied to experimental designs more elaborated than blocking, such as relative validity. In Chapter 5, a rather different

strategy is used to assess the causal role of associability changes in the relative validity effect (Wagner et al., 1968).

1.6. Summary

In the postscript of 1969 to his article entitled *Stimulus validity and stimulus selection in associative learning*, Wagner described it as “consoling” to have a modified continuity vs. noncontinuity controversy in place of the by then antiquated original continuity vs. noncontinuity one (Wagner, 1969b, p. 121). Just under 40 years on, our deficient understanding of the mechanisms driving cue competition makes the matter appear recalcitrant. The proliferation of theoretical accounts in the field of associative learning, as well as the stark contrast between the mechanisms they posit, are indicative of the great challenge facing any research trying to establish what truth lies in each of them. This thesis represents an attempt towards a systematic assessment of the mechanisms involved in cue competition. Within its limited scope, it provides an empirical analysis of two of the most prominent theories in the field: the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988, Chapters 2 and 3) and the attentional theory by Mackintosh (1975, Chapters 4 and 5).

Chapter 2

2. Analysis of the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988)

2.1. Introduction

In the present chapter, I evaluate the accounts of cue competition advanced by the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) and the Rescorla-Wagner (1972) model. Each of the two experiments that follow tests a unique prediction of the Comparator Hypothesis thus far neglected in the associative learning literature. The first of these predictions concerns conditioned inhibition.

It is a well-established fact that intermixed presentations of X+ and XB- trials result in stimulus B becoming a conditioned inhibitor (e.g. Pavlov, 1927). What is precisely meant by B accruing inhibition, and how the latter develops remains a much heated controversy among current theories of associative learning. One strategy that could help to elucidate the nature of inhibition consists in examining the influence that B may exert on X's associative status.

In the context of the Rescorla-Wagner (1972) model, for instance, inhibition is equated with negative associative strength. According to this view, a stimulus will develop negative associative strength when it is presented without consequences in the presence of another cue which otherwise signals the US occurrence. Thus, following conditioned-inhibition training of the form X+ XB-, B will acquire negative associative strength as X gains positive associative strength. In the early stages of training, some of the positive associative strength acquired by X on X+ trials will be lost on XB- trials. As training progresses, however, B will gradually acquire the total amount of negative associative strength available on XB- trials, and as a result X will

be protected from further extinction. Therefore, the Rescorla-Wagner (1972) model anticipates that presenting X after an X+ XB- discrimination will result, other things being equal, in stronger responding to X than if the same training were given in the absence of B.

The notion that inhibitors may protect from extinction the excitors fostering the acquisition of their inhibition cannot be readily derived from the principles of the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988). This follows from the way in which the model deals with conditioned inhibition. According to this account, associations forged in conditioning can be only excitatory. Whether a stimulus serves as an excitor or an inhibitor following conditioning training is determined by the outcome of the comparison process taking place at the time of testing. Only when the excitatory status of the comparator is markedly superior to that of the target stimulus, will the target manifest the properties of a conditioned inhibitor. It further follows that an inhibitor, which will be incapable of activating the US representation, will not influence the comparator process with any excitor with which it has been paired. It is not obvious, therefore, how a conditioned inhibitor like B might protect X from the disruptive effects of the partial reinforcement schedule during an X+ BX- discrimination. Thus, in contrast to the Rescorla-Wagner (1972) model, the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) predicts that a partial-reinforcement schedule for X should lead to similar levels of responding to this stimulus whether or not B co-occurs on nonreinforced trials.

Direct comparison of performances to X following either an X+ X- or X+ XB- schedule, however, raises a number of interpretative difficulties. To mention but one, nonassociative factors such as the partial-reinforcement extinction effect may be expected to have uneven repercussions for responding to X in each of these

treatments. Experiment 1 attempts to settle these contrasting views on the role of conditioned inhibitors with a design that obviates such difficulties.

The second unique prediction advanced by the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) regards the fate of blocking with extended training. In a typical blocking experiment (e.g. Kamin, 1968), a CS A is initially conditioned to signal the US occurrence. Following this training, the same stimulus is presented alongside a second CS, X, and the compound is subsequently paired with the US. At test, X typically evokes little conditioned responding by comparison with a control group in which A does not receive pretraining in isolation.

According to the Rescorla-Wagner (1972) model, a stimulus will only acquire associative strength if it is paired with a surprising US. In the blocking group, however, the US is already predicted at the outset of training with AX by virtue of the preceding training with A. Under these circumstances, the US will no longer support further learning, and X will accordingly fail to enter into an association with it. It is important to note that this should be the case regardless of the amount of training given with AX. Thus, like other traditional accounts of associative learning (e.g. Mackintosh, 1975; Pearce & Hall 1980; Pearce 1987, 1994), the Rescorla-Wagner (1972) model envisages blocking as a permanent phenomenon.

In contrast, the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) predicts that as training proceeds with AX, blocking should tend to disappear. For the following discussion, bear in mind that contiguity is the sole determinant of the associative strength a stimulus acquires. Whether or not a stimulus is accompanied by another CS, all that determines its associative strength is the number of times it has been paired with the US and the nature of the reinforcement schedule. During training with AX in the blocking group, therefore, X will acquire as

much associative strength as it would in the absence of A. At test, however, the measure of associative strength that X will have accrued does not directly translate into performance. Instead, performance to X results from the comparison between its associative strength and the associative strength of A, its comparator stimulus. The greater A's associative strength is relative to that of X, the more A will interfere with the expression of X's associative strength in conditioned responding. Thus, pretraining with A in a blocking experiment enables this stimulus to reduce performance to X to a greater extent in the blocking than the control group. Blocking of X by A should be short-lived, however, as further training with AX will lead to equivalent individual asymptotes of learning for A and X in both groups. Continued training should therefore tend to equate the target-comparator relationship across the blocking and control conditions and eliminate thereby the very imbalance through which the theory accounts for blocking. Thus, contrary to the Rescorla-Wagner (1972) model, the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) predicts that blocking is a transient phenomenon.

The predictions derived from the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) have received support from at least one study in which continued training with AX eliminated blocking (Azorlosa & Cicala, 1988). Unambiguous interpretation of these results, however, is hampered by a confound arising from the sequential structure of the standard blocking design. If, as testing these conflicting predictions implies, training with AX is prolonged so as to secure asymptotic learning, it may as well be expected that the memory of pretraining with A should naturally decay over time. To the extent that blocking depends on the durability of this memory, an attenuation of the effect is conceivable on these grounds alone. To circumvent this problem, Experiment 2 used a single-phase blocking design,

in which trials with A and AX were randomly interspersed. As will become clear shortly, the discrepant predictions of the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) and the Rescorla-Wagner (1972) model over the fate of blocking to X remain, with this design, essentially unaffected.

2.2. Experiment 1

In Experiment 1 (Table 2.1) two groups of rats received training using a Pavlovian appetitive conditioning procedure. The TD group was trained with a true discrimination in which the compound AX was paired with food and BX was paired with nothing. The other group, FP, was trained with a feature-positive discrimination (Pavlov, 1927) in which AX was paired with food, but X by itself had no programmed consequences.

Table 2.1. Design of Experiment 1.

Group	Acquisition training		Probe tests
True discrimination (TD)	AX+	BX-	A-
Feature-positive (FP)	AX+	X-	A-

Thus, the individual schedules of reinforcement for A and X were identical across the groups, which only differed in the presence of B on nonreinforced trials in the TD group. On Sessions 13 to 18, single, nonreinforced test trials with elements A,

B and X were intermixed in the otherwise regular training schedule. The index of the CR evoked by these stimuli was the duration of anticipatory magazine activity in their presence.

The question of interest was whether the presence of B would have any influence upon X's fate as a conditioned excitor in the TD group with respect to the FP group. It should be noted, however, that any interpretation based on the direct comparison of performance to X at test will be hindered by the unequal degree of generalisation decrement that X should undergo across treatments. A more convenient way to compare the groups in terms of X's status as a conditioned stimulus is to measure its capability to overshadow A. Accordingly, the analysis of the results from the test data will focus on test trials with A, whereas those with B and X, of ambiguous interpretation, will be only briefly reported in the text. Comparison of the extent to which A will be overshadowed by X in the two groups provides a straightforward test of the account of cue competition advanced by the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988), set against the background of the Rescorla-Wagner (1972) model.

According to the Rescorla-Wagner (1972) model, an AX+ BX- discrimination will be solved in the following manner. At the start of conditioning, stimuli A and X will share increments of associative strength on reinforced trials in proportion to their relative salience. X, however, will lose part of its associative strength on compound trials with B, which, itself never reinforced, will become a conditioned inhibitor. As training progresses, A, consistently reinforced, will secure the lion's share of associative strength, whereas B will become a sufficiently strong inhibitor to cancel out X's modicum of associative strength on nonreinforced trials. Once this condition is met, so that the summed associative strength of the BX compound is close to

nought, the presence of B will continue to protect X from extinction in the TD group, and as a result X will be in a position to compete with A for an asymptotic share of associative strength. In the FP group, by contrast, X is in B's absence more vulnerable to extinction, and hence, less of a competitor with A for associative strength than in the TD group. It follows from this account that discrimination training on these two conditions should result in subjects in the TD group responding more vigorously to A at test than in the FP group.

In contrast to this prediction, the Comparator Hypothesis of Miller and collaborators ((Miller & Schachtman, 1985; Miller & Matzel, 1988) anticipates no difference in the strength of the CR that A should evoke as a result of the two schedules. Over the course of training, equivalent schedules of reinforcement for A and X across groups should result in their acquiring similar amounts of associative strength. The continuous nonreinforcement schedule on which B is trained in the TD group, on the other hand, should preclude the formation of an association between B and the US. It should be evident that, with no capability to activate the US representation, the probability that B will act as a comparator stimulus for X in the TD group should be minimal. Excluding the influence of B on the expression of the X-US association leaves X as an equally effective comparator stimulus for A in both groups. As a result, the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) predicts that the output of the comparator process that takes A as the target stimulus should be equivalent across the groups, and therefore, that an equally strong CR during A should be observed irrespective of the treatment given.

Experiment 1 tested these conflicting predictions between the Rescorla-Wagner (1972) model and the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988).

Method

Subjects. Twenty-four male, hooded Lister rats (*Rattus norvegicus*) supplied by Harlan Olac (Bicester, Oxon, England) served as experimental subjects. They had previously participated in an experiment on spatial learning, in which they received training in a Morris pool. They were housed in pairs in a light-proof holding room in which the lights were on for 14.5 hr/day. Prior to the start of the experiment, at an age of approximately 5 months, they were gradually reduced to 80% of their free-feeding weights. They were then maintained at these weights throughout the experiment by being fed a restricted amount after each experimental session. Experimental sessions were conducted at the same time each day, during a period when the lights were on in their holding room.

Apparatus. All experimental procedures were performed in eight identically specified operant boxes (24.5 x 23.0 x 21.0 cm) supplied by Campden Instruments Ltd., Loughborough, UK, which were housed in separate light- and sound- attenuating chests. A ventilating exhaust fan in each of the chests provided a background noise of 72 dB (C scale). Three walls of each operant box were constructed from aluminium, while a clear acrylic door served as the fourth wall. The ceiling was translucent white acrylic, and the floor was a grid made of stainless steel rods. The front aluminium wall housed a recessed food magazine (5.0 x 6.0 cm) with its base located 0.5 cm above the grid floor. A 5- Ω loudspeaker located on the ceiling delivered a 10-Hz click at an intensity of 83 dB (C scale). A 240-V, 60-W strip light was mounted above the translucent ceiling, and this light could be illuminated to provide a constant light stimulus. Three 2.8-W lamps covered by 1.5-cm diameter plastic discs were mounted on the front aluminium wall on top of and on either side of the recessed food

magazine. Only the last two, 15.0 cm above floor level and 12.5-cm apart, were used. On reinforced trials, a 45-mg food pellet (traditional formula, P. J. Noyes, Lancaster, NH) could be delivered into the magazine tray by means of a rotatory magazine dispenser. The magazine was covered by a clear acrylic flap hinged at the top, which the animals had to push open in order to gain access to the food pellets as they were delivered. A 1.0-cm deep rectangular frame was attached to the front wall around the opening to the magazine. Set into this frame, 5 mm in front of the magazine flap, were three pairs of photodiode sensors in such a position that horizontal infrared beams were located 10, 20, and 30 mm above the grid floor. Appropriate circuitry permitted a Risc PC microcomputer (Acorn Computers Ltd., Cambridge, England), programmed in Arachnid (Paul Fray Ltd., Cambridge, England), to record the responses upon detecting the interruption of these beams and to control the experimental events.

Procedure. All animals initially received two 30-min sessions of magazine training, during which one food pellet was delivered to the tray at regular 1-min intervals. After the end of each of these sessions the rats remained in the conditioning chambers for another 30-min period for acclimation purposes. In the first session only, the acrylic flaps of the food magazines were taped open. Following magazine training, the subjects were randomly assigned to two groups, labelled True discrimination (TD) and Feature-positive discrimination (FP), of equal size ($n=12$), to receive Pavlovian appetitive conditioning. During acquisition training, the TD group received 10 presentations of each of AX and BX within a session, whereas the FP group received 10 presentations of each of AX and X within a session. In the TD group, presentations of AX, but not BX, were followed by the delivery of a food pellet. The FP group

experienced reinforcement following AX, but there were no programmed consequences following X trials. The trial sequence was random with the constraint that no more than two trials of the same type could occur in succession. All stimulus presentations lasted 10 s, and were separated by a 2-min variable ITI schedule, defined as the duration between the termination of one trial and the onset of the next. The stimuli that served as A and B were two visual stimuli appropriately counterbalanced. One visual cue, a flashing light, was constructed by alternating the illumination of the two bulbs located to the left- and right-hand sides of the food magazine, in such a manner that each bulb remained on for 0.5 s and its offset marked the onset of the other. For the other visual cue, a houselight, the strip light mounted above the translucent ceiling was used to provide a constant and diffuse illumination over the conditioning chamber, which otherwise remained dark when neither of the visual cues was on. In both groups, a train of clicks served as stimulus X. Eighteen acquisition sessions were conducted, each of approximately 1 h in duration. From sessions 13 to 18, three daily, nonreinforced probe trials, one for each of the A, B and X elements, were randomly intermixed with the regular acquisition schedule. On both acquisition and test trials, the cumulative time that the subjects spent with their heads in the magazine area over the 10 s that each trial lasted served as a measure of conditioned responding. Similar pre-CS measures were taken over the 10 s preceding the presentation of each trial to provide an indication of contextual conditioning.

Results. Figure 2.1 shows the results of acquisition training in the true- and the feature-positive-discrimination groups (TD and FP groups, respectively), plotted in 2-session blocks. Superimposed on these data, across the final three 2-session blocks,

are the results of the test trials with A. Because of their direct relevance for comparing performance to A at test, the statistical analysis of the acquisition data was confined to the final three 2-session blocks. Starting with a glance at the acquisition data, it is apparent that both groups had successfully solved their respective discriminations by the time the test trials with A were administered. Further inspection of these data also suggests that, overall, the TD group responded more vigorously than the FP group during both types of trial. These observations were confirmed by a $2 \times 3 \times 2$ ANOVA of individual mean durations of magazine activity with the factors of group, session block and stimulus, which revealed an effect of group, $F(1, 22) = 6.09$, an effect of session block, $F(2, 44) = 9.18$, and an effect of CS, $F(1, 22) = 183.28$, as well as a Session block \times CS interaction, $F(2, 44) = 6.52$. The Group \times Session block, $F < 1$, Group \times CS, $F < 1$, and Group \times Session block \times CS, $F(2, 44) = 1.79$, interactions were not significant.

Focusing now on the test trials with A, a look at the graph reveals higher conditioned responding to this cue in the FP than the TD group. A 2×3 ANOVA of individual mean durations of magazine activity with the factors of group and session block revealed an effect of session $F(2, 44) = 3.45$, but failed to reveal an effect of group, $F(1, 22) < 1$, or a significant Group \times Session block interaction, $F(2, 44) = 2.53$. The p value for the Group \times Session block interaction ($p = 0.09$) was, however, close to the conventional level of significance.

Inspection of the acquisition data suggests a possible explanation as to why the test data with A might have failed to reach significance. Throughout the final stages of training the rate of responding during both reinforced and nonreinforced trials was faster in the TD than the FP group. It is possible that if this difference had not existed then the difference between the groups in terms of responding to A would have been

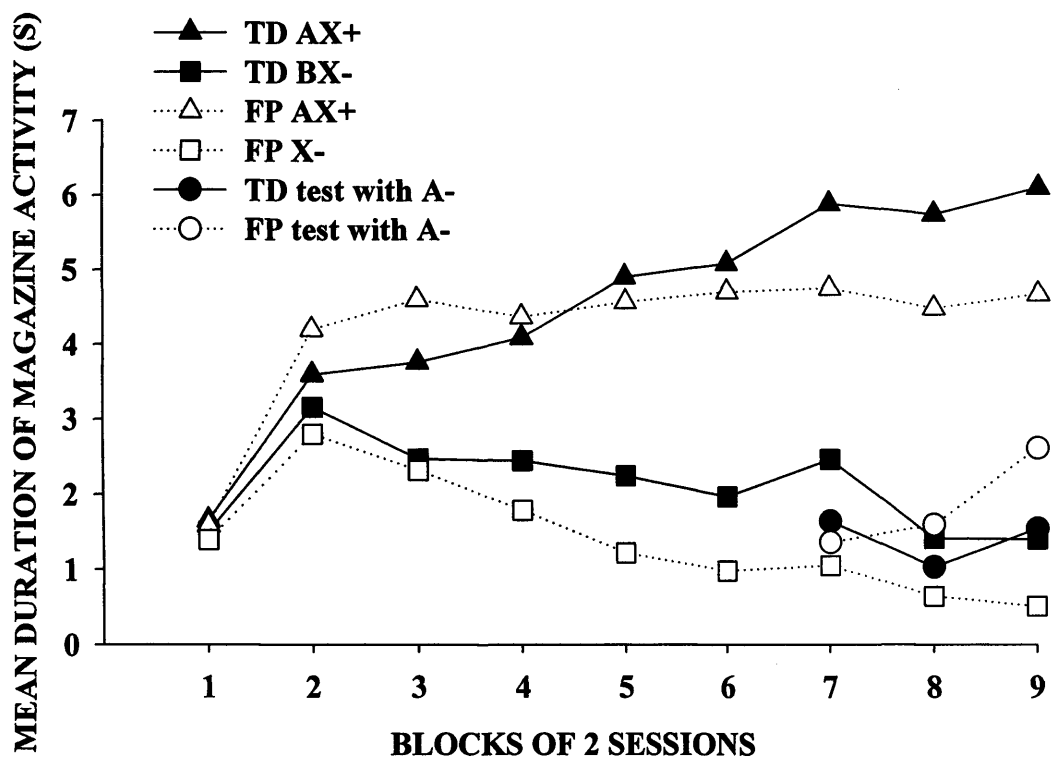


Figure 2.1. Acquisition and test data in the true-discrimination (TD) and feature-positive (FP) groups.

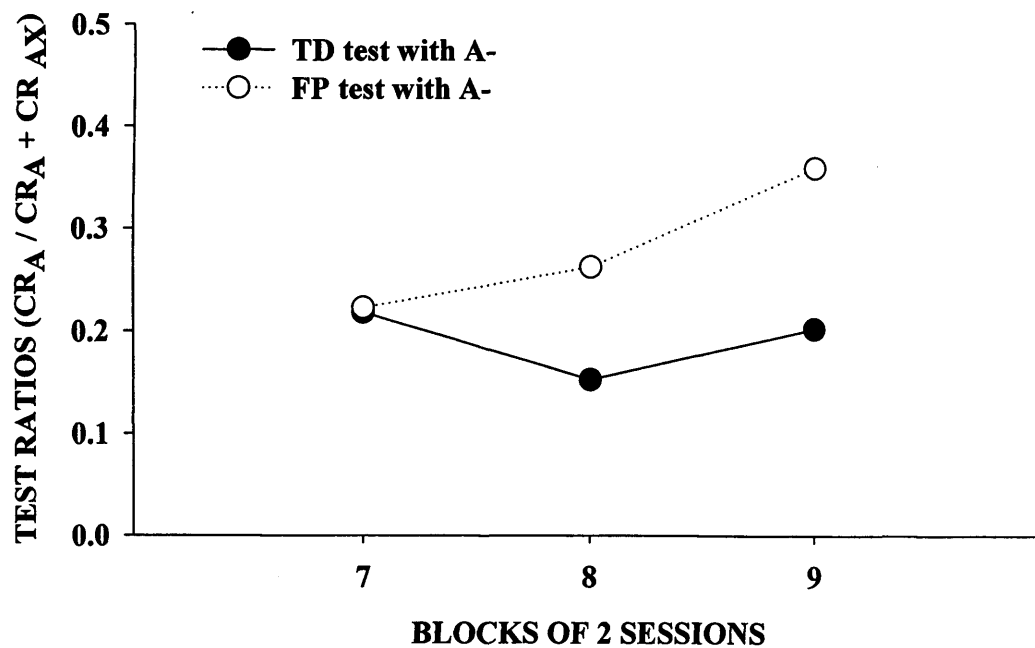


Figure 2.2. Test ratios.

greater than was found. In order to take account of this possible influence of the difference rates of responding during the training stimuli, individual test ratios of the form $CR_A/(CR_A+CR_{AX})$ were calculated for the two groups, where CR_A represents the individual mean durations of magazine activity in the presence of A, and CR_{AX} represents the individual mean durations of magazine activity in the presence of AX+, across the final three 2-session blocks of training. The ratios are plotted in Figure 2.1. Inspection of this panel shows that responding during A was more vigorous relative to AX in the FP than the TD group. A two-way ANOVA of individual ratios with the factors of group and session block statistically confirmed the difference between the groups by revealing a significant effect of session, $F(2, 44) = 3.34$ and, crucially, a Group x Session block interaction, $F(2, 44) = 3.56$. No significant effect of group was revealed, $F(1, 22) = 2.02$. Exploration of the significant Group x Session block interaction using simple main effects analysis revealed an effect of session block for the FP group, $F(2, 44) = 4.97$, but no effect of session block for the TD group, $F(2, 44) = 1.93$. Furthermore, there was a significant effect of group on Session block 3, $F(1, 66) = 6.06$, but no effects of group were revealed on Session blocks 1 and 2, $F_s(1, 66) < 1.68$.

With regard to the test trials administered with the other elements, X and B, the mean duration of magazine activity for the three 2-session blocks combined in the TD and the FP groups were, respectively, 3.24 and 0.79 in the presence of X, and 0.16 and 0.32 in the presence of B. Groups comparisons of individual mean durations of magazine activity during these stimuli across the three 2-session blocks combined revealed a significant difference between the groups in responding to X, $t(22)=3.49$, but no difference in responding to B, $t(22)=1.22$.

Discussion. In Experiment 1 two groups, TD and FP, received training with a true discrimination (AX+ BX-) and a feature-positive discrimination (AX+ X-), respectively. Single, daily test trials with stimuli A, B and X, interpolated in a normal conditioning session, were conducted over the final six sessions of acquisition training. Of particular theoretical interest was the comparison of performance to A between the two groups. The test trials with A revealed a stronger CR during this cue in the FP than the TD group. This difference fell short of significance when the rates of responding during A were compared directly for the two groups. This difference was, however, significant when account was taken of the overall faster rate of responding by the TD than the FP group by comparing response rate ratios. Thus, it appears that the capability of A to elicit a CR was greater after the training given in the FP than the TD group.

One might as well wonder why the overall level of responding in the presence of A at test was so low. Perhaps the answer lies partly in the fact that the stimuli used as A were of the visual modality, which in our lab usually elicit a fairly moderate amount of conditioned responding, at least for magazine approach procedure. A second explanation may be related to the fact that A was presented on a single occasion for test during six consecutive days. Such single presentations may have been particularly vulnerable to the effects of generalisation decrement. In any case, the results are in keeping with the predictions of the Rescorla-Wagner (1972) model, according to which X should be better qualified to overshadow A in the TD compared with the FP group. In support of this account, it was found that responding to X at test was significantly higher in the TD than the FP. Underlying X's differential capability to overshadow A across the groups is the presence of B on nonreinforced trials in the TD group, which acts to shield X from extinction.

The results from the experiment are not so readily explained by the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988). On this account, if X is to reduce at all responding to A, it should do so in a similar manner across the groups. This follows from B's ineffectiveness to serve as a comparator stimulus for X and, thereby, to influence X's effectiveness to serve in turn as a comparator stimulus for A in the TD group. All else being equal, there should be thus no difference between the groups in terms of the strength of the CR elicited by A at test. Experiment 1 disconfirmed this prediction of the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988), and pointed to what appears as a clear inadequacy thereof; namely, its disregard for the role that a continuously-nonreinforced stimulus may play in the fate of the conditioned excitor extinguished in its presence.

2.3. Experiment 2

In Experiment 2 a single-phase blocking design (Table 2.2) served to determine whether blocking is transient as uniquely predicted by the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988). Three groups of rats were trained with the appetitive procedure that was used in Experiment 1. In the Blocking group, intermixed presentations with A and AX were followed by food. One overshadowing control group (Overshadow) received the same number of reinforced presentations with AX as the Blocking group. A second control group (Overshadow-2) experienced a similar overshadowing treatment, except that AX was presented

twice as many times as for the Blocking and Overshadow groups. All groups received nonreinforced presentations of BY. The purpose of the study was to evaluate conditioning to X among the groups as a result of this training. For reasons that will presently become clear, four test trials with X were administered to the three groups on Sessions 10, 20, 30 and 40. As in Experiment 1, the duration of anticipatory magazine activity was used as an index of conditioned responding.

Table 2.2. Design of Experiment 2.

Group	Acquisition training	Test trials
Blocking	A+ AX+ BY-	X-
Overshadow	AX+ BY-	X-
Overshadow-2	AX+ AX+ BY- BY-	X-

According to the Rescorla-Wagner (1972) model, with sufficient training the Blocking group should come to exhibit a deficit in responding to X relative to the control groups, which themselves should not differ. In the Blocking group, presentations with A by itself will allow this CS to continue gaining associative strength once conditioning with AX has reached asymptote. The associative strength of X will be driven to almost zero in order to ensure that the overall associative strength of AX remains close to asymptote. By comparison with the Blocking group, the simple overshadowing treatment in the control groups should, barring some deleterious asymmetry in their relative salience, result in substantial conditioning to

X. At asymptote, moreover, the initial difference in X's associative strength between the Overshadow-2 and the Overshadow groups produced by their unequal number of AX trials in each session will tend to disappear. Over extended training, therefore, the Rescorla-Wagner (1972) model predicts an ever larger blocking effect in the Blocking group as reflected, to a similar degree, by comparison with each of the Overshadow-2 and Overshadow groups.

In contrast, the Comparator Hypothesis of Miller and collaborators (1985, 1988) predicts that the effects of single-stage blocking should disappear over extended training. This prediction may be derived from one basic tenet of the theory: that stimuli present in conditioning undergo independent learning experiences leading to independent learning asymptotes. During the initial stages of conditioning in the Blocking group, reinforced presentations of A by itself will result in this stimulus acquiring associative strength more rapidly than X. Accordingly, if X were tested early in conditioning little responding should be observed in the Blocking group as a consequence of the detrimental comparison with the relatively stronger cue A. This way the theory is able to predict the phenomenon of single-stage blocking. As A and X approach asymptotic learning, however, the model anticipates that A's head-start over X will diminish until its capability of reducing responding to X is no superior to that following asymptotic overshadowing training. If this were the case, then test trials with X after extended training should reveal the abolishment of blocking, as indicated by the absence of a difference in performance between the Blocking and the Overshadow-2 groups. Although a failure to confirm this prediction might appear to challenge the account of blocking provided by the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988), it is worth noting that at any point in training the Blocking group will have received half as many conditioning trials with X

as the Overshadow-2 group. The difference in responding to X could then be attributed to conditioning with this stimulus being more effective in the Overshadow-2 than the Blocking group. One way to test this possibility is to compare performance to X in the Overshadow-2 group with that in the Overshadow group, which received the same number of X presentations as the Blocking group. Evidence of similar performance to X in the two control groups should indicate that conditioning with X has been comprehensive enough in the Blocking group to secure asymptotic learning. A rather more serious challenge for the model would therefore arise if poorer performance to X in the Blocking group persisted beyond the point at which the Overshadow-2 and Overshadow groups come to behave similarly.

The success of the study thus rests on the assumption that, with sufficient training, the Overshadow-2 and Overshadow groups will not differ in their performance to X. Provided this condition is satisfied, Experiment 2 should demonstrate either that blocking is a permanent response deficit in the manner envisaged by the Rescorla-Wagner (1972) model, or an ephemeral one, as derived from the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988). Given the need to track the progress made by X in the control groups, test trials with this stimulus were administered every ten sessions. The results from these tests supplied the necessary data to evaluate the fate of single-stage blocking over extended training in the Blocking group.

Method

Subjects and Apparatus. The subjects were 24 experimentally naïve, male, hooded Lister rats (*Rattus norvegicus*) supplied by Harlan Olac (Bicester, Oxon, England) that were maintained in an identical fashion to the subjects in Experiment 1. The apparatus was the same as in Experiment 1, except for an extra 5-Ω loudspeaker located on the ceiling of the conditioning chamber which delivered a 1-KHz tone at an intensity of 80 dB (C scale).

Procedure. The details of magazine training were identical to those in Experiment 1. Following this phase, the subjects were randomly assigned to three groups of equal size ($n = 8$) to receive Pavlovian appetitive conditioning. A single-phase-blocking group, B, received 8 intermixed presentations of each of A and AX followed by food, making a total of 16 reinforced trials. In the other two, Overshadow-2 and Overshadow groups, reinforced presentations with AX were delivered on 16 and 8 occasions, respectively. In addition, all groups received a number of nonreinforced trials with the compound BY that matched the number of X presentations in each group. The purpose of these trials was to sharpen the generalisation gradients around the reinforced stimuli. Thus, for the Blocking, Overshadow-2 and Overshadow groups, the total number of trials within a session was 24, 32 and 16, respectively. The trial sequence was randomized with the constraint that no more than two trials of the same type could occur in succession. The CS and ITI durations were the same as for Experiment 1. Likewise, the same flashing light and houselight used in Experiment 1 served as the visual stimuli that functioned as A and B. The auditory stimuli that served as X and Y were, respectively, a train of clicks as that used in

Experiment 1 and a tone. Stimuli serving as A and B, but not those serving as X and Y, were counterbalanced. Forty acquisition sessions were conducted. On Sessions 10, 20, 30 and 40, four consecutive test trials with X in extinction were administered following four initial acquisition trials. These test sessions were then completed with the application of the remaining acquisition trials of a regular session. Details regarding the measurement of the conditioned response in the CS and preCS periods were identical in all respects to those of Experiment 1.

Results. The results of acquisition training in the Blocking group are shown in Figure 2.3, whereas those in the Overshadow-2 and Overshadow control groups are shown in Figure 2.4. A glance at the two panels reveals that all three groups had solved their respective discriminations by the third acquisition session. Moreover, there is an indication that the groups did not differ in terms of their performance in the presence of AX from Session 7 onwards, by which time they all had attained asymptote. To determine the truth in these observations, a 3 x 4 ANOVA of individual mean durations of magazine activity during AX with the factors of group and session was conducted for the four sessions that contained test trials with X (Sessions 10, 20, 30 and 40). The results from the ANOVA confirmed similar levels of responding to AX among the groups by revealing nonsignificant effects of group, $F(2, 21) = 2.37$, session, $F(3, 63) = 1.73$, and a nonsignificant Group x Session interaction ($F < 1$). Considering performance in the Blocking group, it is apparent that acquisition of conditioned responding proceeded more gradually in the presence of stimulus A than in the presence of AX. This suggestion was confirmed by a 2 x 4 within-subjects

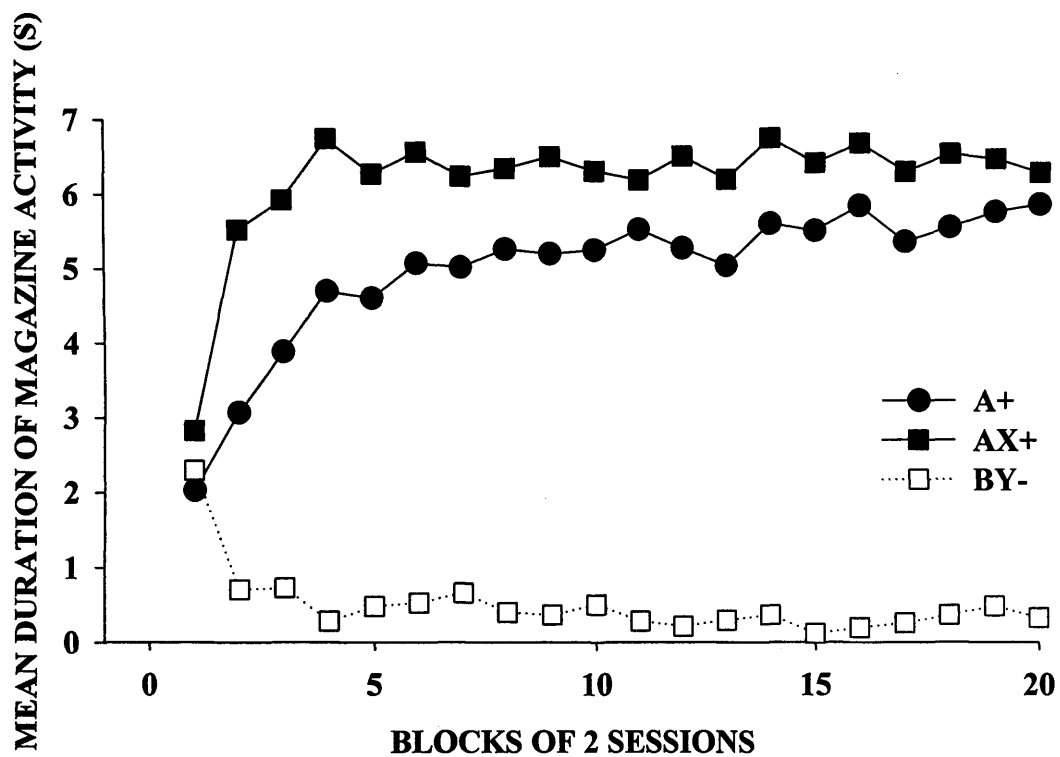


Figure 2.3. Acquisition in the Blocking group.

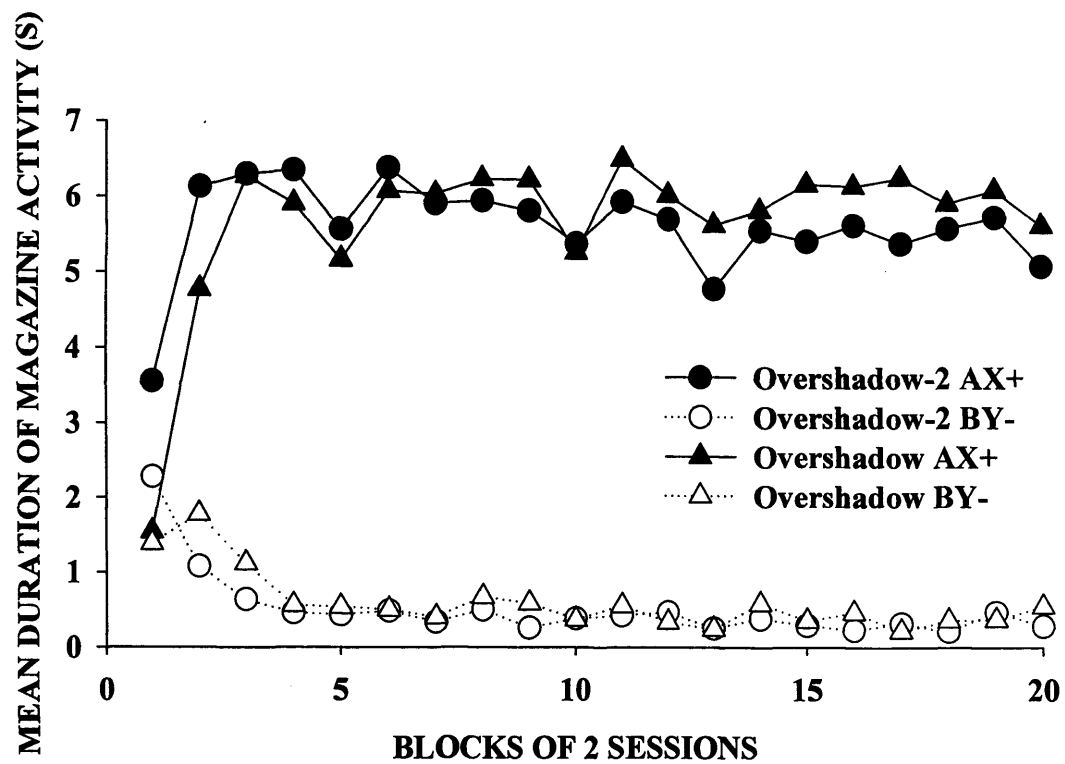


Figure 2.4. Acquisition in the control groups (Overshadow-2 & Overshadow).

ANOVA of individual mean durations of magazine activity during A and AX with the factors of CS and session, also restricted to Sessions 10, 20, 30 and 40. The results from the ANOVA revealed an effect of CS, $F(1, 7) = 22.87$, but failed to reveal an effect of session ($F < 1$), or a significant CS x Session interaction, $F(3, 21) = 2.14$. The results from the test trials with X are plotted in Figure 2.5. Inspection of these data shows the groups responding initially with equivalent strength. Eventually, however, performance in the Blocking group declined relative to the Overshadow-2 and Overshadow groups, which themselves did not differ. In support of these observations, a 3 x 4 ANOVA of individual mean durations of magazine activity with the factors of group and session, revealed a significant Group x Test Session interaction, $F(6, 63) = 2.28$. No effects of group, $F(2, 21) = 1.46$ or test session were

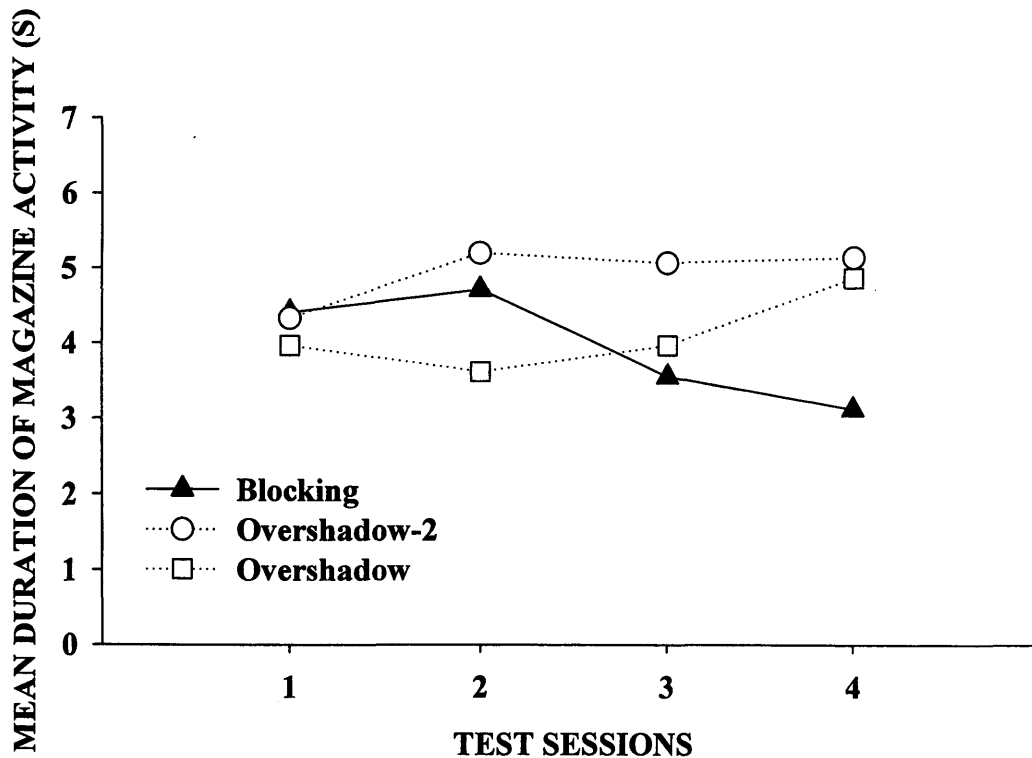


Figure 2.5. Responding during X in the groups across the 4 test sessions.

revealed ($F < 1$). Exploration of the significant interaction using simple effects analysis found no significant difference among the groups on Test Sessions 1-3, $F_s(2, 84) < 1.90$, but an effect of group on Test Session 4, $F(2, 84) = 3.45$. Further analysis of this effect using Tukey's HSD test revealed that the B group differed significantly from both the Overshadow-2 and the Overshadow groups. However, no difference was found between the Overshadow-2 and the Overshadow groups.

Discussion. In Experiment 2 one group of animals, the Blocking group, received training in which randomly-intermixed presentations of A and AX were followed by food. Two control groups, Overshadow-2 and Overshadow, experienced treatments in which AX presentations were likewise reinforced, but their number in the Overshadow-2 group was twice that in the Overshadow group. On Sessions 10, 20, 30 and 40, test trials with X were delivered to all the groups.

In the early stages of training, both the Rescorla-Wagner (1972) model and the Comparator Hypothesis ((Miller & Schachtman, 1985; Miller & Matzel, 1988) predict that presentations of A by itself should block responding to X in the Blocking group. The question of interest was whether A should continue to do so after learning about X has reached asymptote. One indicator that learning about X in the Blocking group has become asymptotic comes from the comparison of the Overshadow group, which received the same number of X presentations as the Blocking group, with the Overshadow-2 group, which received twice as many X presentations as the Overshadow and the Blocking groups. If it could be shown that X evokes a comparable measure of responding at test in the Overshadow and Overshadow-2 groups, then there should be sufficient ground to believe that learning about X in the Blocking group must itself have reached asymptote. Under these circumstances,

evidence for a persistent deficit in performance to X in the latter group relative to the two control groups could not reasonably be attributed to inadequate training. Moreover, it would be equally unlikely that such deficit might be due to A being a stronger comparator for X in the Blocking than the Overshadow-2 groups, since A was presented in both groups on the same number of occasions. An explanation in terms of blocking would thus appear necessary.

Such is the pattern of results revealed on the final test session of Experiment 2. A deficit in performance to X in the Blocking group was indeed observed with respect to the Overshadow-2 group. Crucial for a blocking interpretation was the finding that on the same test session no difference existed between the Overshadow-2 and Overshadow groups. On the basis of this finding, it is plausible to conclude that blocking of X was effective despite conditioning with this cue having reached asymptote.

A ready explanation for single-stage blocking is advanced by the Rescorla-Wagner (1972) model. On this account, blocking should begin to develop once the summed associative strength of AX has reached asymptote. From this point onwards, further gains accrued by A on trials in which it is presented by itself will tend to inflate the summed associative strength of AX over its asymptotic or λ value, resulting in overexpectation of reinforcement. In order to eliminate overexpectation, and therefore adjust the summed associative strength of AX to the maximum reinforcement value supported by the US, X will have to start losing associative strength. Thus, by the time A has become an asymptotic predictor of reinforcement in its own right, X's associative strength will have been driven to nought despite continuous reinforcement.

Not only can the Rescorla-Wagner (1972) model explain the phenomenon of single-stage blocking, but it can also take account of its belated emergence in the present experiment. To do so, it simply needs to assume that the salience of X (a train of clicks) is considerably higher than that of A (two kinds of lights), an assumption supported by previous pilot data from our laboratory. If this were the case, then X should be capable of acquiring substantial amounts of associative strength in all groups early in training. This conclusion is confirmed by the high, equivalent levels of performance in the presence of X for all groups. In the Blocking group, on the other hand, reinforced trials with A by itself, a much weaker stimulus judging from its slow acquisition rate, should be initially of no consequence to X's excitatory properties. Indeed, only with extended training will acquisition with A have progressed sufficiently far to undermine, via overexpectation on AX trials, performance to X significantly. This explains why blocking could only begin to be seen after forty conditioning sessions.

Less capable of accounting for the tardy emergence of blocking is the Comparator Hypothesis of Miller and collaborators (Miller & Schachtman, 1985; Miller & Matzel, 1988). According to this theory, blocking should occur relatively early in training in the Blocking group as a consequence of the higher proportion of A to X trials turning the former stimulus into a deleterious comparator for the latter. As noted previously, however, such head-start to A should endow this cue with only a temporary advantage in reducing responding to X. This is because once X has become an asymptotic predictor of the presence of food, A should be no more detrimental as X's comparator in the Blocking group than following asymptotic overshadowing training of the sort given to the Overshadow-2 and Overshadow groups. If, as the data seem to imply, it is assumed that learning about X has reached asymptote in the

Overshadow-2 and Overshadow groups by the first test session, it follows from this account that no response deficit should be observed in the Blocking group on this or any other of the subsequent test sessions. This prediction is clearly disconfirmed by the outcome of the final test session.

To sum up, the results of Experiment 2 are in keeping with the theoretical account of blocking advanced by the Rescorla-Wagner (1972) model, but pose a challenge for that put forward by the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988).

2.4. General Discussion

The two experiments presented in this chapter tested two unique predictions of the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988), hitherto overlooked. Experiment 1 explored the prediction that a conditioned inhibitor should exert no influence on the response-eliciting properties of the excitator extinguished in its presence during training. Experiment 2, on the other hand, tackled the prediction that blocking should disappear over extended conditioning. For diverse theoretical reasons, as discussed above, the results of these experiments challenge the account of cue competition offered by the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988).

In Experiment 1, one group of rats (TD) received training with a true-discrimination, AX+ BX-, whereas another group (FP) was given training with a feature-positive discrimination, AX+ X-. On the final six sessions of conditioning, one daily test trial with each of A, B and X was delivered randomly intermixed with

the training trials of a regular conditioning session. The purpose of the experiment was to evaluate whether the presence of B would, to some extent, preserve X's excitatory status in the TD by comparison with the FP group. If this is the case, then X should overshadow A to a greater degree in the TD than the FP group, and A should accordingly elicit stronger conditioned responding in the FP than the TD group. Experiment 1 provided some preliminary evidence in support of this contention.

In the discussion to the experiment, I argued that the predictions leading to this pattern of results can be readily derived from the principles of standard associative theory, represented by the Rescorla-Wagner (1972) model. Whatever the merit of this account, the results in Experiment 1 are inconsistent with the theoretical interpretation of cue-competition advanced by the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988). According to this view, similar levels of performance to A should be observed in the TD and FP groups. When A is presented at test in either group, it will evoke a representation of the US and a representation of X, its comparator, which in turn will activate a representation of the US. The comparison process involving the two US representations, described in previous pages, will then determine the extent to which A elicits a conditioned response. Since the strength of both the A-X and X-US associations should be equivalent across the groups, performance to A should be dampened down to the same extent as a consequence of the comparison with X.

Furthermore, the Extended Comparator Hypothesis (Denniston et al., 2001; Savastano et al., 2003) does not appear any better apt to accommodate the present results. This recent version of the theory predicts that the degree to which the comparator stimulus will reduce responding in the presence of the target comes itself

under the control of other stimuli acting as comparators for the comparator. Thus, in group TD (AX+ BX-), stimulus B could serve as a second comparator for X, and hence moderate the extent to which the latter will reduce responding to A. Stimulus B's lack of associative strength, however, renders its entering such a chain of interactions highly unlikely. More importantly, the influence that B might exert in such circumstances would in any event lead to the reversal of the obtained pattern of results, that is, higher performance during A in the TD than the FP group. In summary, neither the original Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) nor its extended version (Denniston et al., 2001; Savastano et al., 2003) seem capable of accounting for protection from extinction when the CS to be protected is that which has driven the inhibitor to acquiring its properties. This conclusion should, however, be taken with caution, as a more trivial explanation for the present results demands consideration. This account relates to the perceptual masking that B might exert over X on BX trials for Group TD. Thus, it is conceivable that the effectiveness of nonreinforced trials with X was reduced by the presence of B. If this is correct, then X will not lose as much associative strength as it does by itself in Group FP, and as a consequence should prove a stronger comparator for A at the time of testing. Weaker performance to A should then follow in Group TD than Group FP on grounds quite different in nature to those advanced by the Rescorla-Wagner (1972) model. The additional assumption of perceptual interaction between B and X would enable the original comparator account (Miller & Schachtman, 1985; Miller & Matzel, 1988) to accommodate the present data equally well. The design of the Experiment 1 does not allow for this possibility to be ruled out and, accordingly, any conclusion regarding the fitness of the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) on this basis can be at best tentative.

In Experiment 2, single-phase blocking training consisting of intermixed, reinforced presentations with A and AX was given to one group of rats (group Blocking). The occurrence of blocking was judged by comparing performance to X in this group with that observed in two overshadowing control groups (groups Overshadow-2 and Overshadow), which received reinforced presentations with AX. According to the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988), with sufficient training A and X should reach similar levels of asymptotic associative strength in all three groups. Under these circumstances, the extent to which A attenuates responding to X by virtue of the comparison process postulated by the theory should be equivalent across treatments. It follows that, at asymptote, no difference in performance to X should be observed among the groups, or, in other words, that blocking should die away over training.

An indication that X must have attained asymptotic associative strength in all groups comes from the comparison of performances to X in the Overshadow-2 and Overshadow. The results from the test trials with X revealed that these groups were statistically indistinguishable from the first test session, suggesting a rapid acquisition function for X. The possibility of ceiling effects concealing actual differences in X's associative strength in the early stages of training is subsequently undermined by the sheer extensiveness of training (forty sessions). If the above analysis is correct, then the observation of equivalent performances to X in groups Overshadow-2 and Overshadow should have been accompanied by equally similar levels of performance to X in the Blocking group. Subjects in the latter group, however, approached the magazine recess significantly less in the presence of X on the final, fortieth test session, providing unequivocal evidence for blocking.

There is no obvious way to reconcile the principles of the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) with this belated emergence of blocking. In fact, such principles seem to bend more naturally towards predicting the reversal of blocking. According to the comparison process postulated by the theory, performance to X at test is directly related to X's associative strength and inversely related to the product of 1) the associative status of its comparator stimulus A, and 2) the strength of the X-A within compound association. On the assumption that the associative strengths of X and A converge over training into common asymptotic values across groups, the X-A within-compound association remains the sole source for asymptotic differences in responding to X among the groups. It is only natural to expect that this association will be weakest in the Blocking group, where A is presented in the absence of X on half the trials, relative to the Overshadow-2 and Overshadow groups, where neither stimulus is presented in each other's absence. To the extent that the A's deleterious influence as a comparator to X hinges on the strength of their within-compound association, the Blocking group should exhibit higher rates of responding to X at asymptote than either the Overshadow-2 and Overshadow groups. Since no independent measures were taken of the X-A within compound association, it is impossible to evaluate this prediction at this stage. At any rate, the persistence of blocking with continued training reported in Experiment 2 is in agreement with the analysis of cue-competition provided by the Rescorla-Wagner (1972) model. As noted previously, this type of account predicts that the deficit in responding to X can only increase as training with single-stage blocking progresses.

Overall, the results of the two experiments reported in this chapter have found scant support for the predictions of the comparator account, while being consistent with the tenets of the Rescorla-Wagner (1972) model.

Chapter 3

3. Further analysis of the Comparator Hypothesis (Miller & Matzel 1988; Denniston et al., 2001)

3.1. Introduction

Chapter 2 yielded some evidence against the account of cue competition put forward by the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988). In this chapter, a further attempt is made to evaluate this hypothesis by testing the prediction it makes regarding blocking (e.g. Kamin, 1968). The two experiments that follow examined the effects that adding the to-be-blocked stimulus (X) may have upon the conditioned properties of the stimulus responsible for blocking (A).

In the majority of blocking experiments, the focus of investigation has been the excitatory properties that stimulus X acquires during the trials with AX, as a consequence of conditioning with A. What has received considerably less attention, however, is the conditioned changes that take place to A as a consequence of conditioning with AX. The latter question acquires special relevance when one considers the rather different predictions concerning these changes that may be derived from the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) and the Rescorla-Wagner (1972) model.

As noted previously, according to the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) account of conditioning, stimuli do not compete with one another for an association with the US. Instead, stimuli acquire the same associative strength irrespective of whether they are conditioned in isolation or in compound with another stimulus (cf. Hebb, 1949). To explain the effects of A+

AX+ training, Miller and his colleagues (Miller & Schachtman, 1985; Miller & Matzel, 1988) have proposed that the retrieval of the association between the target stimulus, in this case A, and the US is modulated by the strength of the association between a comparator stimulus, X, and the US. More specifically, conditioned responding to a target stimulus is proposed to be a function of the associative strength of the target stimulus, less the product of (i) the association between the comparator stimulus and the US and (ii) the association between the target stimulus and the comparator stimulus. According to these principles, conditioning with A by itself will result in the formation of a strong association between A and the US. Interestingly, introducing trials in which AX signals the US should result in a loss of conditioned responding to A. This prediction follows because the strong associations that should form between A and X and between X and the US will result in X coming to serve as an effective comparator for A, thereby reducing the strength of conditioned responding to A. The Upper Panel of Figure 3.1 shows the output of a computer stimulation of the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) which is based upon the equations that were presented by Savastano et al. (2003)¹. Following the initial conditioning with A, in which conditioned responding reaches an asymptote equal to 1, introducing trials with AX+ should, according to the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel 1988), result in a decline in the strength of the CR evoked by A relative to stimulus B. This simulation employed the same parameters used by Savastano et al. (2003), thus the salencies of A B, X and the US were all set to 1 and the learning rate parameters for reinforcement (p1) and nonreinforcement (p2) were set to 0.2 and 0.013, respectively, however, it should be noted that simulations conducted with other

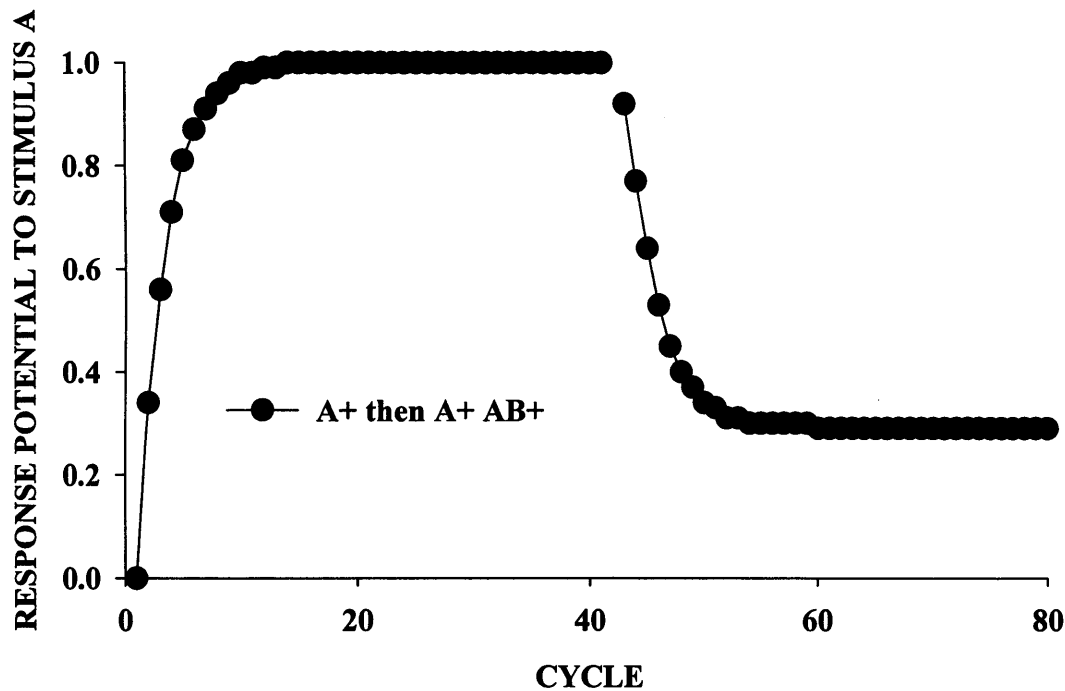
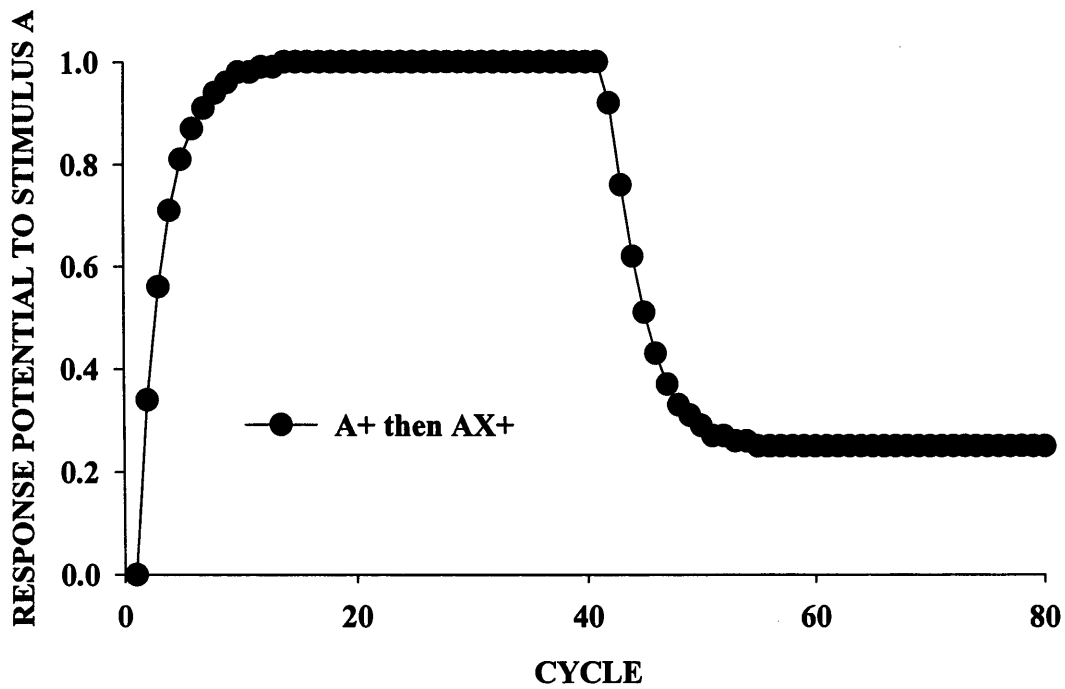


Figure 3.1. Computer simulations of the level of responding in the presence of cue A during A+ trials followed by AX+ trials (Upper Panel) and during A+ trials followed by A+ and AX+ trials (Lower Panel) as predicted by the Comparator Hypothesis, based on the equations provided by Savastano et al. (2003).

parameter values have led to similar outputs. The significance of the simulation that appears in the Lower Panel of Figure 3.1 will be described shortly.

A rather different prediction about the outcome of adding X to A for an A+ AX+ blocking treatment can be derived from the proposals of Rescorla and Wagner (1972). According to this account, conditioning with A by itself will result in the development of an association between A and the unconditioned stimulus (US) which, if sufficient training is given, will reach an asymptote equal to the total amount of conditioning that can be supported by the US (λ). Adding X to A, and pairing this compound with the US, should result in cue competition so that X acquires little or no association with the US and hence the associative strength of AX should, like that of A, equal λ . The implication is that the conditioned properties of A should remain the same throughout this stage: adding X to A should leave A with the same associative strength as a stimulus that is simply paired, in isolation, with the US.

A number of experiments have tested A following conditioning trials with A+ and then AX+. However, for a number of reasons, these studies do not allow us to draw any firm conclusions about whether this training undermines the strength of the CR during A, or leaves it unaffected. Rescorla (1999) reports an appetitive Pavlovian conditioning experiment in which rats initially received trials in which A and B were paired separately with food; then in Stage 2, a compound of A and X was paired with food. In a subsequent test session, there was no difference in the rate of responding to A and B, a result which, at face value, seems to support the view of conditioning provided by the Rescorla-Wagner (1972) model. However, this result can be accommodated by the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) if it is assumed that conditioning with A and B did not reach asymptote in Stage 1. If this were the case then the detrimental effects of adding the comparator,

X, in Stage 2 could have been offset by the additional learning about A that takes place. This would result in conditioned responding to A and to B (which was not presented at all during Stage 2) being rather similar during the final test session, which is exactly what was observed.

In contrast to these results, an experiment by Arcediano, Escobar and Miller (2004, see also: Hall, Mackintosh, Goodall & Dal Martello, 1977) seems to provide support for the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988). In this experiment, rats were first given trials in which a stimulus, A, signalled a shock. Following this training, the animals were split into two groups. For the experimental group, AX was paired with shock, whilst animals in the control group received pairings of BX with shock. In a final test session A evoked more conditioned responding in the control group than in the experimental group, a result which is compatible with the predictions that can be derived from the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988), but not more standard associative models (e.g. Rescorla & Wagner, 1972). Unfortunately, the results reported by Arcediano et al. (2004) may provide, at best, ambiguous support for the Comparator Hypothesis ((Miller & Schachtman, 1985; Miller & Matzel, 1988). This study failed to control for differences in how recently and frequently A had been presented between the experimental groups. It is possible, therefore, that the differences observed between the experimental and control groups were a consequence of, for example, differences in the strength of the CR evoked in the presence of A.

Another explanation for the results from the study by Arcediano et al. (2004) can be derived by referring to stimulus generalisation. It is possible that conditioning with A was close to asymptote by the end of the first stage of training. During the

second stage, X would gain little associative strength in the experimental group, because of blocking by A, whereas both A and X would be expected to gain a reasonable measure of associative strength in the control group. Since A, B and X were all auditory, it is then conceivable that responding during A was stronger in the control group, because it received generalised excitation from two strong sources, than the experimental group, where generalisation from only one weak source was possible (e.g. Rescorla, 1976).

To complicate matters further, an experiment reported by Pearce and Redhead (1999) demonstrated that adding X to A can actually enhance responding to A. In their Experiment 1, two groups of rats were initially given training in which stimulus A was paired with a US. In Stage 2, the experimental group were given paired presentations of AX with a US, whilst the control group continued to receive paired presentations of A with a US. The results from the final test session did reveal a difference in responding to A between the two groups. However, the results were in the opposite direction to that predicted by the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988): conditioned responding to A was higher in the experimental group than in the control group. Unfortunately, this result does not necessarily trouble either the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) or the Rescorla-Wagner (1972) model. Pearce and Redhead (1999) changed the US between Stages 1 and 2 from sucrose solution to food pellets. Without being able to specify exactly how to represent a qualitative change in the US within simulations of either the Rescorla-Wagner (1972) model or the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988), it is difficult generate precise predictions for the outcome of this experiment. These results must remain, therefore, at least provocative, but unfortunately not decisive.

The purpose of the two experiments reported in this chapter, therefore, was to examine the impact of conditioning with an AX compound, following prior conditioning of A, upon the CR evoked by A. This training should, according to the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) have a disruptive effect upon the CR evoked by A, whilst according to the Rescorla-Wagner (1972) model, this training should have no effect upon conditioned responding to A.

3.2. Experiment 3

For Stage 1 of the experiment, rats received appetitive Pavlovian conditioning in which two stimuli, A and B, were paired with food (see Table 3.1). In Stage 2, rats continued to receive pairings of A and B with food, but in addition they also received trials in which a compound of A and X was paired with food. Following this training the animals were presented with test trials with A and B, conducted in extinction. To encourage subjects to differentiate between the experimental stimuli, nonreinforced trials with Y were also included in the first two stages of the experiment.

Table 3.1. Design of Experiment 3.

Stage 1	Stage 2	Test
A+ B+ Y-	A+ AX+ B+ B+ Y-	A- B-

If conditioning in Stage 1 reached asymptote, then the account of conditioning provided by Rescorla and Wagner (1972) predicts that the inclusion of trials with AX in Stage 2 should have no impact upon conditioned responding to A. Therefore, the strength of the CR during A should be equal to that during B. According to the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel 1988), however, the introduction of trials with AX should provide A with an effective comparator stimulus, X, which is also paired with the US. Consequently, there should be a reduction in the strength of the CR during A, but not B, during Stage 2 despite the fact that both of these stimuli are paired with food.

In a conventional blocking experiment, subjects receive conditioning with A followed by conditioning with AX. I decided to adopt the rather different strategy of presenting intermingled trials with A and AX during the blocking stage in order to allow a more detailed test of the predictions made by the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) concerning the experiment. The hypothesis predicts that after the introduction of the trials with AX, there will be a gradual decline in responding to A. It would not be possible to observe this decline during AX trials, because responding will be determined by the properties of both stimuli. However, by occasionally presenting A in the absence of X, it should be possible to observe the impact of the AX trials on responding to A. The Lower Panel of Figure 3.1 shows the results from a computer simulation based on the same principles as the previous simulation in order to determine formally the predictions that follow from the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) for my experimental design. The only difference between the two simulations is that during the second stage of training of the second simulation there were an equal number of trials with A and AX. Despite this change, the simulation

reveals that the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) continues to predict that pairing A with X will result in a loss of responding to A, which should become more evident as training in the second stage of the experiment continues.

The experiment failed to reveal any evidence of a decline in responding to A, relative to that during B, as a consequence of the conditioning trials with AX. One explanation for this outcome is that responding during A was so vigorous that the measure of the strength of conditioning responding was not sufficiently sensitive to detect a modest drop in the propensity to respond to this stimulus. To take account of this possibility, the experiment concluded with a series of extinction trials with A and B, in the hope that differences in performance during these stimuli might be detected when responding during each of them was relatively weak.

Method

Subjects. The subjects were 32 experimentally naive, male, hooded Lister rats that were maintained in an identical fashion to the subjects in Chapter 2.

Apparatus. All details of the apparatus were the same as in Chapter 2, except for the following. Each conditioning chamber had a single, 2.8-W lamp covered by 1.5-cm diameter plastic disc located upon the roof of the magazine. Loudspeakers located in the ceiling could be used to present broadband white noise, a 4-Hz train of clicks and a 100-Hz buzz. All auditory stimuli were approximately 10 dB (scale A) above the background noise level.

Procedure. The details of magazine training were identical to those in Chapter 2. During Stage 1 there were 10 sessions in which A and B were paired with food and Y was not. In each session there were 8 trials with A, 8 with B and 4 with Y. Each of these stimuli was presented for 10 s and, following trials with A and B, a single food pellet was delivered into the magazine. There were no programmed consequences following trials with Y. The mean inter-trial interval (ITI), defined as the duration between the termination of one trial and the onset of the next was 3 min (range: 2-4 min). The stimuli that served as A and B were counterbalanced as the onset of the magazine light, and white noise. The stimulus that served as X was counterbalanced with Y as either a buzz or a train of clicks. The trial sequence was random with the constraint that no more than 2 trials of the same type could occur in succession.

During Stage 2 there were 8 sessions in which A, B and AX were paired with food and Y was not. In each session there were 5 presentations each of Y, A and AX. To equate the number of times that A and B were each paired with food there were 10 presentations of B in each session. Following trials with A, B and AX a single food pellet was delivered into the magazine, there were, once again, no programmed consequences following trials with Y. The details of the ITI and trial sequencing were the same as in Stage 1.

For the final test session, A, AX and Y were each presented once and B was presented twice. Food was presented following trials with A, AX and B, but not after Y. Following these trials there were 10 nonreinforced trials each with A and with B. The details of the ITI and trial sequencing were the same as in Stage 1.

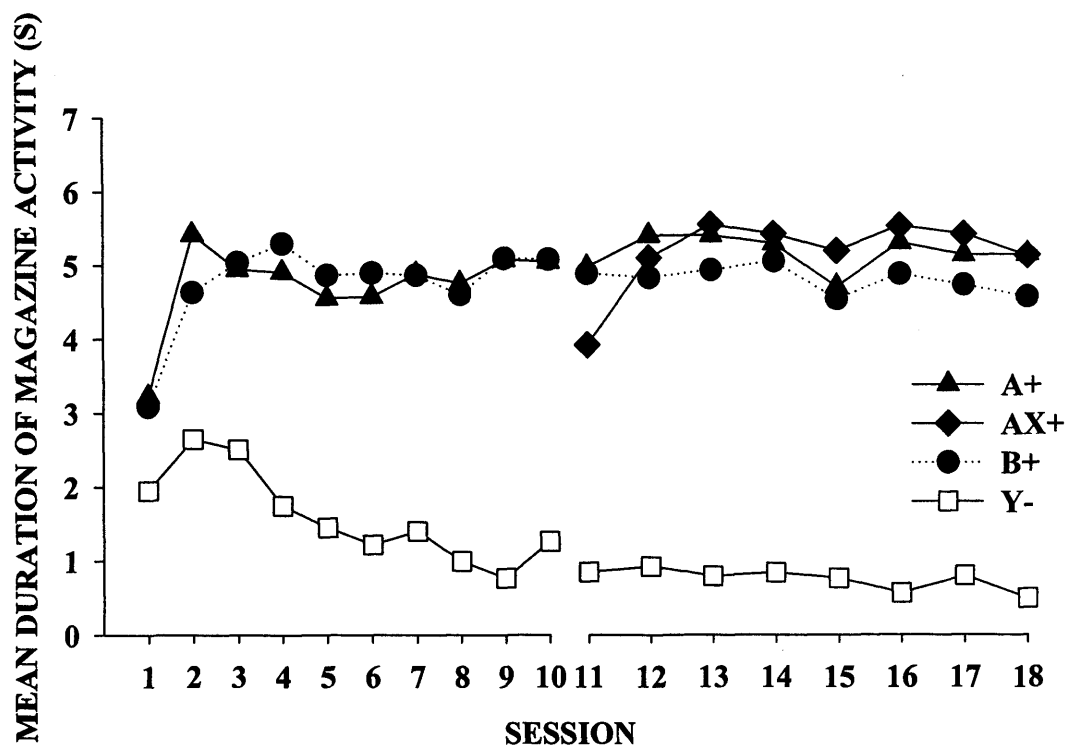


Figure 3.2. Acquisition in Experiment 3.

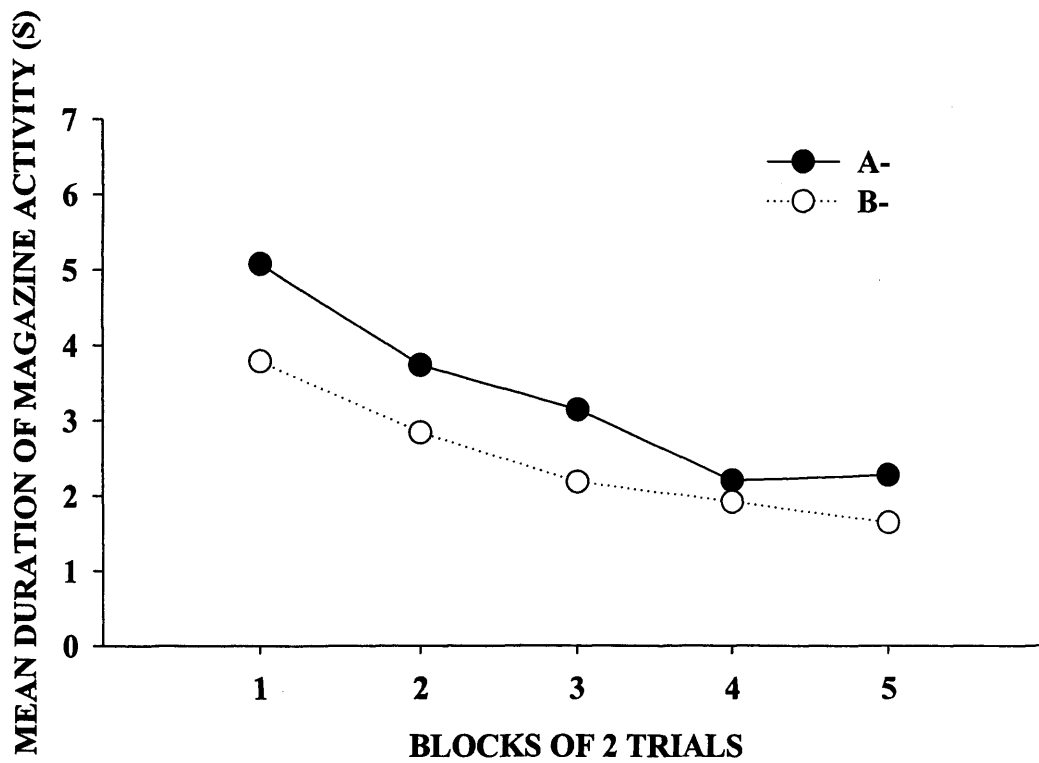


Figure 3.3. Test with A and B in Experiment 3.

Results. The results of the training from Stage 1 are shown on the left-hand side of Figure 3.2. Conditioning proceeded smoothly, and by session 10 there was a clear discrimination between A and Y and between B and Y. This observation was confirmed by a 1-way Analysis of Variance (ANOVA) of individual mean durations of magazine activity to A, B and Y during session 10, $F(2, 62) = 47.35$. Paired t-tests corrected according to the Bonferroni procedure revealed that responding to A and B was higher than to Y, but there was no difference between A and B.

The results of the training from Stage 2 are shown on the right-hand side of Figure 3.2. The discrimination between the stimuli that were reinforced (A, B and AX) and nonreinforced (Y) was evident throughout this stage. Adding X to A resulted in a slight disruption of conditioned responding during Session 11, and there was an indication that conditioned responding to A was stronger than to B throughout this stage. A 2-way ANOVA of individual mean durations of magazine activity with the factors of stimulus and session revealed an effect of session, $F(7, 217) = 2.46$, an effect of CS, $F(3, 93) = 89.51$, and a Session x CS interaction, $F(21, 651) = 3.35$. Analysis of simple main effects revealed significant effects of CS on each session, $F_s(3, 744) > 50.66$. Paired t-tests, again corrected according to the Bonferroni procedure, revealed that responding to A, B and AX was higher than to Y on each session. In addition, responding to AX was significantly weaker than to A on Session 1. No further comparisons were significant.

The results of the final test session, in which A and B were presented in isolation and nonreinforced are shown, in 2-trial blocks, in Figure 3.3. The numerical difference between A and B that was seen in the Stage 2 was present during the test session and was, if anything, enhanced. This observation was confirmed with a 2-way ANOVA of individual mean duration scores with the factors of CS and trial-block

which revealed an effect of CS, $F(1, 31) = 4.26$, confirming that responding to A was stronger than to B, and an effect of trial-block, $F(4, 124) = 44.20$. The CS x Trial-block interaction was not significant, $F(4, 124) = 2.01$.

Discussion. In Stage 1, rats received trials in which stimuli A and B were paired with food. During Stage 2 this training continued, but half of the trials with A were replaced with trials in which AX was paired with food. In a subsequent test session, A and B were presented in extinction and conditioned responding to A was found to be stronger than to B. These results do not support the predictions of the Comparator Hypothesis (Miller & Matzel 1988; Denniston et al., 2001), according to which conditioned responding to B should have been stronger than to A.

3.3. Experiment 4

Experiment 3 demonstrated that following separate conditioning with A and B, conditioning trials with A, B and AX resulted in stronger conditioned responding to A than to B. According to the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988), a consequence of conditioning AX in Stage 2 should be that X will become a comparator for A, resulting in a weakening of the conditioned response to A, relative to B. The results of Experiment 3 therefore failed to confirm this prediction. This conclusion, however, rests upon the assumption that X did in fact become a comparator for A. Granting this assumption requires one to accept two further assumptions: first, that an association was formed between A and X; and

second, that X acquired an association with the US. Unfortunately, Experiment 3 provided no evidence for either of these two assumptions, and furthermore there are grounds for believing that neither of them is warranted. On the one hand, Rescorla and Cunningham (1978; see also Rescorla & Freberg, 1978) have shown that within-compound associations are weakened by presenting one of the elements of the compound in isolation. Thus, the A+ trials in both stages 1 and 2 of Experiment 3 would be expected to weaken the growth of an A-X association in stage 2. On the other hand, there is evidence which suggests X will acquire almost no associative strength when conditioning trials with A precede conditioning trials with AX (e.g. Kamin, 1968, 1969). Without any evidence for an association between A and X, or between X and the US, the results of Experiment 3 do not necessarily provide a challenge to the theory proposed by Miller and his colleagues (Miller & Schachtman, 1985; Miller & Matzel 1988), because these associations are essential if the comparator process is to weaken responding to A.

The purpose of Experiment 4, therefore, was to investigate whether or not these foregoing assumptions were indeed warranted. The first two stages of Experiment 4 were identical to Stages 1 and 2 from Experiment 3: A and B were initially paired with the US and Y was nonreinforced, and this training continued in Stage 2, again with the addition of AX+ trials. Following this training, the rats were split into two groups for a third stage of training. Rats in Group X received nonreinforced trials with X and rats in Group Y received nonreinforced trials with Y. All animals were then given a series of nonreinforced test trials with A and B. A summary of the design of Experiment 4 is shown in Table 3.2.

If X acquired no associative strength during the conditioning trials with AX in Stage 2 then there should be no difference in the strength of conditioned responding to X and

to Y during the extinction trials in Stage 3. If, however, X acquired even some associative strength then conditioned responding should be higher to X than to Y, and we will have some evidence to support the first assumption made above. If the training in Stage 2 resulted in no association forming between A and X, then extinguishing X in Stage 3 for Group X should have no effect on the conditioned response that will be evoked by A when it is subsequently presented in a test session: responding to A should still be greater than responding to B. Failure to confirm this prediction will imply some evidence to support the second assumption.

Table 3.2. Design of Experiment 4.

Stage 1	Stage 2	Stage 3	Test
A+ B+	A+ AX+ B+ B+	X-	A- B-
Y-	Y-	Y-	A- B-

Method

Subjects and Apparatus. The subjects were 32 experimentally naive, male, hooded Lister rats that were maintained in an identical fashion to the subjects in Experiment 3. The apparatus was the same as in Experiment 3.



Procedure. Stages 1 and 2 of the current experiment were identical in all respects to Stages 1 and 2 of Experiment 3. Thus, animals first received 10 sessions in which A and B were paired with food and Y was not. There were subsequently 8 sessions in which A, B and AX were each paired with food and Y was again nonreinforced.

Following this training the animals were split into two groups and were given two sessions of extinction with either X or Y. For the first of these sessions, the first 5 trials comprised two training trials with B and one with each of A, AX and Y. Group X then received 16 nonreinforced trials with X and subjects in Group Y received 16 trials with Y. On the following day subjects in Group X and Y received 20 nonreinforced trials with X and Y respectively.

For the final, test, session all rats received 10 trials each with A and B. The details of this session were identical to the final test session of Experiment 3, except that no warm-up trials preceded the trials with A and B. Any procedural details omitted from this Experiment were the same as for Experiment 3.

Results. The results of the training from Stage 1 are shown on the left-hand side of Figure 3.4. Conditioning proceeded in a similar fashion to Experiment 3 and by session ten there was a discrimination between A and Y and between B and Y. This observation was confirmed by a 1-way Analysis of Variance (ANOVA) of individual mean durations of magazine activity to A, B and Y during session ten, $F(2, 62) = 251.96$. Paired t-tests corrected according to the Bonferroni procedure revealed that responding to A and B was higher than to Y, but there was no difference between A and B.

The results of the training from Stage 2 are shown on the right-hand side of Figure 3.4. In keeping with the results of Experiment 3, the discrimination between the stimuli that were reinforced (A, B and AX) and nonreinforced (Y) was evident throughout this stage and adding X to A resulted in a slight disruption of conditioned responding during session 11. A two-way ANOVA of individual mean durations of magazine activity with the factors of stimulus and session revealed an effect of CS, $F(3, 93) = 203.00$, but no effect of session, $F < 1$. There was however, a significant Session x CS interaction, $F(21, 651) = 5.57$. Analysis of simple main effects revealed an effect of session for AX, $F(7, 868) = 7.97$, which again reflects the disruption of conditioned responding on Session 11, but no effects of session for any of the other stimuli, $F_s(7, 868) < 1.90$. Furthermore, there were significant effects of CS on each session, $F_s(3, 744) > 86.91$. Paired t-tests, again corrected according to the Bonferroni procedure, revealed that responding to A, B and AX was higher than to Y on each session and that in addition, responding to AX was significantly weaker than to A on Session 11. No further comparisons were significant.

The results of Stage 3, in which X and Y were nonreinforced for, respectively, Group X and Group Y are shown in two-trial blocks in Figure 3.5. The mean duration of magazine activity to X in Group X was, at first, greater than the mean duration of magazine activity to Y in Group Y, suggesting that X did acquire an association with the US in Stage 2. However, the effects of nonreinforcement eventually resulted in a reduction in responding to X, so that by the end of this stage, the mean durations of magazine activity to X and Y were at a similar, low, level. These observations were confirmed with a two-way ANOVA of individual mean durations of magazine activity with the factors of Group (X vs. Y) and two-trial block, which revealed significant effects of group, $F(1, 30) = 13.67$, session, $F(17, 510) = 6.26$ and a

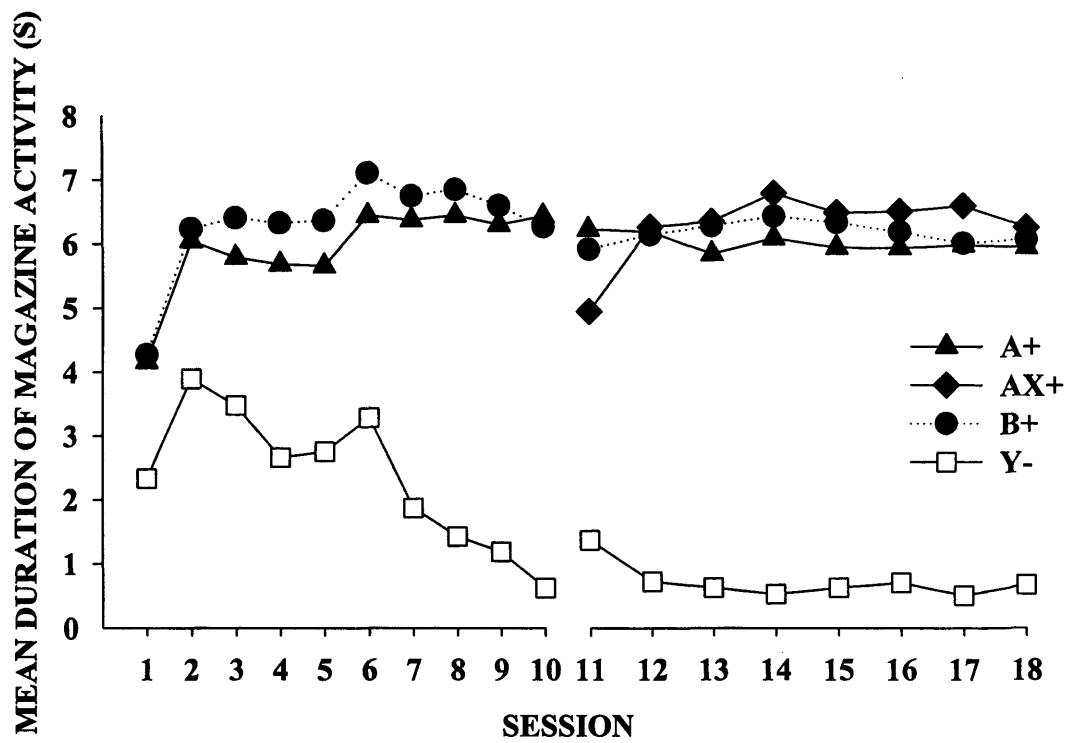


Figure 3.4. Acquisition in Experiment 4.

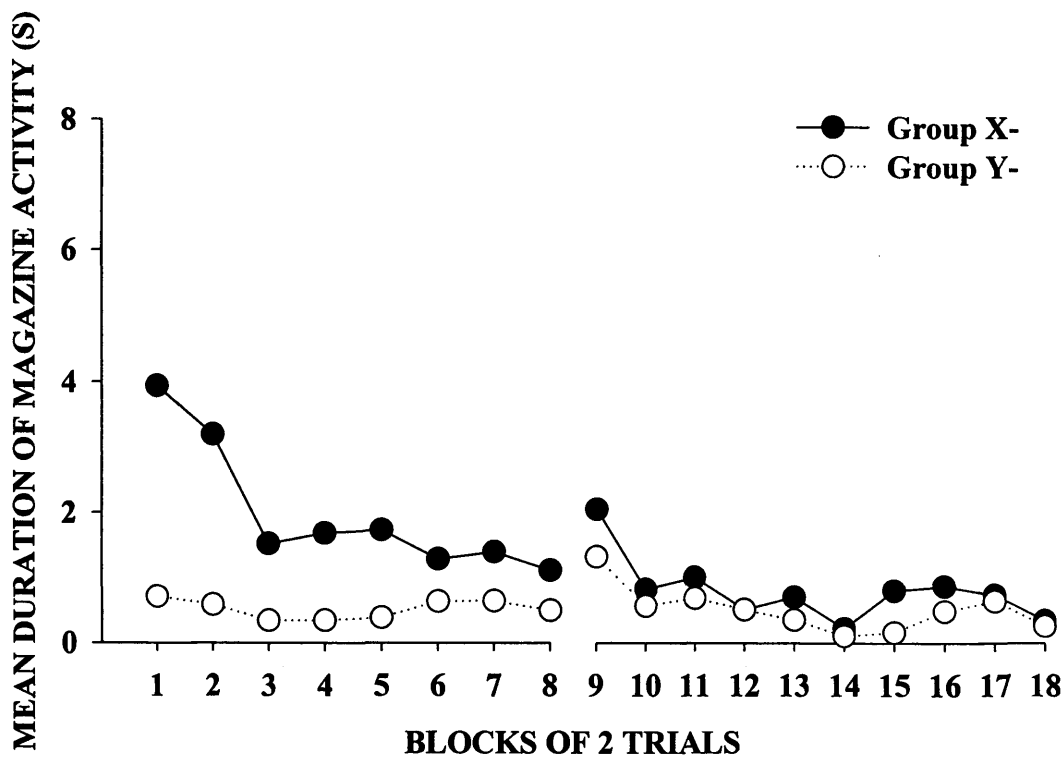


Figure 3.5. Extinction of X and Y in Experiment 4.

significant Group x Session interaction, $F(17, 510) = 3.92$. Analysis of simple main effects revealed an effect of trial-block for Group X, $F(17, 510) = 9.44$, but not for Group Y, $F < 1$, confirming the detrimental effects of nonreinforcement upon X for Group X. Furthermore, conditioned responding to X in Group X was higher than to Y in Group Y on trial-blocks 1 to 5, $F_s(1, 540) > 5.93$, but not on any of the remaining trial-blocks, $F_s(1, 540) < 2.42$.

The results of the final stage in which test trials with A and B were given to both Groups X and Y are shown, respectively, in the Upper and Bottom Panels of Figure 3.6. The results from Group Y reveal that the mean durations of magazine activity to A were greater than to B, a result which replicates the results of Experiment 1. However, this difference was abolished in Group X. With the exception of the first 2-trial-block, there was no indication of any difference in the mean durations of magazine activity to A and B. These observations were confirmed by a three-way ANOVA of individual mean durations of magazine activity with the factors of group (Group X vs. Group Y), CS (A vs. B) and two trial-block, which revealed a significant three-way interaction among these factors, $F(4, 120) = 2.60$. Simple main effects analysis of this interaction revealed a significant CS x Trial-block interaction for Group Y but not for Group X. Further simple effects analysis revealed that, crucially, responding to A was higher than to B on trial-blocks 2 and 3 for Group Y, $F_s(1, 150) > 5.53$, but not on trial blocks 1, 4 and 5, and on none of the trial blocks for Group X, $F_s(1, 150) < 3.29$. Other results from the simple effects analysis revealed a Group x CS interaction on trial-block 2, $F(1, 150) = 3.90$, and a Group x CS interaction that just missed significance on trial-block 3, $F(1, 150) = 3.79$, $p=0.054$. Further simple effects analysis of the interaction on trial-block 3 revealed a difference between Groups X and Y in their conditioned responding to A, $F(1, 300) =$

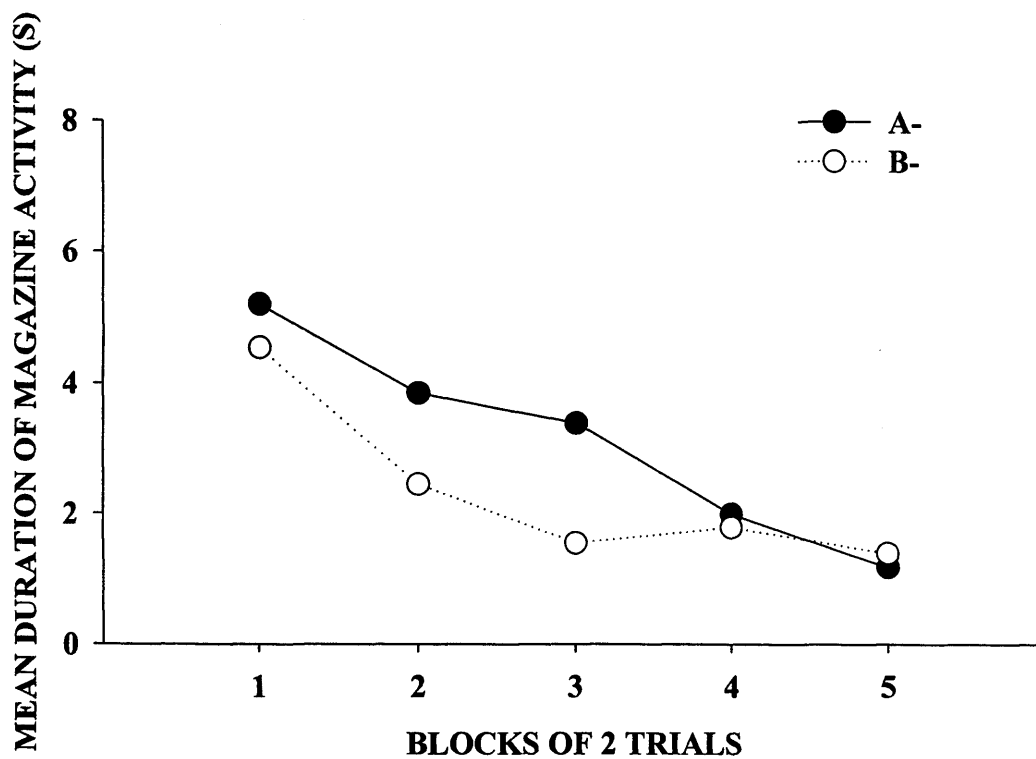
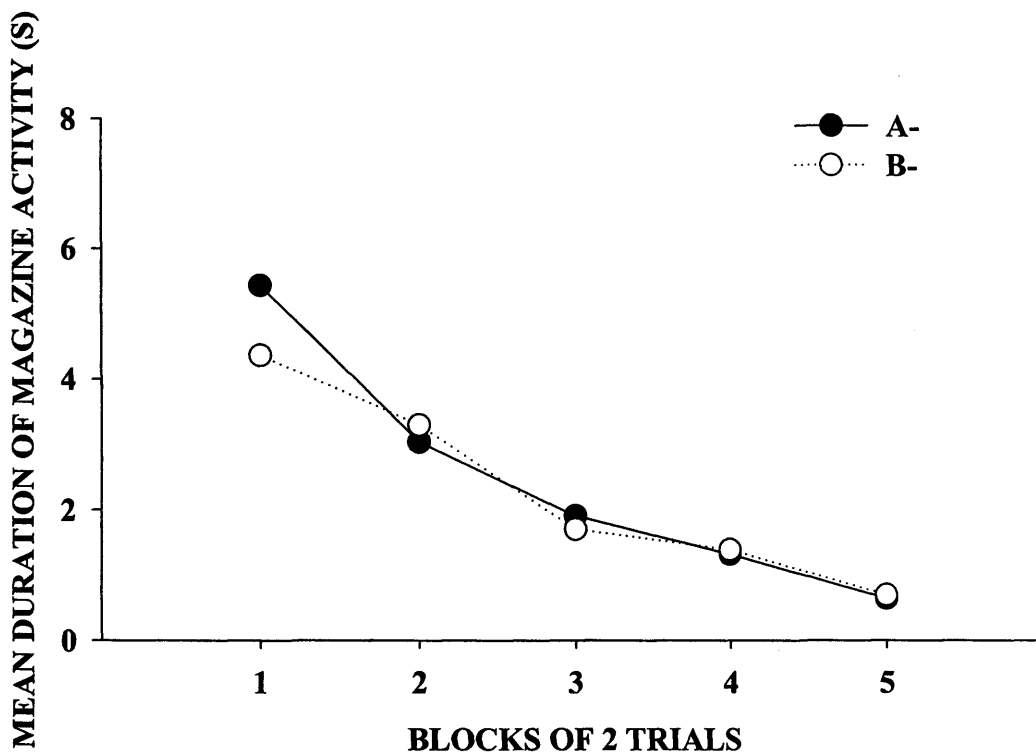


Figure 3.6. Conditioned responding in the presence of A and B in Group X (Upper Panel) and Group Y (Lower Panel) in Experiment 4.

5.36, but not to B, $\underline{F} < 1$. Other results from the overall ANOVA were an effect of trial-block, $\underline{F}(4, 120) = 65.58$ and a Trial-block x CS interaction, $\underline{F}(4, 120) = 2.64$. None of the remaining effects or interactions were significant, $\underline{F}_s(1, 30) < 2.26$ and $\underline{F}(4, 120) < 1$.

Discussion. The purpose of Experiment 4 was twofold. First, it was intended to provide a replication of the results of Experiment 3. The results from Group Y in the current experiment confirmed the reliability of the effect seen in Experiment 3: following conditioning with A, B and AX, conditioned responding to A was higher than to B. Second, it was intended to demonstrate that conditioning with A and AX resulted in the formation of associations between X and the US, and between A and X. Test trials conducted in extinction following the conditioning stages revealed that conditioned responding to X was higher than to Y, suggesting X did have associative strength, and furthermore that extinction of X abolished the superior conditioned responding to A than to B. This final result is difficult to explain without assuming the existence of an association between A and X. According to the proposals of Miller and his colleagues (Miller & Schachtman, 1985; Miller & Matzel, 1988) these associations should result in a weakening of the conditioned response evoked by A relative to B, which has no effective comparator stimulus. The results, however, indicate the opposite: with these associative structures in place, conditioned responding to A is stronger than to B.

It is possible that the stronger conditioned response seen to X relative to Y in Stage 3 of the current experiment did not, in fact, reflect differences in the associative strengths of these stimuli. It is conceivable that X had no associative strength of its

own at all, and instead could only evoke a conditioned response by virtue of activating a representation of the US through its within-compound association with A. Although this alternative is plausible, it seems unlikely. If X had no associative strength at all, why then would conditioned responding to A be stronger than to B in the first place? Another possibility, which is more difficult to refute, is that responding to X was greater than to Y not because it had greater associative strength, but because X was less familiar than Y. Before the extinction trials with X and Y in Stage 3, Y had been presented to each animal 80 times, whereas X had been presented only 40 times. Perhaps magazine activity was greater to X than to Y because this behaviour had had less of an opportunity to habituate (Thompson & Spencer, 1966). Although I cannot refute this possibility with absolute certainty, I do view it with caution for it is possible to make a comparison of the mean durations of magazine activity to X and Y following equal amounts of exposure to these stimuli. For the current experiment this necessitates a comparison of the mean durations of magazine activity to Y for the animals that would go on to be in Group Y on session 11 (1.33 s) with the mean duration of magazine activity to X in Group X during the first 5 trials with X in Stage 3 (3.23 s). The difference between these means was significant, $t(30) = 2.98$, suggesting that X did in fact acquire an association with the US.

One shortcoming with the design of the present experiment is that Group X received extinction trials with an excitatory stimulus, and Group Y received extinction trials with a neutral stimulus. It is possible that a within-compound association did not develop between A and X and that for some unspecified reason nonreinforced exposure to an excitor weakened responding to A to a greater extent than nonreinforced exposure to a neutral stimulus. Perhaps, for example, the effects of nonreinforcement with X generalised to some extent to A. Such an effect would not

be expected in Group Y for which the effects of extinction with Y would be minimal. A weakness with this explanation is that it is not clear why the effects of extinction with X did not also generalise to B in Group X and result in the difference between responding to A and B being sustained. Whatever the merits of this discussion, the facts remains that the conditioning trials with AX augmented responding to A, and this result is contrary to that predicted by the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988).

3.4. General Discussion

In Experiment 3, rats first received trials in which A and B were paired with a US while Y was not. Subsequently, A, B and AX were paired with the US and, on a final test session, conditioned responding was higher to A than to B. The results of Experiment 4 point to the associative properties of X as the source of this effect. Extinction tests revealed that conditioned responding was stronger to X than to the nonreinforced Y, implying that X acquired associative strength during Stage 2, and furthermore, that extinction of X resulted in an equivalent level of conditioned responding to A and B. A natural way of interpreting these results is to appeal to the formation of a within-compound association between A and X during Stage 2. As a consequence of this association, presenting A during the test should result in the activation of the US representation through A's own direct association with the US, as well as indirectly, through A's association with X. The consequence of having the US representation being activated by two sources should be summation, and consequently, an enhancement of conditioned responding. This being the case, then

the current results would differ from more conventional demonstrations of summation (e.g. Pearce, George & Aydin, 2002; Rescorla, 1997) in which performance is enhanced by the combined effects of two physically present CS; the results of Experiments 3 and 4 imply that an analogous result can be observed when a physically present CS is presented in conjunction with an associatively activated one (see: Rescorla, 1980).

These experiments are by no means the first to demonstrate the existence of within-compound associations during a blocking procedure. In a conditioned flavour-aversion experiment reported by Speers, Gillan and Rescorla (1980) rats first received paired presentations of flavours A and B with illness and then in Stage 2, compounds of AX and BY, both of which were again paired with illness. To examine if a within-compound association had developed between A and X and between B and Y, A was paired with illness, whereas B was not; a choice test was then given between X and Y, which revealed a stronger aversion to X than to Y. These results are clearly compatible with the results of Experiment 3, and in particular, Experiment 4. Where the current experiments are novel, however, is in demonstrating that a within-compound association can facilitate the conditioned response evoked by another stimulus. This class of effect (potentiation) is typically observed in stimuli that are of low salience, and which by themselves are poor at supporting conditioning (e.g. Galef & Osborne, 1978; Durlach & Rescorla, 1980). For the current experiments, however, responding was enhanced in a stimulus that had itself been used to support effective, and indeed asymptotic, Pavlovian conditioning. If there is any merit to the preceding analysis, therefore, these results can be said to extend the conditions under which potentiation occurs.

Before concluding that these results constitute a demonstration of potentiation, an account based on the notion of representation-mediated extinction (Holland & Forbes, 1982; Ward-Robinson & Hall, 1996) should be considered. In the first report of this phenomenon, Holland and Forbes (1982) initially gave rats in the experimental group compound presentations of a tone and sucrose. They then paired sucrose with lithium chloride before going on to present the tone by itself with no consequences. At test, rats in this group showed less of an aversion to sucrose than those in a control group in which the initial tone-sucrose pairings were omitted. Holland and Forbes (1982) argued that nonreinforced presentations of the tone in the experimental group should evoke the representation of sucrose. The activation of such representation in the absence of the US (lithium chloride) could thus account for the extinction of the previously conditioned aversion to this flavour.

It seems thus sensible to ask whether a similar process of mediated extinction could not at least in part explain the results reported in this chapter. According to this interpretation, during extinction of the added cue X (Experiment 4) the representation of its associate A is activated in the absence of the US, resulting in the (partial) extinction of the A-US association. This explanation would seem particularly plausible if the CR evoked by A following extinction of X had been *less* vigorous than that recorded in the presence of B. That this was not so casts some doubts over the merit of this sort of analysis in the circumstances at hand. Indeed, what the experiments in this chapter show is that extinction of the added cue X returns —after having boosted them— the excitatory properties of A to a level comparable to those of the control cue B. It appears more reasonable, therefore, to interpret the outcome of extinguishing X as that of subtracting the potentiated component from A's CR than as

that of simply extinguishing an otherwise unpotentiated CR through mediated extinction.

In any event, the results of the current experiments are contrary to the predictions that can be derived from the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988). According to this account, introducing conditioning trials with AX in Stage 2 should result in associations forming between X and the US, and between A and X. These associations should ensure that the conditioned response elicited by A will be weakened. By contrast, our results indicated that as a consequence of this training, conditioned responding to A was enhanced. The results of Experiment 4 provided a replication of the results of Experiment 3 as well as some evidence for associations between X and the US, and between A and X. Thus, the presence of these associations does not appear to influence conditioned responding to a target stimulus—in this case A—in the way envisioned by Miller and his colleagues (Miller & Schachtman, 1985; Miller & Matzel, 1988). I now turn to analyse the significance of these results for the extended version of the theory (Denniston et al., 2001; Savastano et al., 2003).

The predictions derivable from the Extended Comparator Hypothesis (Denniston et al., 2001; Savastano et al., 2003) do not differ in essence from those of the original formulation (Miller & Schachtman, 1985; Miller & Matzel, 1988). It may be timely to recall that this more recent version of the theory incorporates the assumption that the target-comparator and comparator-US associations are themselves modulated by comparator processes involving higher-order comparator stimuli. One obvious candidate to serve as a higher-order comparator in the present case is the experimental context. Taking the context (C) into account, the training regime administered in Experiments 3 and 4 would thus be AC+ BC+ ACX+ C-. According

to the Extended Comparator Hypothesis (Denniston et al., 2001; Savastano et al., 2003), on AC+ trials C should serve as a comparator for A's comparator X, thereby reducing the extent to which the latter is in turn able to reduce responding during AC+ trials. As a consequence, the effect of the context will be a recovery of the strength of responding during AC+ trials, i.e. an attenuation of blocking. The question of interest is whether, in addition to anticipating a certain recovery from blocking during A+ trials, this extended version of the theory can go as far as to predict that following training with A+ and B+, conditioning with A+, B+ and AX+ will result in stronger conditioning to A than to B. Simulations of the Extended Comparator Hypothesis (see Figure 3.7) conducted (again using the equations provided by Savastano, et al., 2003) with the assumption that the experimental context can engage in the comparator process, are unable to generate this prediction. Like the predictions derived from the original Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988; see Figure 1 in Chapter 1), these simulations revealed that the CR in the presence of B should be stronger than that in the presence of A. It should be noted, however, that Miller and his colleagues (e.g. Arcediano et al., 2004, p. 1110; Stout et al. 2003, p. 91; Wheeler & Miller, 2005. p. 475) have on a number of occasions explicitly ruled out the contribution of the experimental context as a comparator stimulus in experiments in which multiple discrete cues are conditioned in compound. This position is justified by the low salience usually attributed to the context and the extinction it is supposed to suffer during the inter-trial interval, making its contribution minimal as a higher-order comparator. All the same, even in the extreme case in which the salience of the context is assumed to be maximal, the extended version of the theory can only at best predict that no difference in responding in the presence of A and B will follow, as shown in Figure 3.7.

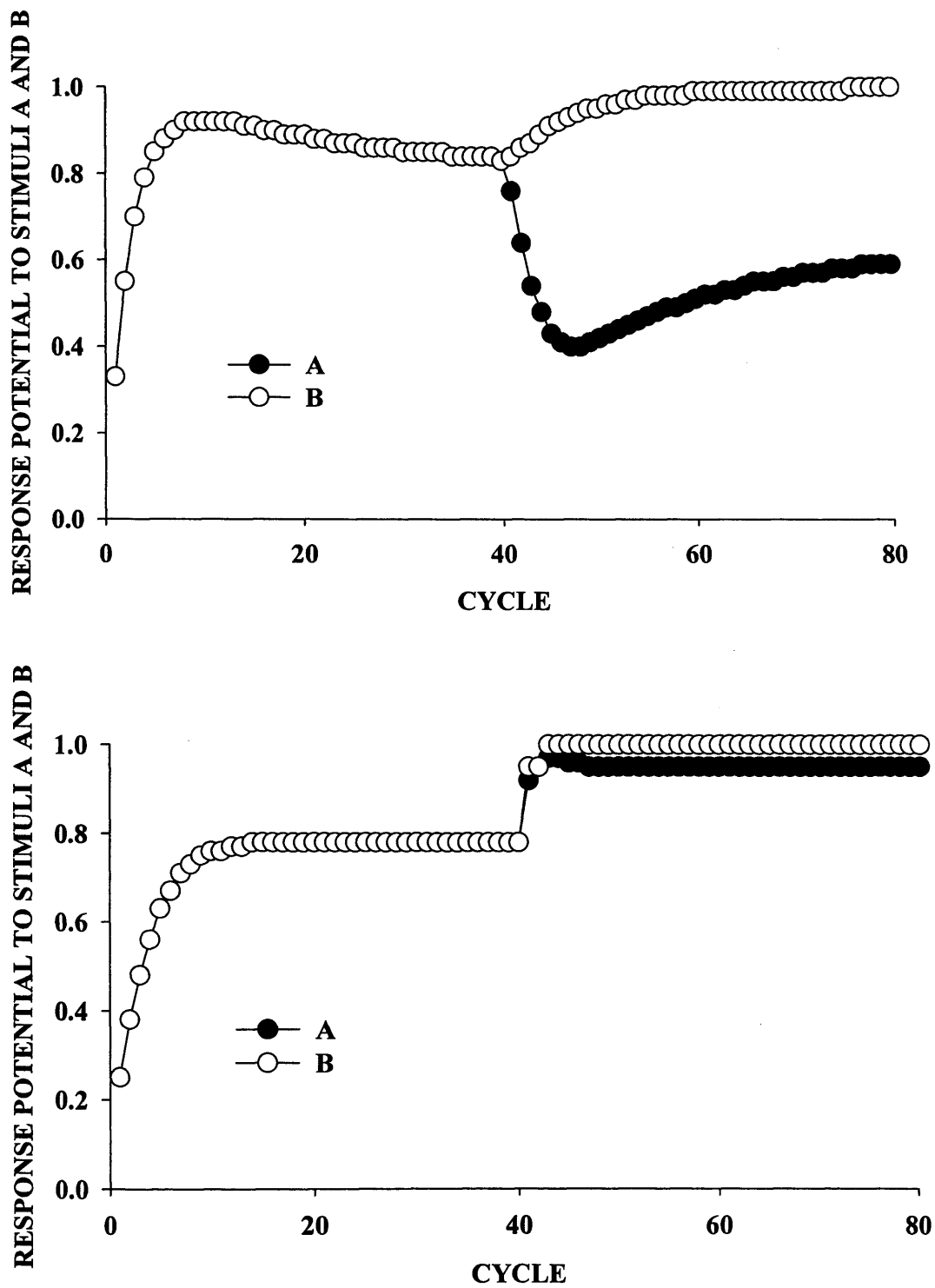


Figure 3.7. Computer simulations of the level of responding in the presence of cue A during AC+ trials followed by AXC+ trials and during BC+ trials when C (context) has moderate (Upper Panel) and maximum (Lower Panel) salience, based on the equations provided by Savastano et al. (2003).

When taken at face value, these results would also seem to be problematic for the theory of conditioning proposed by Rescorla and Wagner (1972). Recall that according to this theory, conditioning with A and then an AX compound should result in cue competition, preventing (or at least limiting) any additional association forming between A and the US in Stage 2. Consequently, conditioned responding to A should be no higher than to B, which was conditioned in isolation. However, as discussed earlier, the results of Experiment 4 imply that the association between the CS and the US was not the only association that influenced performance to A, the status of the association between A and X also influenced conditioned responding to A. The role of associations between neutral stimuli in studies of Pavlovian conditioning has been considered by a number of authors on a number of occasions (e.g. Rescorla & Cunningham, 1978), and these studies have motivated a number of theorists to incorporate structures within models of learning to account for so-called within-compound associations (e.g. McLaren & Mackintosh, 2001; Pearce, 2002; Wagner, 1980). The results of Experiment 3, therefore, do not necessarily provide the same degree of difficulty for standard associative models as they do for the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988). It should go without saying that if the association between A and X or the association between X and the US is undermined in any way then conditioned responding to A should, according to these accounts, be equal to B. This is of course the result we observed in Experiment 4.

Notes

1. Note that in order to simulate the predictions of the original Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) with the equations included in Savastano et al.'s (2003) article on the Extended Comparator Hypothesis the influence of the context was ignored. Thus, only one comparator stimulus (X) was considered for the target cue A.

Chapter 4

4. Assessment of a novel technique for the study of associability changes

4.1. Introduction

The two preceding chapters have provided some evidence against the account of cue competition put forward by the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988). The most distinctive feature separating this type of explanation from other theories of learning (e.g. Rescorla-Wagner, 1972; Mackintosh, 1975; Pearce & Hall, 1980; Pearce, 1987, 1994) is the proposal that cue competition reflects a failure to express rather than encode the target association. As discussed in Chapter 1, however, there is little agreement even among theorists who support the encoding-deficit view of cue competition as to the nature of the processes involved (Rescorla-Wagner, 1972; Mackintosh, 1975; Pearce & Hall, 1980; Pearce, 1987, 1994). As part of the present analysis of the mechanisms underpinning cue competition, I now evaluate the view that cue-competition effects are caused by a reduction in associability, as formalised in the classical theory of Mackintosh (1975). To this end, the two following chapters analyse the account of the relative validity effect (Wagner et al., 1968) provided by this theory.

The rules governing changes in stimulus' associability as advanced in Mackintosh's (1975) theory were expounded in some detail in Chapter 1. For present purposes, it should be sufficient to recall the psychological interpretation behind this account, according to which the best available predictors of a trial outcome will be endowed with high associability, whereas relatively poorer predictors will be endowed with low associability. Crucially, because associability determines the size

of trial-by-trial changes in associative strength, the associative strength of stimuli that are poor predictors of reinforcement will be modified more slowly.

In a relative validity experiment (e.g. Wagner et al., 1968), A is the best predictor of reinforcement in the true-discrimination group (AX+ BX-), and there should be a gain in its associability, which will enable it to acquire substantial amounts of associative strength. In contrast, the common element X, by virtue of its being a poorer predictor of the occurrence of reinforcement will lose associability and fail to secure a significant measure of associative strength. In the control, pseudo-discrimination group (AX+/- BX+/-), the associability of X will remain relatively high. Indeed, although all elements signal reinforcement with equal probability, B is twice as often paired with reinforcement as either A or X. When performance in the presence of X by itself is then compared between the two groups, the model predicts stronger responding in the pseudo-discrimination than the true-discrimination group (i.e. the relative validity effect (Wagner et al., 1968)).

For the purpose of analysing this account of relative validity (Wagner et al., 1968), the following strategy has been adopted. In the present chapter, a novel behavioural technique is used to investigate associability changes in true discriminations. One important outcome from this chapter is that X does not appear to lose associability during an AX+ BX- discrimination under certain circumstances. Drawing on these results, Chapter 5 addresses the question of whether associability changes can provide a complete account of relative validity (Wagner et al., 1968).

The introduction of a novel behavioural technique to measure associability changes was motivated by the inadequacy of the methods traditionally used (see Chapter 1, p. 50-51 for a discussion). Although some evidence (Mackintosh & Little, 1969; George & Pearce, 1999) have been gathered to support the contention that true-

discrimination training leaves the relevant stimuli (A and B) with higher associability than the irrelevant one (X), the crudeness of the methods to date affords little scope for further exploring the precise circumstances leading to changes in associability.

In the IDS-EDS technique, regarded by many as providing the clearest evidence of associability changes in a discrimination, (Turrisi, Shepp & Eimas, 1969, Mackintosh, 1974; George & Pearce, 1999), subjects are initially given discriminative training involving two stimulus dimensions. For the solution of the discrimination, stimuli from one of these dimensions are relevant (i.e. they predict with accuracy either the presence and absence of reinforcement), whereas stimuli from the other dimension are irrelevant (i.e. their presentation is uncorrelated with the presence or absence of reinforcement). For the test discrimination, all subjects receive a discrimination involving new stimuli from the dimensions previously used. For half of them, the solution of the test discrimination rests upon the use of stimuli from the previously relevant dimension (IDS group), whereas for the remaining subjects it rests upon the use of stimuli from the previously irrelevant dimension (EDS group). A number of experiments have found that the EDS group solves the test discrimination with greater difficulty than the IDS group, suggesting that the dimensions involved emerge with different associabilities from the initial discrimination (Shepp & Eimas, 1964; Mackintosh & Little, 1969; Pearce, Redhead & George, 1998; George & Pearce, 1999).

One shortcoming of the IDS-EDS method is that the effect relies on the associability of original stimuli generalising towards the test stimuli from the same dimension. The conditions that promote within-dimension generalisation of associability remain an open empirical question, the nature of which might in time

account for the frequent failures to report an IDS-EDS effect (e.g. Couvillon, Tennant & Bitterman, 1976; Hall & Channell, 1985).

It seems clear from the previous discussion that any attempt to investigate the role of associability changes in the relative validity effect (Wagner et al., 1968) should begin by developing novel methods capable of affording a more direct measure of associability changes in true-discrimination learning. One constraint imposed on any such endeavour is that the stimuli that are relevant to the solution of a true discrimination may differ from those irrelevant not only in terms of the attention they command (Mackintosh, 1975), but also in terms of the responses they elicit. When the stimuli in question are subsequently transferred into a new discrimination, attempts to assess their associability may be confounded by the different responses, or strength of responding, that occur in their presence. Any attempt to compare the associability of several stimuli directly seems thus to start off with an apparently impossible requirement: equating the stimuli's associative strengths before their tendencies to form subsequent associations may index changes in associability.

One solution to this conundrum consists in embedding the stimuli under study in compounds that are matched in their overall associative strengths. The basic design of the next experiment is shown in Table 4.1. In the first stage, subjects receive training with two true discriminations AX+ BX- and CY+ DY-, where A, B, C and D on the one hand, and X and Y on the other belong to different dimensions. Following this training, all subjects are transferred to a new discrimination in which the compound AY is reinforced, whereas AX and CY are not reinforced. Note that all stimuli present in the transfer discrimination form part of the original discrimination, each compound being composed of one previously relevant (A or C) and one previously irrelevant element (X or Y). Furthermore, the summed associative strength

of all three compounds can be assumed to be the same. The question of interest is whether responding to AX and CY will extinguish at the same rate.

Table 4.1. Basic experimental design used throughout this chapter.

Acquisition training	Test discrimination
AX+ BX- CY+ DY-	AY+ AX- CY-

According to a non-attentional type of account such as the Rescorla-Wagner (1972) model, performance during AX and CY during the test discrimination will be identical, provided the salience of the elements is equivalent. This prediction follows because changes in the associative strength of the elements in each of AX and CY during nonreinforced trials rely on error terms of equal magnitude. A similar prediction, albeit through the operation of rather different processes, can be derived from the configural theory of Pearce (1987, 1994, 2002). In contrast, the theory of Mackintosh (1975) predicts that responding during AX will extinguish more slowly than during CY ⁽¹⁾. According to this account, the solution of the true discriminations in the first stage will result in more attention being paid to A and C than to X and Y. Consequently, the discrimination between AY+ and CY- will be relatively easy to solve as it is based on A and C, to which the animal is already attending; whereas the discrimination between AY+ and AX- will be more difficult, as it is based on X and Y that have come to be ignored.

Experiments 5 and 6 test these conflicting predictions and report some evidence for the suitability of the technique described for revealing associability

changes. Experiment 7 investigates the mechanism responsible for these changes; in particular, whether they take place at a central or peripheral level.

4.2. Experiment 5

Two groups of pigeons initially received training with two true discriminations of the form AX+ BX- and CY+ DY- using autoshaping. The stimuli for the discriminations consisted of two squares that were presented side by side in the centre of a television screen situated behind a transparent response key. One of the squares was filled with one of four colours, and the other square was filled with one of four patterns. For the Colour-Relevant group, four colours were consistently followed by either the presence (A and C) or the absence of reinforcement (B and D), whereas two patterns were uncorrelated with the presentation of reinforcement (X and Y). For the Pattern-Relevant group, these relations were reversed, so that four pattern were reliable predictors of the presence (A and C) or omission of food (B and D), whereas two colours (X and Y) were poor predictors of the trial outcome.

Once the two groups had mastered their discriminations, a test discrimination was introduced with food presented after AY, but not after AX or CY. The counterbalancing of the stimuli in Stage 1 allowed for the identity of the compounds presented during the test discrimination to be the same for all the subjects. If associability changes occur as a result of true-discrimination learning, then the component discrimination based on the previously irrelevant elements (AY+ AX-) should be solved with greater difficulty in both groups than the component discrimination based on the previously relevant elements (AY+ CY-).

Method

Subjects. The subjects were 32 experimentally naïve male and female adult homing pigeons (*Columba livia*) that were housed in pairs and had free access to grit and water. They were gradually reduced to 80% of their free-feeding weights and were maintained at this level by being fed a restricted amount of food following each experimental session. They were kept in a light-proof room in which the lights were on for 14.5 hr each day. Training took place on successive days, at the same time, during periods when the lights were on in the pigeons' holding room.

Apparatus. The experimental apparatus consisted of eight 30 x 35 x 33 cm (Height/Width/Depth) pigeon test chambers. Each contained a 5-cm high x 4.5-cm wide clear acrylic panel which was hinged at the top. Pecks on the panel were detected by a reed relay that was operated whenever a bar-magnet attached to its lower edge was displaced by a distance greater than 1 mm. The midpoint of the panel was 22 cm above the floor of the chamber, and 16 cm from the front of the chamber. A Saka 7" TFT colour television (Model No.: T-V710) with a 15.5 x 9 cm screen was located 6 cm behind the acrylic panel. Food was delivered by operating a grain feeder (Colbourn Instruments, Lehigh Valley, PA) with an opening that measured 5.0 cm x 6.0 cm located in the same wall as the response key. The midpoint of the opening was 7 cm above the chamber floor and 16 cm from the front of the chamber. The feeder was illuminated whenever grain was made available. The chambers were permanently lit during all experimental sessions by a 2.8-W bulb operated at 24 V, located 2.5 cm above the top of the chamber's acrylic ceiling. A Risc PC (Acorn Computers Ltd., Cambridge, England) programmed in Arachnid (Paul Fray Ltd., Cambridge, England)

was used for the control of the experimental events, recording of responses and the generation of the stimuli on the TFT monitors.

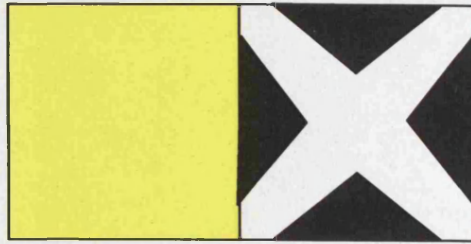
Stimuli. The stimuli were presented on the TFT monitor, which was otherwise dark. The stimuli consisted of two squares (2.0 cm x 2.0 cm) that contained either a colour or a pattern that were joined together along a vertical side. The midpoint of this vertical side coincided with the centre of the TFT screen. On half of the trials the colour stimulus was presented in the left square and the pattern was presented in the right square, for the remaining trials the positions of the colours and patterns was reversed. The four colours used were red, yellow, green and blue. The four patterns were black and white vertical stripes, black and white horizontal stripes, a black and white checkerboard, and a white St Andrews cross on a black background.

Procedure. The subjects first received three sessions in which they were trained to eat food whenever it was presented by the hopper. They were then given four sessions of autoshaping in which a white circle with a diameter of 16 mm was presented in the middle of the TFT monitor for 10 s. There were 45 trials in a session, the mean interval between the start of each trial was 60 s (range = 35 – 85 s), and food was made available by the hopper whenever the white circle was removed from the TFT screen. Whenever food was made available, it was presented for 4 s. The following thirty-two sessions comprised training with two true discriminations after the manner shown in Table 4.2. For animals in the Colour-Relevant group, food was presented following trials with compounds of red and St Andrew's cross (AX) and green and vertical lines (CY), but not after trials with yellow and St Andrew's cross (BX) and

Colour-Relevant group



AX+



BX-



CY+

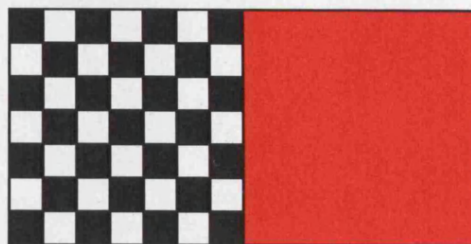


DY-

Pattern-Relevant group



AX+



BX-



CY+



DY-

Figure 4.1. Stimuli used in Experiment 5.

blue and vertical lines (DY). For subjects in the Pattern-Relevant group, food was presented following trials with compounds of St Andrew's cross and red (AX) and vertical lines and green (CY), but not after trials with checkerboard and red (BX) and horizontal lines and green (DY). There were 40 trials within each session, 10 for each trial type. The stimuli were presented in a random sequence, with the constraint that no more than three trials of the same type could occur in succession.

Table 4.2. Stimuli used during the acquisition and test phases in the Colour-Relevant and Pattern-Relevant groups

Acquisition training Colour-Relevant group	Stimulus identity
AX+ BX- CY+ DY-	red & cross + yellow & cross - green & vertical + blue & vertical -
Acquisition training Pattern-Relevant group	
AX+ BX- CY+ DY-	cross & red + checks & red - vertical & green + horizontal & green -
Test	
AY+ AX- CY-	red & vertical + red & cross - green & vertical-

On the following day, training on the transfer discrimination commenced. For each of twenty-two sessions all animals received trials in which a novel compound composed of red and vertical lines was paired with food (AY), whereas the previously reinforced compounds formed with red and St Andrew's cross (AX for the Colour-Relevant group and CY for the Pattern-Relevant group) and green and vertical lines

(CY for the Colour-Relevant group and AX for the Pattern-Relevant group) were nonreinforced. There were 40 trials within each session, 20 trials with the reinforced compound and 10 each with the two reinforced compounds. Other details were the same as in the previous stage.

Results. For the purpose of analysing the acquisition of the two true discriminations in the Colour-Relevant and Pattern-Relevant groups during Stage 1, the mean rates of responding per minute during all reinforced and nonreinforced trials were computed for each bird. Scores across the 32 sessions were then combined in 2-session blocks, as shown in Figure 4.2. Inspection of this figure suggests that the discriminations during Stage 1 were mastered more readily by the Colour-Relevant than the Pattern-Relevant group (see also: George & Pearce, 1999; Mackintosh & Little, 1969; and Hall & Channell, 1985). In order to validate this observation, a three-way analysis of variance (ANOVA) of individual mean responses per minute with the factors of group, CS (reinforced or nonreinforced) and session block was conducted for the 16 session blocks. The results from the ANOVA revealed an effect of session block, $F(15, 450) = 4.50$ and CS, $F(1, 30) = 167.28$, and significant Group x CS, $F(1, 30) = 23.01$, Session block x CS, $F(15, 450) = 33.98$, and Group x Session block x CS, $F(15, 450) = 9.80$, interactions. There was no significant effect of group, $F(1, 30) = 1.06$, or Group x Session block interaction, $F(15, 450) = 1.20$. Analysis of the significant three-way interaction, using tests of simple main effects, confirmed that the two groups differed in the readiness with which they solved their discriminations. Whereas the Colour-Relevant group showed evidence of having solved their discriminations from the second 2-session block onwards, $F_s(1, 480) > 24.78$, it took six 2-session blocks for the Pattern-Relevant group to start responding at significantly

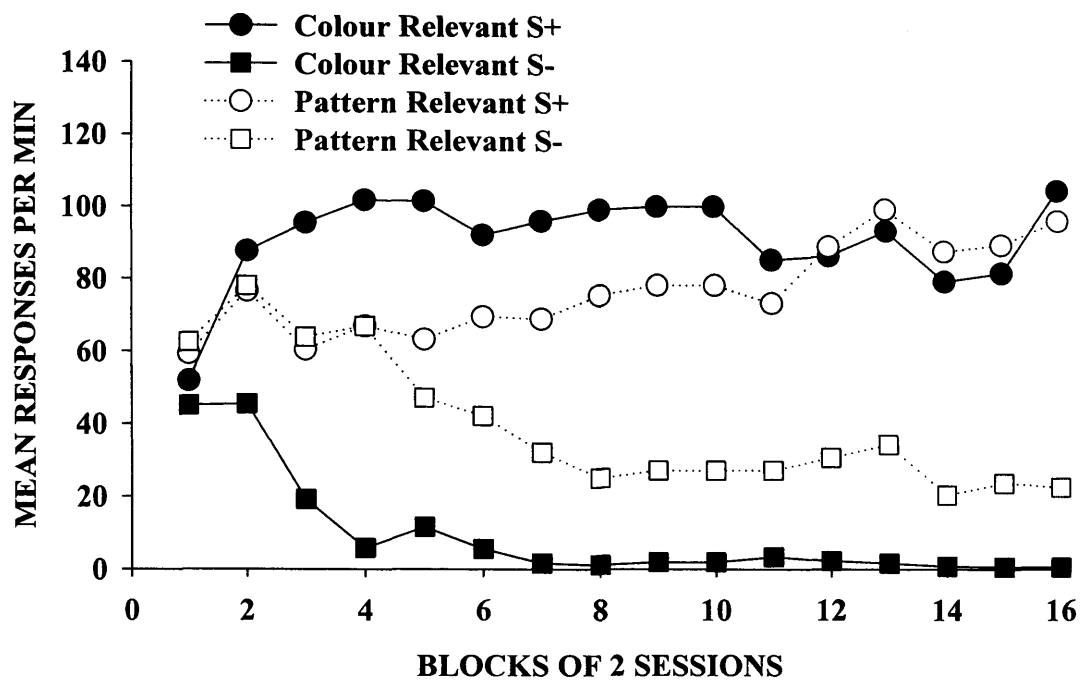


Figure 4.2. Acquisition in Experiment 5

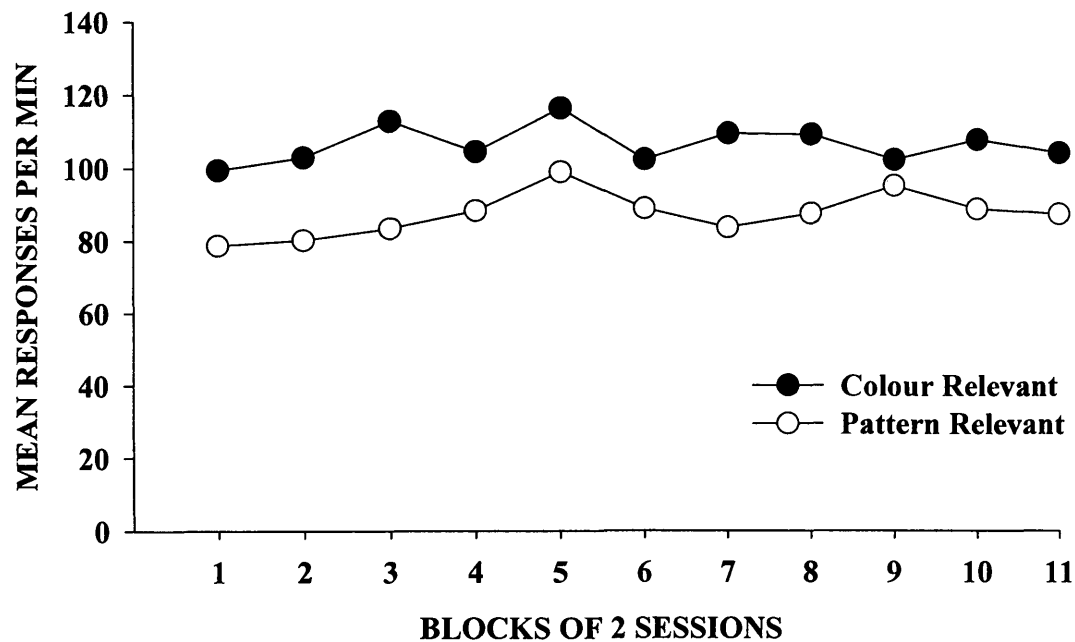


Figure 4.3. Reinforced compound during the test discrimination in Experiment 5.

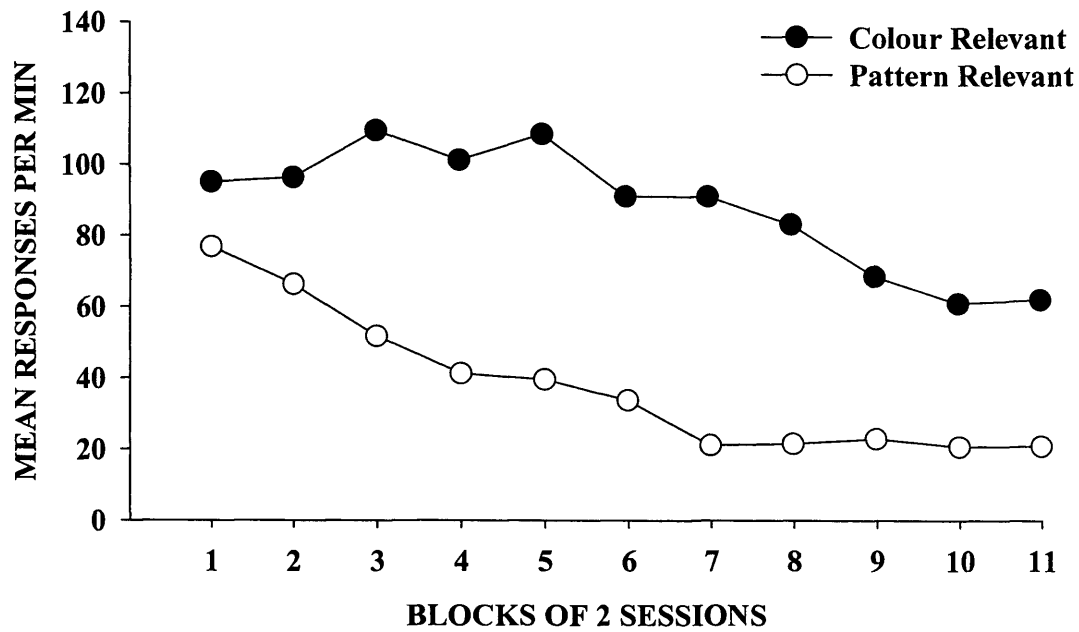


Figure 4.4. Non-reinforced compound with a pattern as the unique element during the test discrimination in Experiment 5.

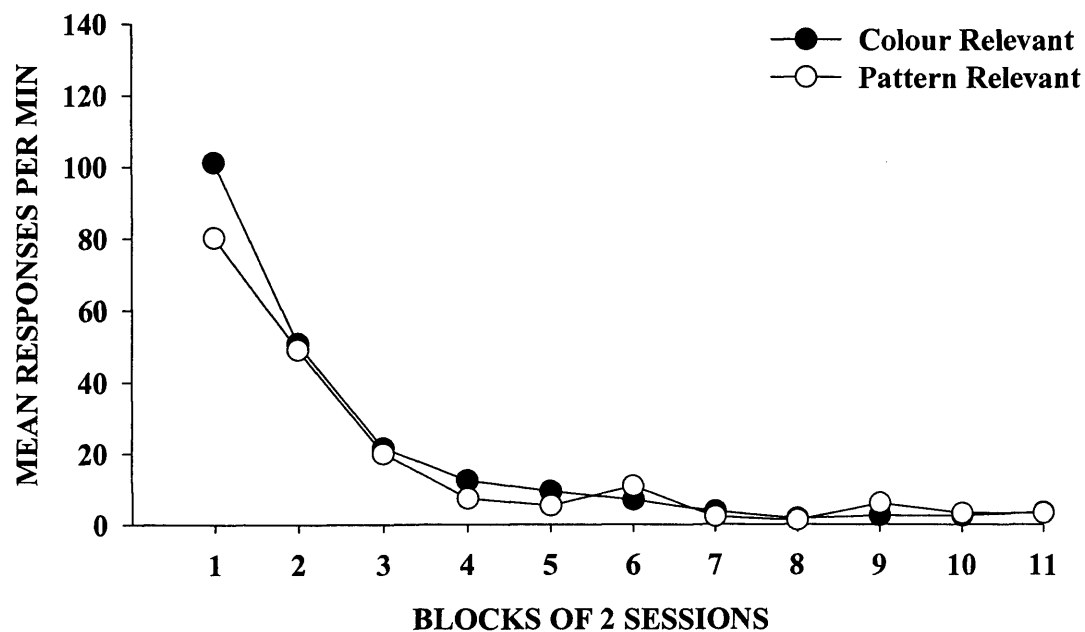


Figure 4.5. Non-reinforced compound with a colour as the unique element during the test discrimination in Experiment 5.

different rates in the presence of reinforced and nonreinforced CSs, $F_s(1, 480) > 29.46$.

The different rates at which the Colour-Relevant and the Pattern-Relevant groups solved their true discriminations have important implications for the analysis of the results from the AX+ AX- CY- discrimination. Recall that A and C were relevant in Stage 1 and X and Y were irrelevant. Suppose that A and C are colours, and X and Y are therefore patterns; the discrimination in Stage 2 between, say, AX+ and CY- is likely to be acquired more readily than between AX+ and AY- because of the more rapid learning that is possible with colours than with patterns. Given such an outcome it might then be difficult to determine whether there were any changes in the associability of the colours or patterns as a result of the training in Stage 1. In view of this problem, a between-group analysis was adopted when examining acquisition of each of the colour-based and pattern-based discriminations. Before meaningful between-group comparisons can be made, however, it is necessary to show that the mean rate of responding on reinforced trials by the end of Stage 1 is equivalent in both groups. A further simple main effects analysis revealed that the groups did not differ significantly in this respect from session block nine onwards, $F_s(1, 960) < 3.07$.

In order to analyse the acquisition of the test discrimination (AY+ AX- CY-), the groups were compared in terms of the readiness with which they solved the discrimination based on the colours on the one hand, and the discrimination based on the patterns on the other. If the associability of the two dimensions was modified in Stage 1, then the Pattern-Relevant group should find the discrimination based on the patterns easier than the Colour-Relevant group. By contrast, the Colour-Relevant group should solve the discrimination based on the colours more readily than the Pattern-Relevant group. At the outset of Stage 2, all three compounds in each group

should have a similar amount of associative strength. It would be reasonable to expect, moreover, that the associative strength of the reinforced compound will remain high in both groups throughout the test phase. The readiness with which the pattern-based and colour-based discriminations are solved is therefore indexed by the rate at which responding during the nonreinforced compounds extinguish across the groups. If the pattern-based discrimination is solved more readily by the Pattern-Relevant (AY+ CY-) than by the Colour-Relevant group (AY+ AX-), then extinction for the compound with a pattern as the unique element (relative to the reinforced compound) should proceed more quickly in the former group. Conversely, if the colour-based discrimination is solved with greater difficulty by the Pattern-Relevant (AY+ AX-) than the Colour-Relevant group (AY+ CY-), then the compound with a colour as the unique element (relative to the reinforced compound) should extinguish more readily in the latter group.

Figures 4.3, 4.4 and 4.5 show the groups' performance in the presence of the reinforced compound, the nonreinforced compound with a pattern as the unique element and the nonreinforced compound with a colour as the unique element across the eleven 2-session blocks of Stage 2. A glance at Figure 4.3 suggests differences in baseline rates of responding to the reinforced compound from the outset of Stage 2 across the groups. This apparent difference, however, was not confirmed by statistical analysis. A group x session block ANOVA of individual mean number of responses per minute in the presence of this compound across the eleven session blocks of Stage 2 revealed no significant effect of group, $F(1, 30) = 2.70$, or Group x Session block interaction, $F < 1$. The effect of session-block was significant, $F(10, 300) = 1.91$. This result is important if comparisons of the rates of extinction for the pattern-distinct and the colour-distinct compounds are to be meaningful indices of associability changes.

Turning to the Figure 4.4, it is evident that performance in the presence of the nonreinforced compound that had a pattern as the unique element extinguished faster in the Pattern-Relevant than the Colour-Relevant group. This observation was confirmed by a group x 2-session block ANOVA of individual mean number of responses per minute during the compound across the eleven 2-session blocks of Stage 2, which revealed significant effects of group, $F(1, 30) = 24.53$, 2-session block, $F(10, 300) = 18.73$, and a significant Group x 2-Session block interaction, $F(10, 300) = 4.56$. An exploration of this interaction by means of simple main effects analysis revealed that the groups differed significantly in their rates of responding during this compound from the second 2-session block onwards, $F_s(1, 330) > 2.14$. This is the type of result one would expect if the associability of the patterns was higher in the Pattern-Relevant than the Colour-Relevant group as a result of training in Stage 1.

Inspection of Figure 4.5 clearly indicates that this effect was not symmetrical. Indeed, both groups appear to have reduced responding in the presence of the compound with the colour as the unique element at the same rate, suggesting no changes in the associability of the colours. A group x 2-session block ANOVA of individual mean number of responses per minute during the colour-distinct compound across the eleven 2-session blocks of Stage 2 confirmed this observation by revealing no significant effect of group, $F < 1$, or Group x 2-Session block interaction, $F(10, 300) = 1.26$. An effect of session was found significant, $F(10, 300) = 86.35$.

Discussion. Experiment 5 reveals that the novel technique that was developed for studying associability changes was useful. Following training on two true discriminations involving patterns and colours, the associability of patterns that were

relevant to their solution was greater than of patterns that were irrelevant, a result compatible with Mackintosh's (1975) theory. A similar outcome was not observed with the colours that were used, which implies that the discrimination training resulted in a change of associability for stimuli belonging to one dimension, but not the other.

I shall return to the different results that were obtained with the colours and patterns shortly. For the present, it is important to consider the explanation for the finding that for the compound with a pattern as the unique element, responding extinguished more readily in the Pattern-Relevant than the Colour-Relevant group. I have suggested that this outcome was a consequence of the associability of the pattern being high in the Pattern-Relevant group –because it had been relevant to the solution of the discriminations—and low in the Colour-Relevant group –because it had previously been irrelevant to the solution of the discriminations. There is, however, another explanation for this finding. The design of Experiment 5 ensures that some patterns (vertical lines and St. Andrew's cross) were presented twice as often during Stage 1 in the Colour-Relevant than the Pattern-Relevant group. During the subsequent test discriminations, the more rapid rate of extinction with the compound that had a pattern as the unique element in the Pattern-Relevant group might therefore have been a consequence of the lesser exposure to the patterns in this group relative to the Colour-Relevant group. One purpose of the next experiment was to evaluate this explanation for the results of Experiment 5.

The greater ease with which the Colour-Relevant than the Pattern-Relevant group solved the discrimination in Stage 1 suggests that the discriminability of the four colours that were selected was greater than of the four patterns. Perhaps, therefore, it is possible to reduce the associability of a stimulus during a

discrimination only if it is similar to other stimuli. In view of this possibility, the design of the next experiment differed from the one just described in a second way. The four colours that were used were more similar to each other than those used for the present experiment.

4.3. Experiment 6

The design of the present experiment was similar to the above study, except that steps were taken to ensure that the groups received similar exposure to the patterns and colours that were presented during Stage 1. The training given in Stage 1 is summarised in Table 4.3, which shows that both groups received four discriminations involving four colours and four patterns. Stimuli A through D were colours for the Colour-Relevant group, and patterns for the Pattern-Relevant group, whereas W through Z were patterns for the Colour-Relevant group, and colours for the Pattern-Relevant group. In addition, for the reason just given, the colours in the present experiment were changed to two shades of green, and two shades of blue. The assignment of these colours to the compounds in Stage 1 ensured that two shades of the same hue were used for the colour-based discrimination in Stage 2. A further change between the two experiments is that the patterns were red and white, rather than black and white.

Table 4.3. Example of the training received in the Colour-Relevant and Pattern-Relevant groups in Experiment 6.

Colour-Relevant	Stimulus identity	
AX+ BX- CY+ DY- AW+ BW- CZ+ DZ-	dark green & vertical + light green & horizontal + dark green & cross + light green & checks +	light blue & vertical - dark blue & horizontal - light blue & cross - dark blue & checks -
Pattern-Relevant		
AX+ BX- CY+ DY- AW+ BW- CZ+ DZ-	vertical & dark green + horizontal & light green + vertical & light blue + horizontal & dark blue +	cross & dark green - checks & light green - cross & light blue - checks & dark blue -
Test		
AY+ AX- CY-	dark green & horizontal + dark green & vertical -	light green & horizontal -

Method

Subjects and Apparatus. The subjects were 32 experimentally naïve adult homing pigeons from the same stock and maintained in the same way as for Experiment 5. Following autoshaping, the pigeons were matched on their rate of responding to the white circle and assigned in equal numbers to the four groups. The apparatus was the same as for Experiment 5.

Stimuli. The stimuli were based upon the stimuli used in Experiment 5. The four colours used were dark blue, light blue, dark green and light green. The four patterns

were red and white vertical stripes, red and white horizontal stripes, a red and white checkerboard, and a red St Andrew's cross on a white background. All other details were identical to Experiment 5.

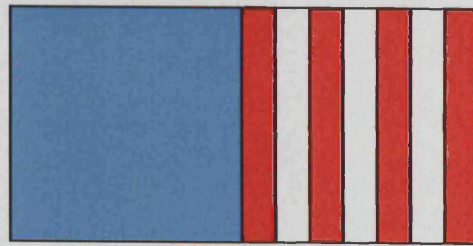
Procedure. Magazine training and autoshaping with a white circle were the same as for Experiment 5. Following autoshaping all subjects received thirty-two sessions of conditioning, with four true discriminations. For half of the animals in the Colour-Relevant group, the discriminations are shown in the upper panel of Table 4.3. For the remaining animals in the Colour-Relevant group the reinforcement contingencies were reversed. For half of the animals in the Pattern-Relevant group, the discriminations were those shown in the middle panel of Table 4.3. For the remaining animals in the Pattern Relevant group the reinforcement contingencies were reversed. There were ten presentations for each trial type within a session, which consisted of 80 trials.

For the following nine test sessions, training was carried out on the transfer discrimination. Half of the animals (see lower panel of Table 4.3) in the Colour-Relevant and Pattern-Relevant group received reinforced trials with the novel compound dark green and horizontal lines (AY), and nonreinforced trials with the compounds dark green and vertical lines (AX for the Colour-Relevant group and CY for the Pattern-Relevant group) and light green and horizontal lines (CY for the Colour-Relevant group and AX for the Pattern-Relevant group). The remaining animals received reinforced presentations of the novel compound dark blue and St Andrew's cross (AY) and nonreinforced presentations of dark blue and checkerboard (AX for the Colour-Relevant group and CY for the Pattern-Relevant group) and light blue and St Andrews cross (CY for the Colour-Relevant group and AX for the

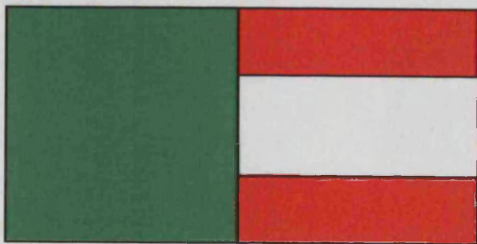
Colour-Relevant group



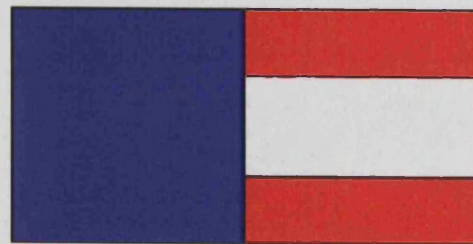
AX+



BX-



CY+

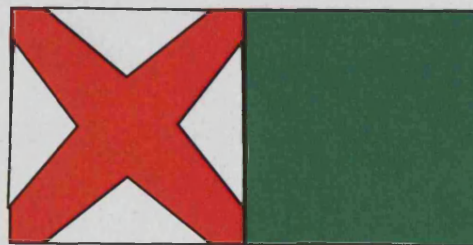


DY-

Pattern-Relevant group



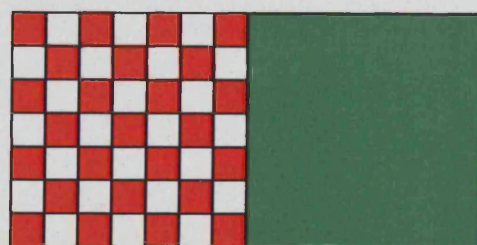
AX+



BX-



CY+



DY-

Figure 4.6. Sample of the stimuli used in Experiment 6.

Pattern-Relevant group). There were sixty-four trials within each session, thirty-two trials with the reinforced compounds AY and sixteen with each of the nonreinforced compounds AX and CY. All other details of the procedure were the same as in Experiment 5.

Results. The results from Experiment 6 were analysed in the same manner as those of Experiment 5. Figure 4.7 shows the mean rates of responding per minute during all reinforced and nonreinforced trials calculated for each subject across the sixteen 2-session blocks. Despite the changes made in stimulus identity relative to Experiment 5, the Colour-Relevant group again appears to have solved their discriminations more readily than the Pattern-Relevant group. A three-way analysis of variance (ANOVA) of individual mean number of responses per minute with the factors of group, CS (reinforced v nonreinforced) and session block lent support to this observation. The results from the ANOVA revealed an effect of CS, $F(1, 30) = 57.23$, and significant Group x CS, $F(1, 30) = 4.42$, Session block x CS, $F(15, 450) = 14.67$, and Group x Session block x CS, $F(14, 450) = 10.06$ interactions. The effects of group, $F(1, 30) = 4.07$, Session block, $F(15, 450) = 1.39$, and the Group x Session block interaction, $F(15, 450) = 1.33$, were not statistically significant. Exploration of the three-way interaction with tests of simple main effects revealed that responding to reinforced and nonreinforced stimuli in the Colour-Relevant group differed significantly from the first Session block of acquisition training, $F(1, 480) > 7.68$, whereas performance to the two kinds of stimuli in the Pattern-Relevant group did not begin to differ until the sixth Session block, $F(1, 480) > 4.45$.

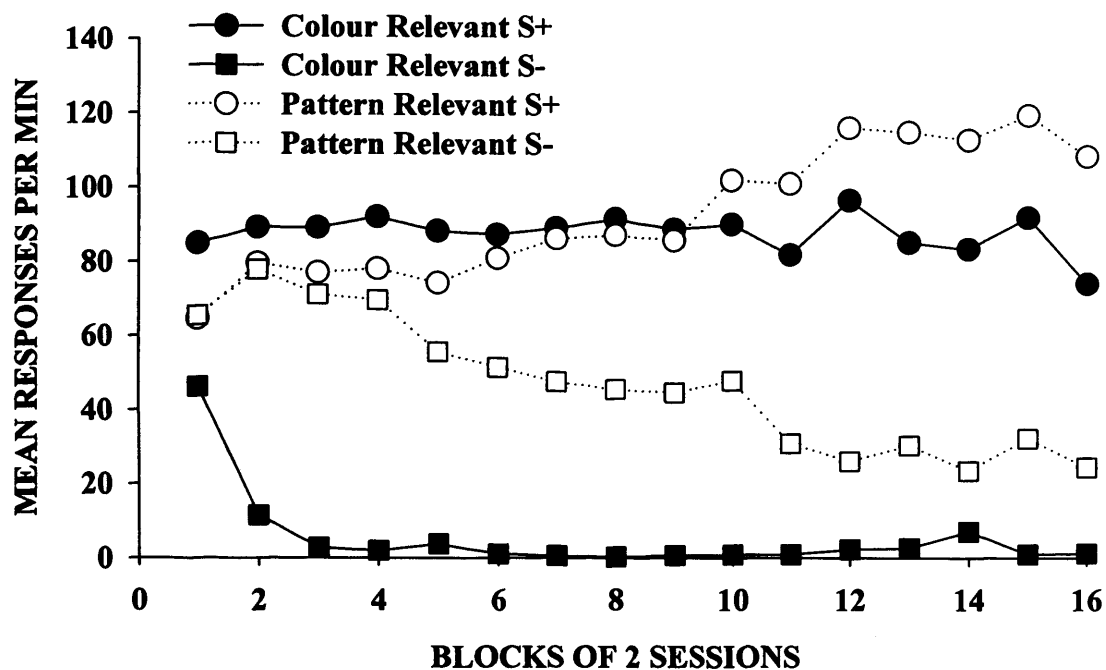


Figure 4.7. Acquisition in Experiment 6.

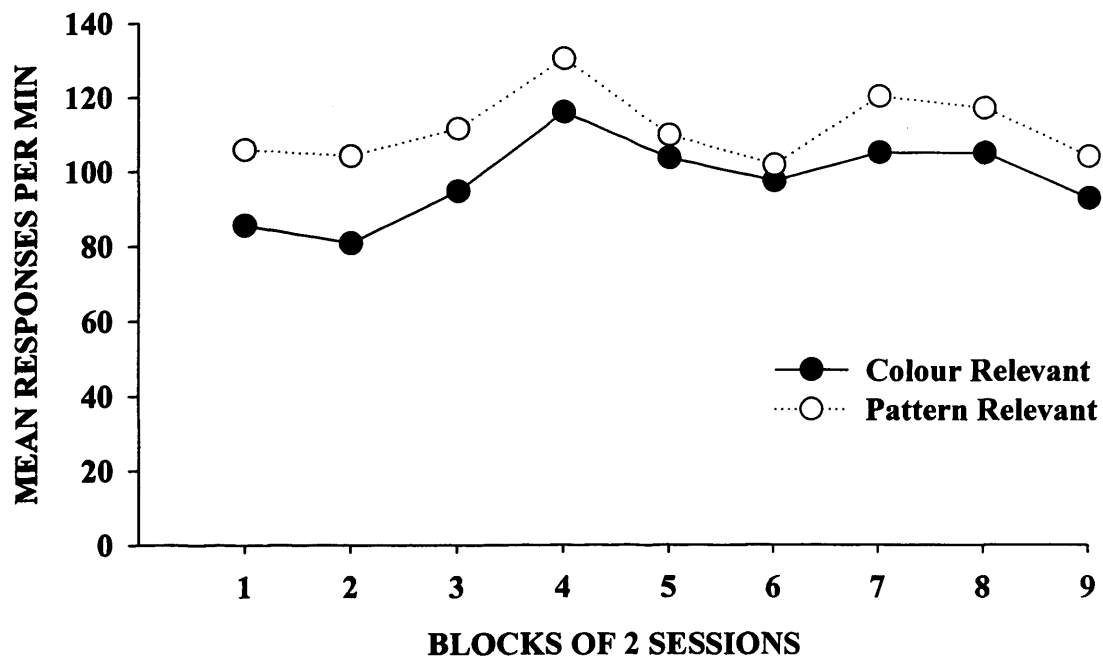


Figure 4.8. Reinforced compound during the test discrimination in Experiment 6.

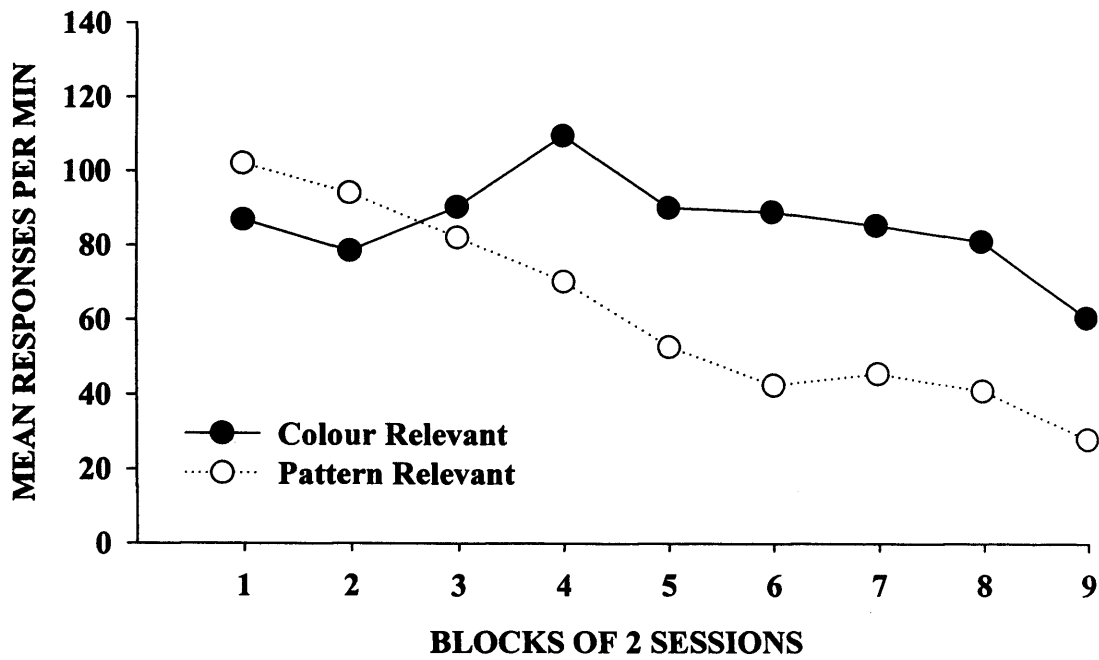


Figure 4.9. Non-reinforced compound with a pattern as the unique element during the test discrimination in Experiment 6.

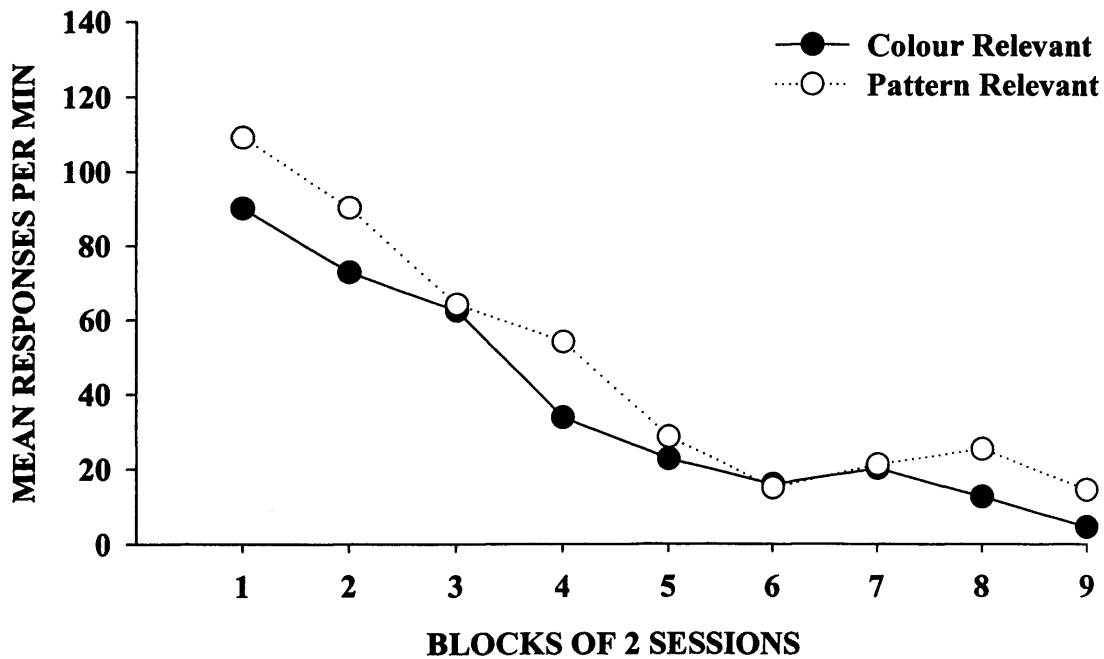


Figure 4.10. Non-reinforced compound with a colour as the unique element during the test discrimination in Experiment 6.

This undesired replication of Experiment 5's results on acquisition forces once more the adoption of between-group comparisons when analysing the results from the test stage. As for Experiment 5, therefore, the results from the transfer discriminations based on colours and patterns were analysed separately by comparing the rates of extinction during each nonreinforced compound across the groups. If the results from Experiment 5 are replicated, then the birds should withhold their responses in the presence of the compound with a pattern as the unique element faster in the Pattern-Relevant group (CY) than in the Colour-Relevant group (AX). Additionally, in view of the failure of the present experiment to reduce the discriminability of the colours relative to the patterns, no difference between the groups should be expected in their rates of extinction in the presence of the compound with a colour as the unique element (CY for the Colour-Relevant group and AX for the Pattern-Relevant group).

Again, this analysis requires that potential between-group differences following acquisition training are not confounded with baseline differences in the rate of responding during reinforced stimuli by the end of Stage 1. A further simple main effects analysis based on the previous ANOVA revealed equivalent levels of performance in the groups during reinforced trials from 2-session blocks 1 to 15, $F(1, 960) = 3.93$, but significantly higher levels of responding for the Pattern-Relevant than the Colour-Relevant group on the final session block, $F(1, 960) = 4.42$. Although not statistically significant, this tendency continued during the transfer discriminations, as suggested in Figures 4.8, which shows the performance of the groups in the presence of the reinforced compound during Stage 2. An ANOVA of individual mean number of responses per minute during this compound across the nine sessions of Stage 2 revealed no effect of group, $F < 1$, and no significant Group x

Session block interaction, $\underline{F} < 1$. The effect of session was significant, $\underline{F}(8, 240) = 2.31$.

In any case, this tendency for the Pattern-Relevant group to respond more vigorously on reinforced trials than the Colour-Relevant group poses no problem for the interpretation of the results from Stage 2. Indeed, if this tendency is to account by itself for any difference in the rates of extinction in the presence of the compound with a pattern as the unique element, then responding in the Pattern-Relevant group should take longer to extinguish than in the Colour-Relevant. In fact, in keeping with the results from Experiment 5, the opposite was found, as shown in Figure 4.9. An ANOVA of individual mean number of responses per minute during the nonreinforced compound with the pattern as the unique element across the nine sessions of Stage 2 revealed an effect of session, $\underline{F}(8, 240) = 5.44$, and a significant Group x Session interaction, $\underline{F}(8, 240) = 3.07$. The effect of group was not significant, $\underline{F}(1, 30) = 2.11$. A simple main effects analysis conducted to explore the significant interaction revealed that the groups differed on session 6, $\underline{F}(1, 270) = 4.89$. Marginal levels of significance were also found on Sessions 7 and 8 ($p = 0.059$ and $p = 0.058$, respectively).

Figure 4.10 shows the performance of the groups in the presence of the nonreinforced compound with a colour as the unique element. As for Experiment 5, it appears that the change of colours for the present experiment did not enable their associability to be changed. An ANOVA of individual mean number of responses per minute during this compound across the nine sessions of Stage 2 confirmed this observation by revealing no effect of group, $\underline{F}(1, 30) = 1.26$, or Group x Session interaction, $\underline{F} < 1$. The effect of session was significant, $\underline{F}(8, 240) = 37.93$.

Discussion. Experiment 6 provided a replication of the results from Experiment 5. Its significance lies, therefore, in the consistency it confers to the behavioural technique employed. Moreover, this consistency was observed even when steps were taken to ensure that both groups received equal exposure to all the stimuli that were used. The experiment did not succeed, though, in its attempt to equate the relative salience of colours and patterns. As was the case in Experiment 5, birds in the Pattern-Relevant group showed greater proficiency upon transfer at solving the discrimination based on the patterns than birds in the Colour-Relevant group. However, the two groups solved the test discrimination based on the colours with equal ease. I shall consider the associability changes observed with the patterns shortly. For the present, I will turn to the apparent failure to modify the associability of the colours is interpreted in the light of the theory of Mackintosh (1975).

Taken at its face value, the apparent failure to modify the colours associability is problematic for the theory of attention proposed by Mackintosh (1975). Regardless of the high intrinsic discriminability of the colours, which the theory captures by assigning a low value to the parameter S (see p. 45-46), the different contingencies with reinforcement undergone by the colours across groups should ensure significant changes in the associability parameter α . Two possibilities present themselves as to why the Color-Relevant and Pattern-Relevant groups did not differ in their ability to solve the colour-based test discrimination. The first of these presumes a special difficulty in altering the associability of those stimuli when it is already high. The second is that the test employed lacked sufficient sensitivity to reveal an existing difference in the associability of the colours. The former interpretation can be found in the theorising of Sutherland and Mackintosh (1971, p. 491; see also: LePelley, 2004, Suret & McLaren, 2003), who argue that it is hard to alter the associability of a

stimulus when it starts out with a high value. As a consequence, the rate at which the discrimination between the compounds that differ in terms of the colours proceeds should not differ in the Pattern-Relevant and Colour-Relevant groups. Support for the assumption that the associability of the colours was high can be found in the speed with which the Colour-Relevant group solved its true discriminations in Stage 1.

For the latter possibility, that a difference existed in the associability of the colours but it was hard to detect, a few explanations can be devised. For instance, it is conceivable that the change of context promoted by the introduction of the transfer discriminations acts, by some unspecified process, to restore the associability of stimuli with high salience, such as the colours. Alternatively, the colour-based test discrimination may have been solved so rapidly by both the Colour-Relevant and the Pattern-Relevant groups that it was impossible to detect differences between the groups. Although it is impossible to reject this explanation, the gradual rate at which extinction proceeded in the presence of the compound with colour as the unique element in both groups renders this possibility unlikely. Indeed, at a rate of ten presentations of this compound per test session, eight sessions for performance to reach floor would seem sufficient to reveal any differences in associability between the groups.

Of greater theoretical interest are the results obtained with the pattern-based discrimination. Although I have hitherto argued for an interpretation in terms of changes in the associability of the patterns, other theoretical accounts must be considered. I shall defer an exposition of these to the discussion of all the experiments in the chapter, however, until a more trivial possibility has been examined: that the associability changes were an artefact generated by orienting responses. Most attentional theories (Mackintosh, 1975; LePelley, 2004; Kruschke, 1992; Pearce,

George and Redhead, 1998) conceive of associability changes as operating upon the central representation of the stimulus. However, much earlier, Spence (1937, 1940) suggested that the locus for changes in associability might be more peripheral. More specifically, he noted that in order to solve a discrimination, the animal must evoke "...responses which lead to the reception of the appropriate aspects of the total environmental complex on the animals sensorium" (Spence, 1937, p.432). That is, as a consequence of being relevant to the solution of a discrimination, an orienting response is directed towards the relevant stimulus (see also: Wyckoff, 1952). This proposal provides a simple explanation for why learning following an intradimensional shift is superior to learning following an extradimensional shift (see: George & Pearce, 1999 for a more recent discussion of this approach). This notion can also provide a straightforward explanation for the results of Experiments 5 and 6 by assuming that the birds in the Colour-Relevant group were less likely to look at the patterns in the pattern-based test discrimination than were the birds in the Pattern-Relevant group.

To explain the failure to change the associability of colours, it would have to be further assumed that both groups were unable to prevent themselves from looking at the square containing the colours throughout the experiment.

4.4. Experiment 7

To avoid the possibility of orienting responses creating the illusion of associability change, Mackintosh (1965a, 1965b) proposed the use of integrated stimuli. For example, if an animal were trained on a true discrimination in which the

components of, say, the AX compound were integrated rather than separate (i.e. a square containing dark-green vertical lines, rather a square of dark green next to a square containing vertical red and white lines) then it would not be easy for a receptor-orienting act to be directed at one of these components without also being directed at the same time to the other. The purpose of Experiment 7 was to explore whether the solution of a true discrimination with spatially integrated stimuli would generate the same type of associability effect that was seen in Experiments 5 and 6. If it does, then it would imply that the locus of the associability change does not reside at the peripheral, response level, but may in fact be operating upon a central stimulus representation, providing support for the theory proposed by Mackintosh (1975).

For Experiment 7, therefore, pigeons were trained on the true discriminations used in Experiment 6 (AX+ BX- CY+ DY- AW+ BW- CZ+ DZ-), in which either a colour or a pattern was the relevant dimension to their solution. For half the birds, the colours and patterns were presented separately on the monitor, as in Experiments 5 and 6 (groups Colour-Relevant-Separate and Pattern-Relevant-Separate). For the remaining birds these two dimensions were integrated (groups Colour-Relevant-Integrated and Pattern-Relevant-Integrated). Once the true discriminations had been mastered, the test discrimination was again introduced with food presented after AY, but not after AX or CY.

Method

Subjects and Apparatus. The subjects were sixty-four experimentally naïve adult homing pigeons from the same stock and maintained in the same way as for

Experiments 5 and 6. Following autoshaping, the pigeons were matched on their rates of responding to the white circle and assigned in equal numbers to the four groups. The apparatus was the same as for the two previous experiments.

Stimuli. The stimuli were based upon those used in Experiment 6. For the Colour-Relevant-Separate and Pattern-Relevant-Separate groups, the stimuli were identical in all respects to the stimuli used in Experiment 6, with the exception that the patterns were coloured red and black, rather than red and white. For the Colour-Relevant-Integrated and Pattern-Relevant-Integrated groups the stimuli were contained within a single square (2 cm x 2 cm) that was located in the centre of the monitor. Within this square were presented the same patterns that were shown to the Colour-Relevant-Separate and Pattern-Relevant-Separate groups. The colour of the pattern was black and whichever colour accompanied the same pattern in the corresponding separate group. For example, one of the patterns seen by the Colour-Relevant-Separate group was a dark-green square adjacent to red vertical lines on a black background. For the Colour-Relevant-Integrated group, the corresponding stimulus was a single square of vertical lines that were dark green on a black background.

Procedure. Magazine training and autoshaping with a white circle were the same as for the two previous experiments. In each of the subsequent twenty-four sessions, all animals received true-discrimination training that was identical to the training given to the birds in Experiment 6, with two exceptions: there were sixty-four trials per session, and integrated stimuli were used for the trials in the Colour-Relevant-Integrated and Pattern-Relevant-Integrated groups. Following this training, animals

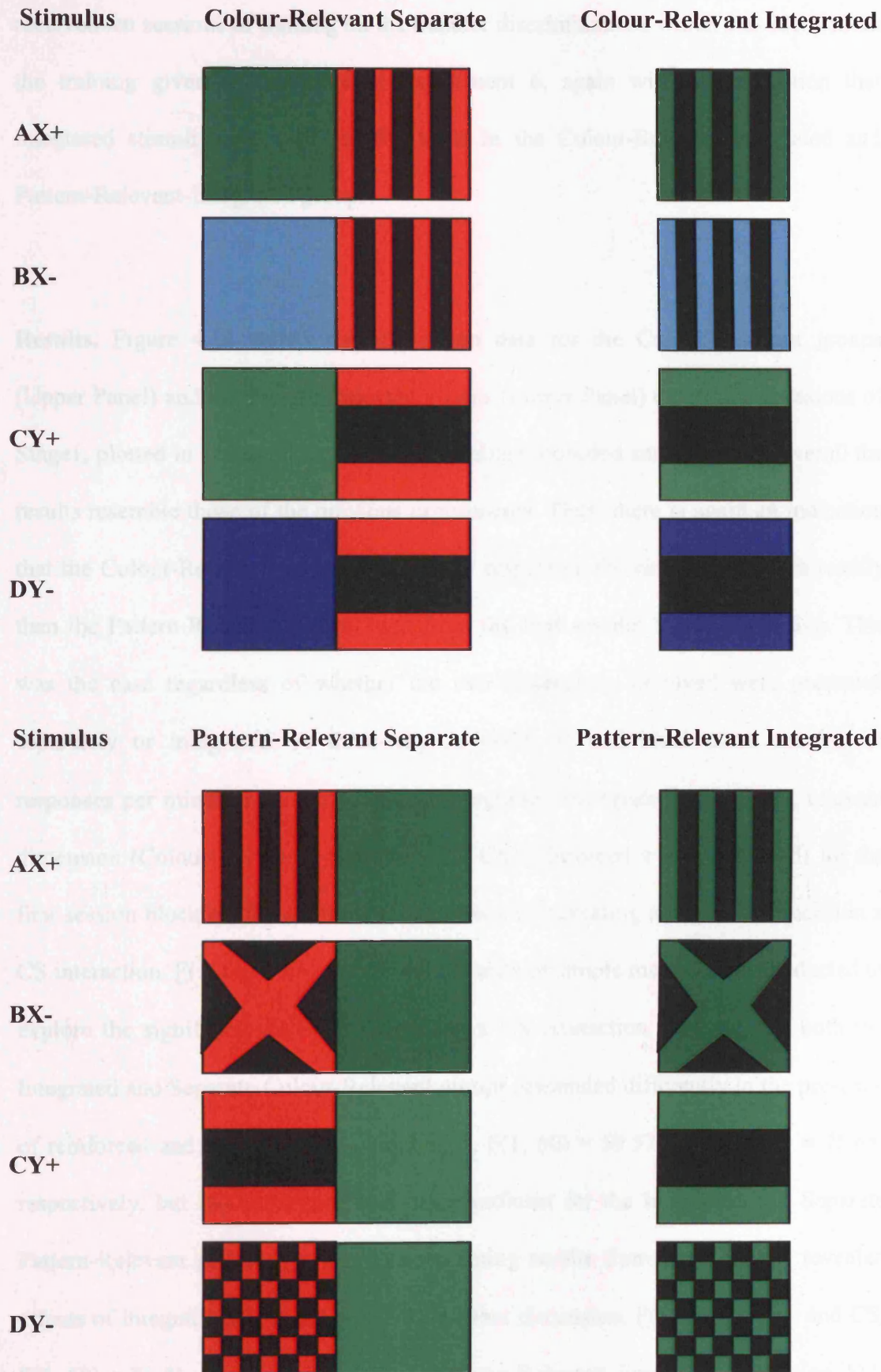


Figure 4.11. Sample of the stimuli used in Experiment 7.

received ten sessions of training on the transfer discriminations which was identical to the training given to the birds in Experiment 6, again with the exception that integrated stimuli were used for the trials in the Colour-Relevant-Integrated and Pattern-Relevant-Integrated groups.

Results. Figure 4.12 shows the acquisition data for the Colour-Relevant groups (Upper Panel) and the Pattern-Relevant groups (Lower Panel) over the 24 sessions of Stage1, plotted in 2-session blocks. Conditioning proceeded smoothly, and overall the results resemble those of the previous experiments. Thus, there is again an indication that the Colour-Relevant groups solved their respective discriminations more readily than the Pattern-Relevant groups right from the first session block of training. This was the case regardless of whether the two dimensions involved were presented separately or integrated. A three-way ANOVA of individual mean number of responses per minute with the factors of integration (integrated v separate), relevant dimension (Colour- v Pattern-Relevant) and CS (reinforced v nonreinforced) for the first session block confirmed these observations by revealing a Relevant dimension x CS interaction, $F(1, 60) = 63.64$. Subsequent tests of simple main effects conducted to explore the significant Relevant dimension x CS interaction revealed that both the Integrated and Separate Colour-Relevant groups responded differently in the presence of reinforced and nonreinforced compounds, $F(1, 60) = 59.57$ and $F(1, 60) = 75.65$, respectively, but this difference was not significant for the Integrated and Separate Pattern-Relevant groups, $F_s < 1$. The remaining results from the ANOVA revealed effects of integration, $F(1, 60) = 17.98$, relevant dimension, $F(1, 60) = 6.03$, and CS, $F(1, 60) = 71.21$, and a significant Integration x Relevant dimension interaction, $F(1, 60) = 6.53$.

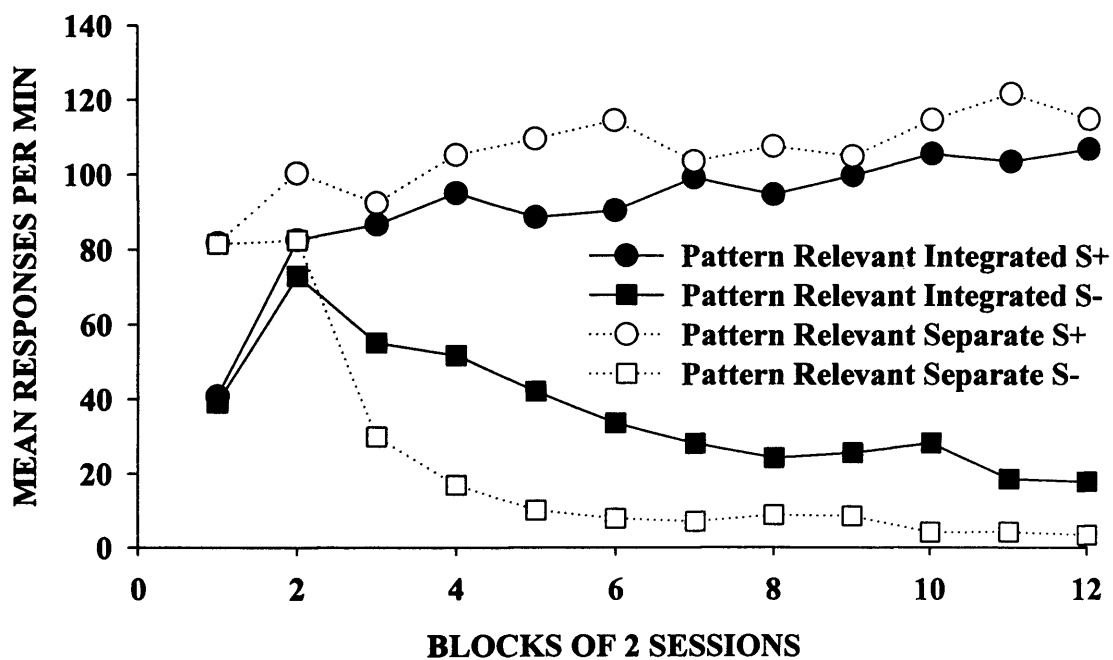
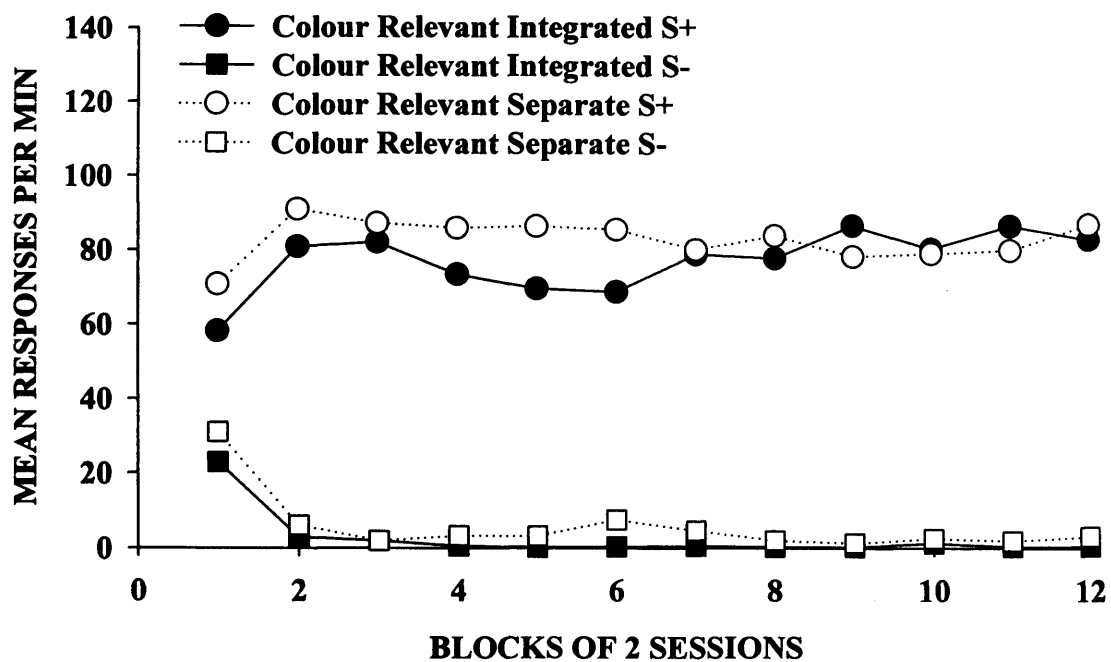


Figure 4.12. Acquisition in the Colour Relevant Integrated and Colour Relevant Separate groups (Top Panel) and Pattern Relevant Integrated and Pattern Relevant Separate groups (Bottom Panel) in Experiment 7.

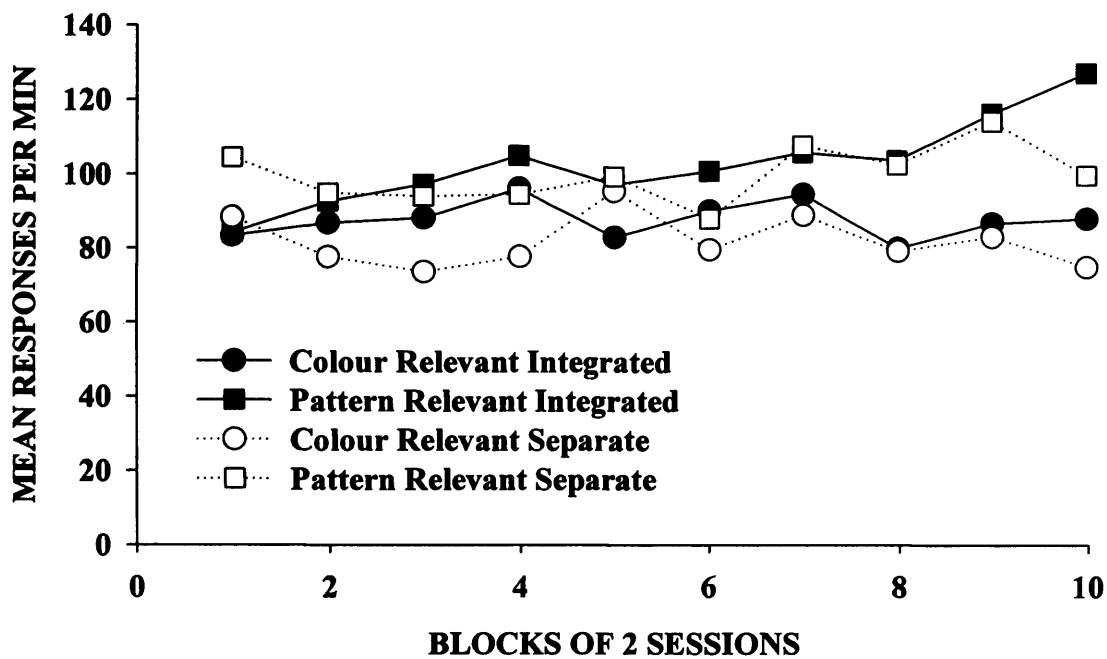


Figure 4.13. Reinforced compound during the test discrimination in Experiment 7.

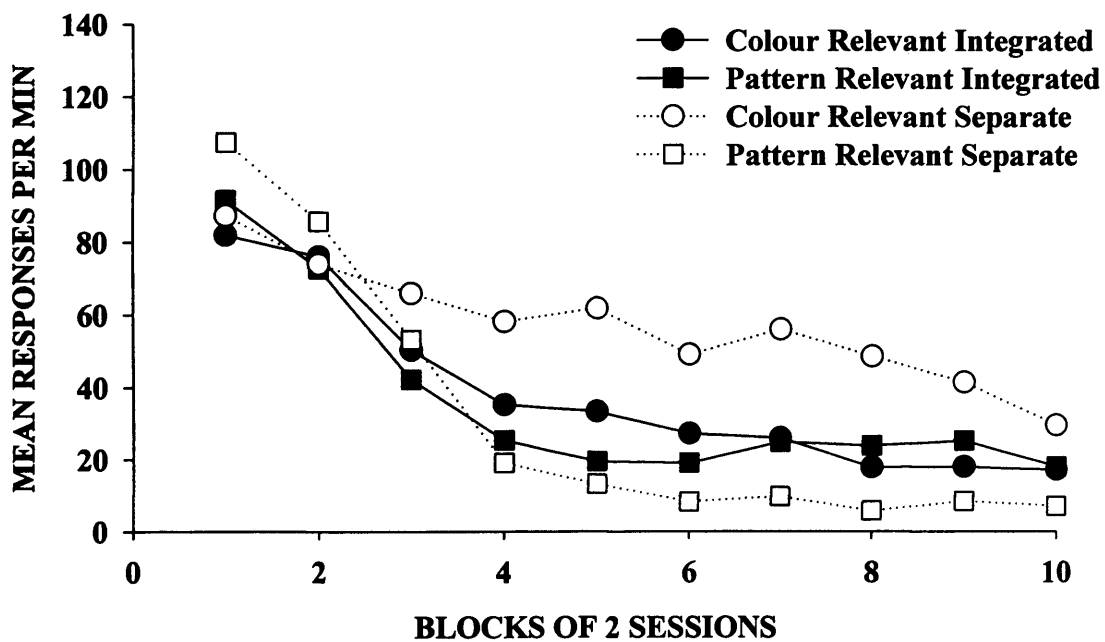


Figure 4.14. Non-reinforced compound with a pattern as the unique element during the test discrimination in Experiment 7.

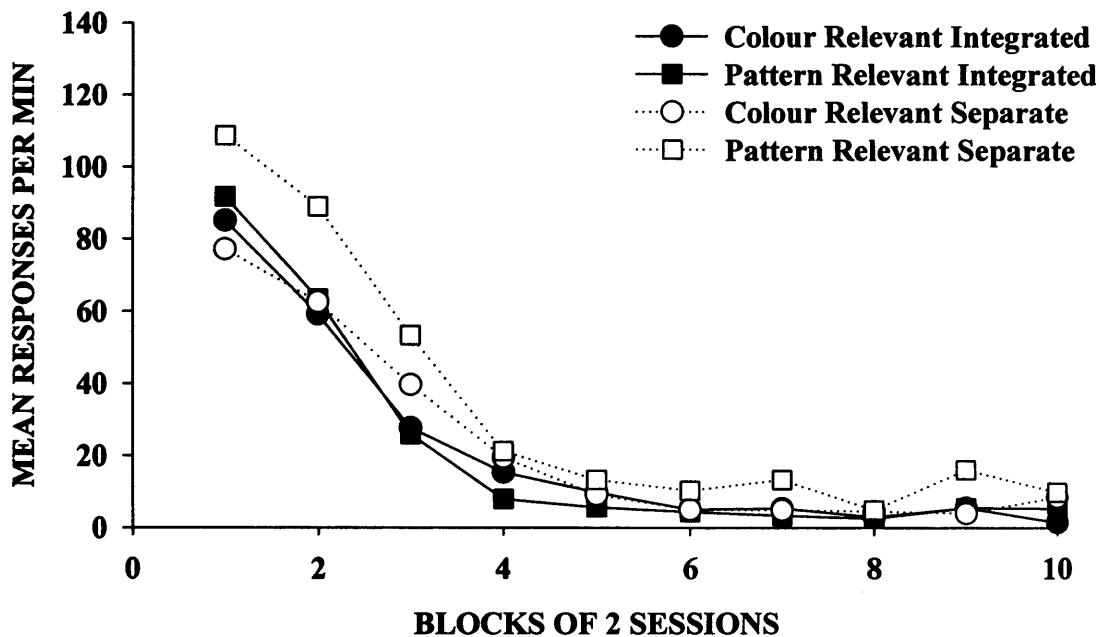


Figure 4.15. Non-reinforced compound with a colour as the unique element during the test discrimination in Experiment 7.

Neither the Integration x CS nor the Integration x Relevant dimension x CS interactions were statistically significant, $F < 1$.

Moreover, a comparison of the performance of the two Colour-Relevant groups, and for the two Pattern-Relevant groups, indicates that integrating the colours and patterns retarded the acquisition of the discrimination for the Pattern-Relevant groups, but not for the Colour-Relevant groups. These observations were confirmed by two separate three-way ANOVA of individual mean number of responses per minute across Stage 1. One ANOVA was based on the Colour-Relevant groups, whereas the other on the Pattern-Relevant groups. Both ANOVAS included the factors of Integration, CS and 2-Session block. For the Pattern-Relevant groups, this analysis most notably revealed an Integration x Session block x CS interaction, $F(11,$

330) = 2.17. Other results from this ANOVA were significant effects of Session block, $\underline{F}(11, 330) = 5.07$, and CS, $\underline{F}(1, 30) = 79.82$, and significant Integration x Session block, $\underline{F}(11, 330) = 5.65$, and Session block x CS, $\underline{F}(11, 330) = 36.90$ interactions. Exploration of the three-way interaction by means of simple main effects analysis revealed that the two Pattern-Relevant groups differed significantly in their discriminative performance on Session blocks 4, 5 and 6, $\underline{F}(1, 360) > 5.84$. By contrast, in the ANOVA for the Colour-Relevant groups the only significant results were the effect of CS, $\underline{F}(1, 30) = 88.94$, and the Session block x CS interaction, $\underline{F}(11, 330) = 8.66$. No effects of integration or session block, $\underline{F}s < 1$, and more important, no significant Integration x Session block, $F(11, 330) = 1.30$, Integration x CS, $\underline{F} < 1$, or Integration x Session block x CS interaction, $\underline{F}s < 1$, were found.

Before analysing the data for the transfer discriminations, it should be noted that, in keeping with the previous experiments, numerically faster terminal levels of performance on reinforced trials were observed for the Pattern-Relevant groups (Bottom Panel) than for the Colour-Relevant groups (Top Panel). A three-way ANOVA of individual mean responses per minute in the presence of the reinforced compounds, which included the factors of integration, relevant dimension and session block, confirmed this observation by revealing a Relevant dimension x Session block interaction, $\underline{F}(11, 660) = 2.76$. Exploration of this interaction using tests of simple main effects revealed a difference between the Colour-Relevant and Pattern-Relevant groups in their rate of responding during reinforced compounds on the final three session blocks, $\underline{F}s(1, 720) > 3.91$. The remaining results from the ANOVA were a significant effect of session block, $\underline{F}(11, 660) = 8.12$, no significant effects of integration, $\underline{F} < 1$, or relevant dimension, $\underline{F}(1, 60) = 2.52$, and no significant Integration x Relevant dimension, $\underline{F} < 1$, Integration x Session block, $\underline{F}(11, 660) =$

1.70, and Integration x Relevant dimension x Session block, $\underline{F} < 1$, interactions. As mentioned in the Results section for Experiment 6, faster responding during the reinforced trials by the Pattern-Relevant than the Colour-Relevant groups, though of some interest in itself, constitutes no challenge for the interpretation of the pattern-based test discrimination. If a similar pattern of results to that of Experiments 5 and 6 emerges, then slower responding in the presence of AX should be ultimately observed for the Pattern-Relevant than the Colour-Relevant group *despite* faster responding in the former at the outset of Stage 2.

The results from the ten sessions of training with the transfer discriminations are shown in Figures 4.13, 4.14, 4.15. As for Experiments 5 and 6, the performance of the groups in the presence of the reinforced compound (Figure 4.13), the compound with a pattern as the unique element (Figure 4.14) and the compound with the colour as the unique element (Figure 4.15) has been plotted and analysed separately. Inspection of Figure 4.13 reveals that performance in the presence of reinforced compound initially proceeded uneventfully. Towards the final sessions, however, there was again a tendency for the Pattern-Relevant groups to respond more vigorously than the Colour-Relevant groups. This observation received no statistical support in a three-way ANOVA for the ten sessions of Stage 2 that included the factors of integration, relevant dimension and session. The sole significant result from the ANOVA was an Integration x Session interaction, $\underline{F}(9, 540) = 1.98$. The effects of integration, $\underline{F} < 1$, relevant dimension, $\underline{F}(1, 60) = 1.27$, and session, $\underline{F}(9, 540) = 1.66$, as well as the interactions between Integration x Relevant dimension, $\underline{F} < 1$, Relevant dimension x Session, $\underline{F}(9, 540) = 1.57$, and Integration x Relevant dimension x Session, $\underline{F} < 1$, were not significant.

The results of primary interest are those from extinction with the compound that contained a pattern as the unique element, which yielded significant differences between the Colour-Relevant and Pattern-Relevant groups in Experiments 5 and 6. A glance at the response rates for the Colour-Relevant and Pattern-Relevant groups (Figure 4.14) indicates that this result was again replicated, but only for the Pattern-Relevant-Separate and Colour-Relevant-Separate groups. For the groups trained with integrated stimuli, responding during this compound extinguished at similar, relatively rapid rates. These observations were confirmed by a three-way ANOVA for the ten sessions of the test stage with the factors of Integration, Relevant dimension and Session, which revealed a significant Integration x Relevant dimension x Session interaction, $F(9, 540) = 2.66$. Subsequent analysis of simple main effects conducted to explore the three-way interaction revealed significant Integration x Relevant dimension interactions on Sessions 7 and 8, $F(1, 600) > 4.25$. This analysis also revealed that the Separate Colour-Relevant and Pattern-Relevant groups differed in their rates of responding during the compound on Sessions 4-9, $F(1, 600) > 4.54$, whereas the Integrated Colour-Relevant and Pattern-Relevant groups did not differ in this respect on any one session, $F_s(1, 600) < 1$. Other significant results from the ANOVA were the effect of session, $F(9, 540) = 52.15$, and a Relevant dimension x Session interaction, $F(9, 540) = 4.18$. The effects of integration, $F < 1$, and Relevant dimension, $F(1, 60) = 2.47$, and the Integration x Relevant dimension, $F(1, 60) = 1.78$, and Integration x Session, $F < 1$, interactions were not statistically significant.

Finally, inspection of Figure 4.15 suggests that, in keeping with the results from Experiments 5 and 6, extinction in the presence of the compound with a colour as the unique element proceeded at similar rates for all four groups. A three-way ANOVA for the ten sessions of Stage 2 with the factors of Integration, Relevant

dimension and Session revealed no effects of integration, $F(1, 60) = 2.89$, or Relevant dimension, $F(1, 60) = 1.55$, and no significant Integration x Relevant dimension, $F(1, 60) = 1.65$, Integration x Session, $F < 1$, Relevant x Session, $F(9, 540) = 1.26$, or Integration x Relevant dimension x Session, $F < 1$, interactions. The effect of session was statistically significant, $F(9, 540) = 97.47$.

Discussion. The results from Pattern-Relevant-Separate and the Colour-Relevant-Separate groups replicate the findings from the previous experiments. The implication is that it is possible to alter the associability of pattern stimuli by making them relevant or irrelevant with respect to colours for the solution of a discrimination. The results from the remaining two groups in the experiment demonstrate that this influence on the associability of patterns fails to occur when the patterns and colours that were used for the discriminations are presented in the same location. The results do not follow naturally from theories of learning that assume that associability changes to stimuli reside at the level of a central stimulus representation (e.g. Kruschke, 1992; LePelley, 2004; Mackintosh, 1975; Pearce, George and Redhead, 1998). According to this class of theory, it should be possible for just the pattern of a stimulus to lose associability even when the pattern of the stimulus is presented integrated with colour. The results of the current experiment, however, are perfectly in accordance with an account of associability which suggests it acts at a more peripheral level (e.g. Spence, 1936; 1937; Wyckoff, 1952). According to these accounts, orienting responses are directed towards relevant stimuli and away from irrelevant stimuli, however if the irrelevant and irrelevant stimuli occupy the same spatial location then such response selection will be hampered.

4.5. General Discussion

The three experiments in the chapter employed a novel technique for detecting changes in the associability of stimuli after they had been either relevant or irrelevant to the solution of a true discrimination, AX+ BX-. Higher associability was reported for previously relevant than irrelevant stimuli when they were patterns, but not when they were colours. Moreover, this difference emerged only when patterns and colours were presented separately on the monitor, rather than integrated in the same location.

One criticism that can be directed at studies of associability changes during discrimination learning is that they do not enable one to discriminate between the view that these changes affect the specific stimuli encountered during training (Mackintosh, 1975) or the entire stimulus dimensions involved (Sutherland & Mackintosh, 1971). With a method such as the IDS-EDS, this criticism is particularly difficult to counter because the transfer discrimination features novel stimuli from the same dimensions as the stimuli initially presented during training. Thus, the very strategy on which the method is based precludes knowledge of whether associability changes reflect some generalisation of associability between the stimuli initially encountered and the novel, test members along the same dimension or, alternatively, they involve tuning in or out entire dimensions. By providing a direct index of the associability changes undergone by specific stimuli, the behavioural technique here introduced should be capable of specifying in future studies which of these processes operates. Rather than using two sets of stimuli from two different dimensions, such an experiment would require the use of a set of stimuli from a single dimension.

A shortcoming of the experiments presented in this chapter is that they fail to reveal whether the associability of the patterns increases in the Pattern-Relevant

group, decreases in the Colour-Relevant group, or both. In fact, this has been long regarded as a common limitation to studies of acquired distinctiveness since the seminal studies by Lawrence (1949; Mackintosh, 1975; Hall, 1991). It is therefore important to highlight that the present method offers an opportunity to address this issue by incorporating a control group initially trained on a AX+ BW- CY+ DZ-discrimination, followed by the AY+ AX- CY- test discriminations. With this training, the two dimensions involved are equally relevant for the solution of the discrimination. Performance by this group at the transfer discriminations could thus provide a baseline against which to compare the associability changes resulting from true-discrimination training. In the experiments just reported, for example, inferior performance by this control group at the pattern-based test discrimination relative to the Pattern-Relevant group would indicate a rise in the associability of the patterns in the latter. Similarly, superior performance by this control group at the pattern-based discrimination in comparison with the Colour-Relevant group would indicate a drop in the associability of the patterns in the latter.

I have hitherto claimed that the behavioural technique presented in this chapter is capable of revealing associability changes, should they indeed take place. However, there are other explanations for why animals should apparently find a AY+ AX-discrimination more difficult than a AY+ CY- one after the training that I used. If it is assumed that the associative strengths of A and Y remain relative constant during the test phase (high for A and low for Y), and the solution of the AY+ AX- CY-discrimination therefore depends on the rates at which X and C, respectively, gain inhibition in one case and lose associative strength in the other, then two possibilities present themselves.

First, it is conceivable that the speed at which stimuli gain inhibition is slower than that at which they lose associative strength. This proposal only makes sense, of course, if the distances to be travelled by the stimuli in terms of associative change are equivalent, as could well be the case with the method under analysis. For example, in order to solve the pattern-based test discrimination in the experiments previously reported, the pattern that is the unique element of the nonreinforced compound must lose excitatory strength in the Pattern-Relevant group, and acquire inhibitory strength in the Colour-Relevant group. If stimuli acquire inhibition less readily than they extinguish excitation, then the difference reported in this chapter (Separate groups) with the pattern-based discrimination would naturally follow. To my knowledge, no empirical studies exist to evaluate this hypothesis. In any case, it is not easy to see how the rates of extinction and acquisition of inhibition could be compared in a manner that is not confounded with differences in associability. This follows because stimuli positioned at various points of the associative strength scale will inevitably differ in their relevance as predictors of a reinforcer. Even so, the null result observed in the first two experiments, and in the separate groups of the third experiment, with the colour-based test discrimination goes some way to challenge the foregoing analysis. If the results observed for the pattern-based discrimination are to be accounted for solely in terms of the development of faster extinction than inhibition, then a similar effect should have obtained for the colour-based discrimination. The fact that these two processes evolved with identical readiness across the groups for the latter discrimination detracts forcefully from this argument. Additionally, it serves to disarm the second alternative to an associability account for the present results.

This second explanation derives from the two articles by Rescorla (2001, 2002), which provided the rationale for the basic design used throughout this chapter.

In one of these experiments (Rescorla, 2001; see Table 4.4), animals received separate conditioning trials with two stimuli, A and C, intermixed with nonreinforced presentations of two other stimuli, B and D. Following this stage, further conditioning or, alternatively, extinction trials were administered with the compound formed by A and B. At test, conditioned responding during the compound AD was compared with that during BC.

Table 4.4. Basic design of Rescorla's (2001) experiments.

A+ C+	AB+	AD vs BC
B- D-	AB-	

If the Rescorla-Wagner (1972) model is correct, and changes in associative strength depend upon the summed error term of the CSs present, then A and B should each undergo identical changes in associative strength during AB trials. Accordingly, responding in the presence of AD and BC should therefore be expected to be the same. If, however, changes in associative strength are a function of the stimulus' associability (α), as suggested by Mackintosh's (1975) theory, then AB trials should modify A's associative strength to greater extent than B's, whether in conditioning or extinction. The results from the experiments contradicted both of these predictions. Instead, Rescorla found that extinguishing AB resulted in the animals responding more strongly in the presence of BC than AD, suggesting greater losses in associative strength for A than B. In contrast, conditioning with AB led to stronger responding to BC than AD, indicating greater gains of associative strength for B than A. Similar

results were found when B and D were established as conditioned inhibitors in the first stage (Rescorla, 2002). In view of these results, Rescorla argued that the magnitude of the changes in associative strengths accruing to the elements of a compound appears to be determined by the individual discrepancies between their own associative strengths and the outcome of the trial.

This finding has clear repercussions for the interpretation of the between-groups difference in the rate of extinction with the compound that had a pattern as the unique element. At the outset of the transfer discriminations, the associative strength of the nonreinforced pattern upon which the pattern-based discrimination hinges is higher in the Pattern-Relevant than the Colour-Relevant group. The discrepancy between this element's initial associative strength and the outcome on those trials ($\lambda = 0$) should accordingly be greater in the Pattern-Relevant than the Colour-Relevant groups. If the rate of extinction of the compound with a pattern as the unique element is assumed to depend to some extent on the discrepancy between that pattern's excitatory strength and the absence of reinforcement following those trials, then the present results can be accommodated by an account that makes no appeal to associability changes. As anticipated above, however, this extension of Rescorla's proposal is unable to predict why there were no comparable differences across the groups in the rate of extinction of the compound with a colour as the unique element.

Another alternative interpretation of the present results to an account appealing to associability changes relies on the notion of acquired distinctiveness or equivalence as championed by Honey and Hall (e.g. Honey & Hall, 1989; Hall, 1991, 1996). According to this view, stimuli along a dimension can undergo acquired distinctiveness or equivalence not because of a rise or drop in associability, but as a result of becoming associated with distinct or common associates, respectively. Thus,

following the training AX+ BX-, CY+ DY-, AW+ BW-, CZ+ DZ-, presented in the first phase of the preceding experiments, the dimension composed by cues A, B, C and D should experience a measure of acquired distinctiveness, while the dimension composed by W, X, Y and Z should suffer acquired equivalence. This prediction follows because cues A and C should become discriminable from cues B and D by virtue of being associated with different outcomes (the presence of reinforcement for A and C and its absence for B and D), while W, X, Y and Z should become less discriminable due to their being associated with a common outcome (partial reinforcement). If a transfer discrimination were then presented involving all these stimuli, it follows from this account that subjects would find it harder to discriminate between, say, W or X and Y or Z than between A or C and B or D. In the absence of some further assumption, however, it is difficult to predict with this account why, at test, the AY+ AX- discrimination should proceed less readily than the AY+ CY- discrimination. Even though, following the above reasoning, some measure of acquired distinctiveness is expected between X and Y, the same should apply to A and C, both equally associated with the presence of reinforcement.

A further alternative to the associability-based analysis of the results here reported can be derived from Stimulus Sampling Theory (SST; Estes, 1950, 1955, 1959). Rather than a monolithic body of theory, Stimulus Sampling Theory can best be characterised as a family of theories resting on a common set of fundamental assumptions (Neimark & Estes, 1967). According to this kind of account, all cues an animal experiences in a learning situation consist of a set of elements (Guthrie, 1935; Estes, 1950). On any given trial, only a subset of the elements forming the cues that impinge on the organism's receptors is sampled. The specific subset of elements being sampled fluctuates randomly from trial to trial. Those elements that happen to

be sampled acquire or lose, in an all-or-nothing manner, associative strength in accordance with their contingency with reinforcement and nonreinforcement. At the outset of conditioning, a small proportion of the elements sampled will possess associative strength, and consequently the strength of the CR evoked will be weak. As training progresses, however, the proportion of elements with associative strength being sampled will, as a matter of probability, increase, leading to stronger CRs. Thus, although one-trial learning is the rule at the elemental level, the nature of the sampling process ensures that the acquisition of a CR should typically be gradual.

It is worth considering what bearing the foregoing assumptions may have upon the interpretation of the current chapter's data, particularly where the colour and pattern dimensions were presented side by side. During acquisition, for instance, the Colour-Relevant groups consistently solved their respective true discriminations with greater ease than the Pattern-Relevant groups, and I attributed this finding to the colours being more salient than the patterns for the species used. A rather different explanation is provided by Stimulus Sampling Theory. According to this account, only a subset of the elements which form the colour and the pattern cues will —with equal probability—be sampled on any given trial. It is clear, however, that colours and patterns differed in complexity (compare the relative simplicity of the colour stimuli, made of uniform square patches varying in hue, with the more complex pattern stimuli). Perhaps, therefore, a relatively small number of elements constituting the colour cues will need to acquire associative strength and be concurrently sampled in order for the discrimination based on the colours to emerge during acquisition. In contrast, a larger proportion of elements endowed with associative strength will need to be sampled for the discrimination based on the patterns to manifest itself in behaviour.

It is less clear, however, how this kind of account may be brought to bear on the superior ease with which the Pattern-Relevant groups, relative to the Colour-Relevant groups, consistently solved the test discrimination based on the patterns. Comparisons between the groups during this stage were indeed concerned with the readiness with which test discriminations relying on a *common* stimulus dimension (i.e. patterns) were solved. Because sampling is, as stated above, assumed to be a random process, the probability that the elements which form the pattern representations will be sampled is the same in both groups. With no further assumptions⁽²⁾, therefore, the theory incorrectly predicts no difference in their performance during the test phase.

Whatever the merit of the preceding accounts in dealing with the apparent associability changes reported, the results from Experiment 7 demonstrate that a much simpler explanation in terms of orienting responses is available. The majority of attentional theories of learning assume that such changes in associability take place at a central level (e.g. LePelley, 2004; Mackintosh, 1975; Sutherland and Mackintosh, 1971). The results from Experiment 3 failed to support this claim because there was a change in the associability of the training stimuli when their components were presented separately for the true discrimination, but not when they were integrated. If changes in associability take place centrally, then it should not matter how the stimuli are presented in order for such changes to take place. These results, therefore, strongly suggest that the locus of the associability changes reported in this chapter lies at the response level rather than at the encoding level.

In view of this conclusion, it is relevant to consider whether all the demonstrations of associability changes in pigeons are subject to the same interpretation. Experiments conducted with pigeons in our laboratory have previously

demonstrated changes in associability, but they employed stimuli that were presented in different locations, which makes it impossible to reject the possibility that they were a consequence of a change in the extent to which subjects oriented towards relevant stimuli and away from irrelevant stimuli (e.g. George and Pearce, 1999; Pearce, George & Redhead, 1998). There is a successful demonstration of the ID-ED effect by Mackintosh and Little (1969) which used integrated stimuli of lines that varied in both orientation and colour. It is hard to see how birds could orient in a way that would allow them to perceive, say, the orientation of a line without noticing its colour. Perhaps, therefore, this study provides evidence that associability changes take place at the encoding level. This conclusion should, however, be treated with a measure of caution because both Couvillon, Tennant & Bitterman (1976) and Hall & Channell (1985) have failed to replicate the effect by Mackintosh and Little (1969). Until this effect has been shown to be reliable, it would be prudent to keep an open mind as to whether associability changes in the pigeon occur centrally.

If the changes in the associability of patterns reported in Experiments 5 and 6 were a consequence of subjects learning to orient towards some stimuli rather than others, then the mechanism that accounts for this change in orienting activity needs to be specified. An obvious candidate is that these responses were acquired through instrumental conditioning (e.g. Wyckoff, 1952). If a subject is exposed to an AX+ BX- discrimination and orients towards stimulus A when it is presented, then the occurrence of food consistently at the end of the trial would be expected to strengthen this response. On the other hand, a similar response directed towards X will fail to be strengthened to the same extent because of the partial reinforcement schedule with which it is associated. An intriguing implication of this analysis is that even though B is relevant to the solution of the discrimination, subjects will orient less towards this

stimulus than to X, or A, because it is never followed by food. To my knowledge this prediction has never been tested. Another implication of this account of the orienting response is that it should apply to any visual stimulus and not just patterns. Of course, this implication is challenged by the failure to find a difference in the associability of colours when they were relevant or irrelevant to the solution of the true discrimination. One possible explanation for this result is that the coloured stimuli were so salient that they elicited an unconditional orienting response that could not be modified by experience. This explanation is similar in logic to that provided by Sutherland and Mackintosh (1971; see also LePelley, 2004; Suret & McLaren, 2003), who proposed that it is difficult to alter the associability of stimuli that have a high conditioned, or unconditioned salience.

An explanation of the present results in terms of orienting responses, however, is not theoretically committed to an instrumental account. It would be sufficient to assume that stimuli with high associative strength will, as part of their classically-elicited responses, be more likely to be gazed at than those having low associative strength. At the start of acquisition training, pigeons in both groups will spend more time looking, and therefore, learning about, the colours than the patterns due to their higher unconditioned salience. In the course of training, occasional inspection of the patterns will bring the birds in the Pattern-Relevant-Separate group, but not the Colour-Relevant-Separate group, into contact with the relevant contingencies with reinforcement, ultimately leading to the solution of the true discriminations in the former group. These birds will gradually come to spend more time looking at the patterns and will, upon transfer, be at an advantage with respect to the birds in the Colour-Relevant-Separate group, for which the initial tendency to look at the colours could only have intensified during acquisition.

It should be noted that it does not necessarily follow from the results of Experiments 5, 6, and those with the Separate groups in Experiment 7 that the changes in associability recorded were a consequence of alterations in a disposition to orient towards patterns. Changes in associability of the sort envisaged by Mackintosh (1975) might have been responsible for these findings. If this possibility is correct then there remains the problem of explaining why the attentional effects were not observed when the colours and patterns were presented together. It is no doubt possible to devise an explanation for this result by appealing to a central attentional mechanism. However, for it to be satisfactory, this explanation would need to be supported by additional evidence casting doubt on the claim that the associability effects we have reported were driven by changes in orienting activity.

To conclude, by using a novel method the present experiments have confirmed that changes in associability to stimuli take place during true discrimination learning. However, these changes were restricted to stimuli which are of low rather than high unconditioned salience, and they appear to be a consequence of changes in a disposition to orient towards certain stimuli. It remains to be determined whether changes in associability in pigeons can also operate at a more central level.

Notes

(1) It would be more precise to say that this prediction is *in the spirit* of the theory of Mackintosh (1975). Strictly speaking, the theory is unable to predict extinction of the compound AX because to do so X would need to become a conditioned inhibitor, and no adequate mechanism for developing inhibition was provided by Mackintosh. For a more recent extension of the theory that overcomes this problem see LePelley (2004).

(2) Note that more recent versions of Stimulus Sampling Theory (e.g. LaBerge, 1995) have eliminated the randomness of the sampling process originally assumed in the theory in order to accommodate the body of data suggesting the involvement of some sort of attentional mechanism. In connection to the present experiments, such versions of SST have no difficulties in yielding predictions equivalent to those made by other attentional theories of associative learning, such as Mackintosh's (1975).

Chapter 5

5. Analysis of the attentional theory of Mackintosh (1975)

5.1. Experiment 8

The three studies in Chapter 4 provided evidence indicating that associability changes in pigeons can occur to patterns trained in the presence of colours during true discrimination learning. Experiment 7 of the same chapter lent support to the contention that, at least with the parameters and species used, associability changes appear to be the consequence of receptor-orienting acts, rather than of some central mechanism modulating attention. At any rate, one clear result of these experiments was the failure to reduce the associability of the colours when they were the irrelevant dimension in a true discrimination, whether presented separately or integrally with the patterns. In the absence of a central or peripheral mechanism to modulate the processing power devoted to colours, we can now consider whether changes in associability are responsible for the relative validity effect in pigeons (Wasserman, 1974).

According to Mackintosh's (1975) theory, relative validity results from a failure to attend, and thereby learn about, the common element X in the true discrimination (AX+ BX-), but not in the pseudo discrimination (AX+/- BX+/-). More specifically, the associability of X will decrease in the true discrimination because X is a worse predictor of reinforcement than A and of nonreinforcement than B. By contrast, in the pseudo discrimination X is no worse a predictor of reinforcement or its absence than A or B, and as a consequence its associability will remain relatively stable. If this theory provides a complete account of relative validity, no such effect should be observed if it is impossible to reduce the associability of X. The purpose of

the single experiment contained in this chapter is to test this proposal using a similar autoshaping procedure to that of Experiment 6 in Chapter 4. The design of the experiment is shown in Table 5.1.

Table 5.1. Experimental design of Experiment 8.

Group	Acquisition	Test
Colour-True	vertical & blue + horizontal & blue –	blue? green?
Colour-Pseudo	vertical & green + horizontal & green –	
	vertical & blue +/- horizontal & blue +/-	
	vertical & green +/- horizontal & green +/-	
Pattern-True	blue & vertical + green & vertical-	vertical? horizontal?
Pattern-Pseudo	blue & horizontal + green & horizontal -	
	blue & vertical +/- green & vertical +/-	
	blue & horizontal +/- green & horizontal +/-	

Two groups of pigeons were trained on two true discriminations of the form AX+ BX- and AY+ BY-. For the Colour-True group, the common elements X and Y were colours and A and B were patterns, whereas for the Pattern-True group X and Y were patterns and A and B were colours. Two control groups received pseudo-discrimination training of the form AX+/- BX+/- AY+/- BY+/- . The Colour-Pseudo group was trained with X and Y as patterns and A and B as colours. For the Pattern-Pseudo group, X and Y were colours and A and B were patterns. In all groups,

patterns and colours were presented side by side on the television screen. This choice of separate over integrated compound stimuli was imposed by the need to present subsequently the common elements of the compounds for test. Following acquisition training in the four groups, intermixed test trials with X and Y by themselves were administered. If X and Y elicited significantly stronger conditioned responding in the Pattern-Pseudo and Colour-Pseudo groups than the Pattern-True and Colour-True groups, respectively, then this would constitute evidence for relative validity.

If a relative validity effect should be found, a second purpose of the present experiment was that of comparing the magnitude of this effect when X and Y are patterns with that when they are colours. From the studies in the previous chapter, we know that the separate presentation of these dimensions on a display leads to the patterns, but not the colours, showing an associability change. As mentioned in Chapter 4, these experiments do not allow us to determine whether these changes are the consequence of an increase in the propensity to look at the patterns for the Pattern-Relevant groups, a decrease in that propensity for the Colour-Relevant group, or both. It is conceivable, however, that in the present experiment the tendency to look away from the patterns could, by the end of training, be higher in the Pattern-True group, for which the colours were relevant, than in the Pattern-Pseudo group, for which the colours were as relevant as the patterns. Insofar as this conjecture is correct, it would not be surprising to find the response-evoking properties of the patterns to be lower in the Pattern-True than the Pattern-Pseudo group, producing an artefact of relative validity.

A rather more telling result would come from the observation of relative validity between the Colour-True and Colour-Pseudo groups. Experiments in Chapter 4 consistently showed that a similar training to that given in the Colour-True group

does not produce any impairment in the colours' capability to serve in the solution of a transfer discrimination. At test, a response deficit in the presence of the colours for the Colour-True but not the Colour-Pseudo group would therefore imply that the relative validity effect does not depend on a change in the associability of the irrelevant cue.

Method

Subjects and Apparatus. The subjects were thirty-two experimentally naïve adult homing pigeons from the same stock and maintained in the same way as for the experiments in Chapter 4. Following autoshaping, the pigeons were matched on their rate of responding to the white circle and assigned in equal numbers to the four groups. The apparatus was the same as for the experiments in Chapter 4.

Stimuli. Four of the stimuli used for Experiment 6 in Chapter 4 were used in the current experiment. These were, for all groups, dark blue and vertical lines, light green and vertical lines, dark blue and horizontal lines and light green and horizontal lines. When one component of a compound stimulus was presented in isolation for testing (e.g. vertical lines), it was presented in the centre of the monitor.

Procedure. Magazine training and autoshaping with a white circle were conducted in the same manner as for the experiments in Chapter 4. In each of the subsequent seventeen sessions, birds in the Pattern-True and Colour-True groups received training that was identical to the training given, respectively, to the subjects in the

Colour-Relevant and Pattern-Relevant groups in Experiment 6 with the exception that only four trial types with AX+, BX-, AY+ and BY- were used. Animals in the Colour-Pseudo and Pattern-Pseudo groups were exposed to exactly the same stimuli, but for these groups each compound was paired with food on half of the trials. Again, there were a total of 64 trials per session, with each of the four stimuli presented 16 times. The first thirty-two trials of session eighteen were identical to the previous training sessions. For the remainder of the session, the stimuli that served as X and Y were each presented three times, and were nonreinforced. The reason for administering so few test trials was to avoid the confounding influence of the partial-reinforcement extinction effect (e.g. Haselgrove, Aydin & Pearce, 2004) on the interpretation of the results. For the Colour-True and Colour-Pseudo groups X and Y were colours, and for the Pattern-True and Pattern-Pseudo groups, they were patterns. Any procedural details omitted were the same as for Experiment 6 in Chapter 4.

Results. For the purpose of plotting and analysing the data from the acquisition stage, the mean of reinforced trials and the mean of nonreinforced trials were calculated for the Pattern-True and Colour-True groups over the nine 2-session blocks. For the Pattern-Pseudo and Colour-Pseudo groups, the mean of all four partially-reinforced trials was calculated. The Upper and Lower Panels of Figure 5.1 show the results from this stage, in 2-session blocks, for the Pattern-True and Pattern-Pseudo groups and for the Colour-True and Colour-Pseudo groups, respectively. It is evident that the Pattern-True and Colour-True groups both solved their respective discriminations and, consistently with previous experiments (see Chapter 4), these discriminations were easier when based upon different colours (Pattern-True group) than different patterns

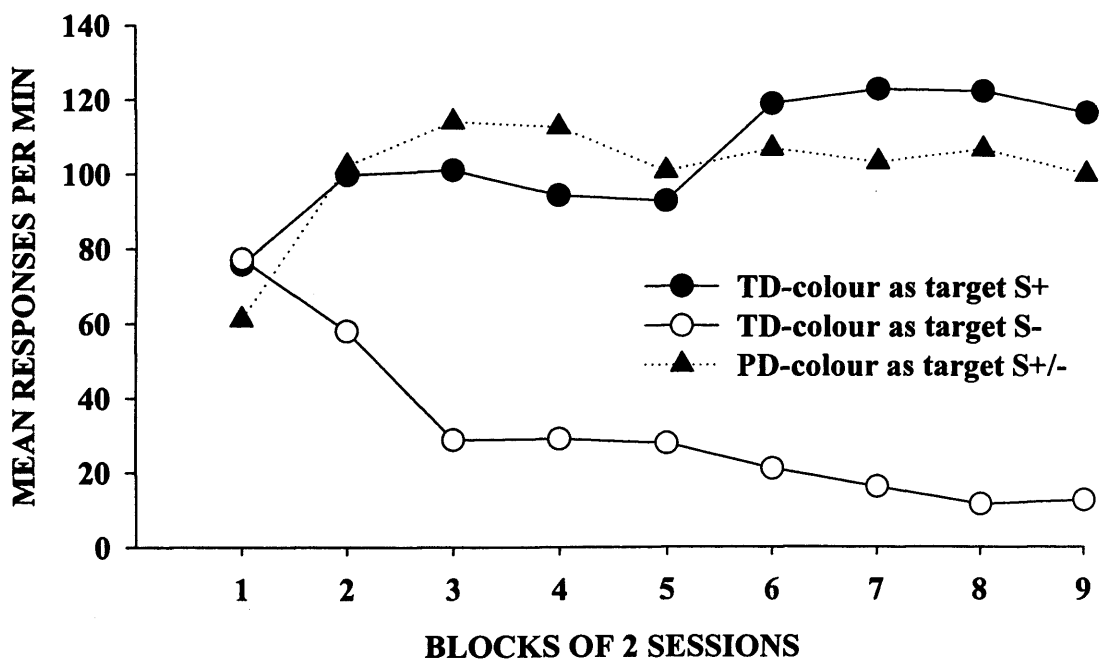
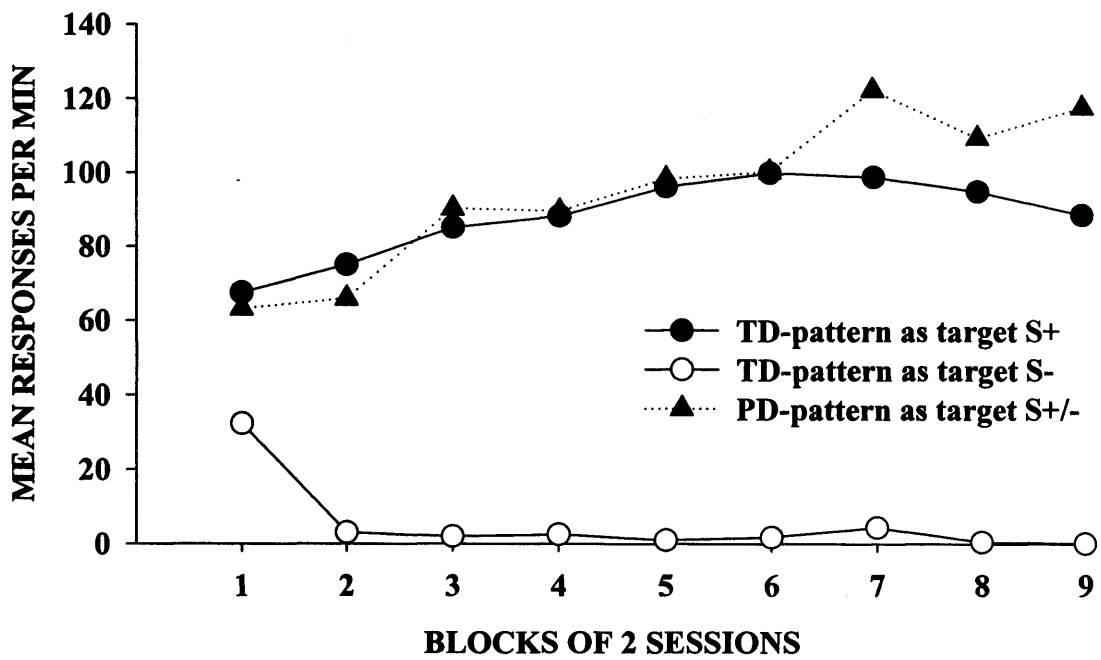


Figure 5.1. Acquisition in the Pattern-True and Pattern-Pseudo groups (Upper Panel) and the Colour-True and Colour-Pseudo groups (Lower Panel) in Experiment 8.

(Colour-True group). These observations were confirmed by a three-way ANOVA of individual mean number of responses per minute with the factors of group (Pattern-True or Colour-True), CS (reinforced or nonreinforced) and session block for the nine session blocks of training, which revealed a significant Group x Session block x CS interaction, $F(8, 112) = 2.11$. Analysis of simple main effects conducted to explore this interaction, however, revealed no significant Group x CS interaction on any session block, $F_s(1, 126) < 2.55$. Further results from this analysis also revealed that the Pattern-True group (colour-based discriminations) performed differently in the presence of reinforced and nonreinforced compounds on each of the nine session blocks, $F_s(1, 126) > 4.71$, and so did, except on the first session block, $F < 1$, the Colour-True group (pattern-based discriminations), $F_s(1, 126) > 6.67$. The remaining results from the ANOVA were an effect of CS, $F(1, 14) = 72.88$, and a significant Session block x CS interaction, $F(8, 112) = 14.38$. The effect of group just failed to reach the conventional level significance, $F(1, 14) = 4.30$, $p = 0.057$. Moreover, the effect of Session block and the Group x Session block and Group x CS interactions were not significant, $F_s < 1$.

A glance at the Pattern-Pseudo and Colour-Pseudo groups indicates that for both conditioning proceeded smoothly. When their performance is compared with that on the reinforced trials in their corresponding true-discrimination groups, there is a suggestion towards the end of training that the Pattern-Pseudo group responded more vigorously than the Pattern-True group, whereas the Colour-Pseudo group responded more weakly than the Colour-True group. In order to investigate these relations, a three-way ANOVA of individual mean number of responses per minute during the reinforced compounds for the two true-discrimination groups and the partially reinforced compounds for the two pseudo-discrimination groups was conducted for

the nine session blocks. The ANOVA included the factors of discrimination (True or Pseudo), test stimulus (Pattern or Colour) and session block. The most relevant outcome from this analysis was a significant Discrimination x Test stimulus x Session block interaction, $F(8, 224) = 2.07$. Exploration of the three-way interaction by means of simple main effects analysis revealed, however, that there was no significant Test stimulus x Discrimination interaction on any session block, $F_s(1, 252) < 2.77$. Furthermore, neither the Pattern-True differed from the Pattern-Pseudo group, $F_s(1, 252) < 2.25$, nor the Colour-True differed from the Colour-Pseudo groups, $F_s(1, 252) < 1.02$, any session block. This analysis also revealed that the Pattern-True and Colour-True groups did not differ in their rates of responding on any session block, $F_s(1, 252) < 2.06$, and neither did the Pattern-Pseudo and Colour-Pseudo groups, $F_s(1, 252) < 3.50$. Other results from the ANOVA were an effect of session block, $F(8, 224) = 11.35$, but no significant effects of test stimulus or discrimination, $F < 1$, and no significant Test stimulus x Session block, $F(8, 224) = 1.42$, Test stimulus x Discrimination or Discrimination x Session block, $F_s < 1$, interactions.

The results from the test with the common elements X and Y appear in Figure 5.2. The combined mean number of responses per minute in the presence of both stimuli was calculated and collapsed across test trials. Responding was faster in the pseudo-discrimination groups than in the true-discrimination groups, irrespective of whether colours or patterns were being presented. These relative validity effects with both patterns *and* colours were confirmed by a two way ANOVA of individual mean number of responses per minute, which included the factors of discrimination (True or Pseudo) and test stimulus (Colour or Pattern). This analysis revealed an effect of discrimination, $F(1, 28) = 29.73$, but no effect of test stimulus, $F < 1$, and no interaction between these factors, $F(1, 28) = 2.70$.

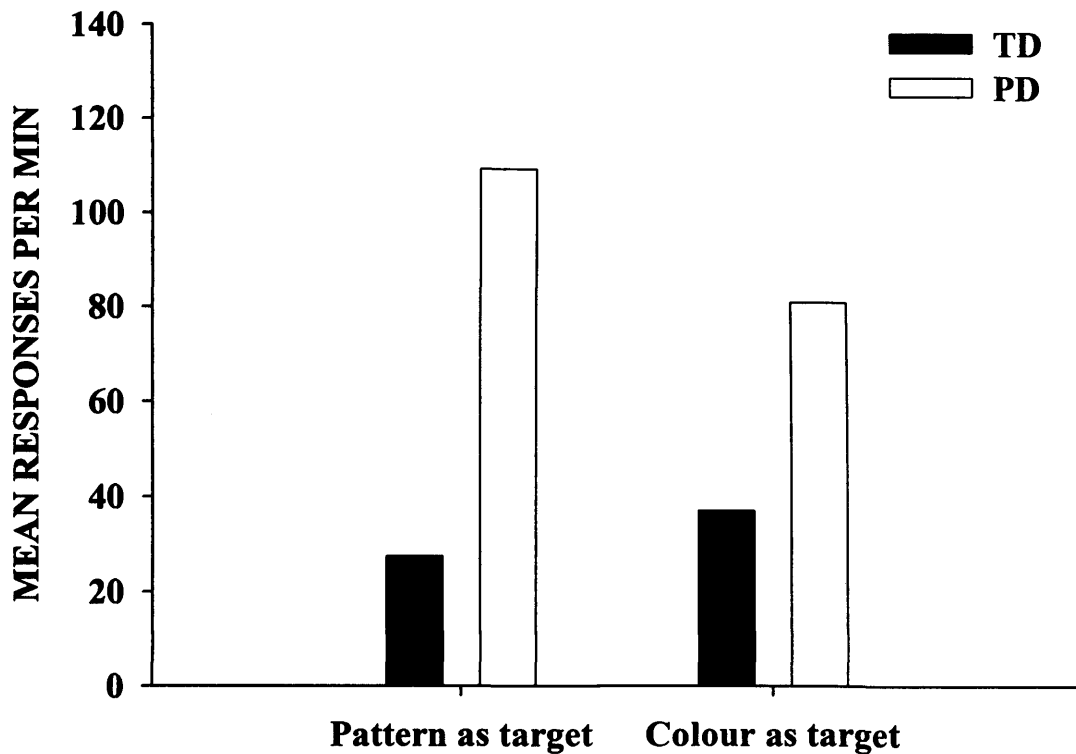


Figure 5.2. Responding during the test with the common elements (X & Y) in Experiment 8.

Discussion. The results from this experiment have implications for our understanding of the relative validity effect. Wagner et al. (1968) compared the effects on conditioning with X of a true discrimination, AX+ BX- and a pseudo-discrimination, AX+/- BX+/- . In keeping with the similar design that was employed here, they found that responding during X was stronger after training with the pseudo-discrimination than the true discrimination (Wasserman, 1974 for a report of the effect in pigeons). According to the theory of Mackintosh (1975), this effect is a consequence of X being paid less attention in the true than the pseudo-discrimination group. However, it was shown in Chapter 4 that when X is a colour its associability is not reduced by training

with an AX+ BX- discrimination. On this basis, therefore, the theory of Mackintosh (1975) predicts that responding to the colours serving as X should have been the same during the test trials with the Colour True and Colour Pseudo groups. The finding that responding was lower in the Colour True than the Colour Pseudo group clearly implies that some factor other than a change in associability is responsible for relative validity.

The foregoing conclusion is based on the assumption that the design of the current experiment offered the same opportunity for associability changes to occur as the designs of the experiments in Chapter 4. It should be noted, however, that the latter involved eight different training trials (AX+ BX-, CY+ DY-, AW+ BW-, CZ+ DZ-), whereas the current experiment only used four different training trials (AX+ BX-, AY+ BY-). It is thus conceivable that because of the different training there was a loss of attention to X and Y in the present experiment both when they were patterns (Pattern-True group) and when they were colours (Colour-True group). This possibility cannot be fully dismissed in the absence of further studies. However, a suggestion that the associability processes that occurred to these stimuli were in all probability the same across all these studies comes from the following fact. That is, that in accordance with the results from the experiments in Chapter 4, the acquisition of the true discriminations in the current experiment was acquired more readily when colours rather than patterns were the relevant stimuli.

A feature of the theory of Mackintosh (1975) is that the acquisition of associative strength by a stimulus is determined by the discrepancy between its own associative strength and the value of the reinforcer with which it is paired. Other theories assume that the associative strength of X will be related to the discrepancy between the combined associative strength of all the stimuli that are present on a trial

and the value of the reinforcer (e.g. Rescorla & Wagner, 1972; Pearce, 1987, 1994). These theories are able to explain the relative validity effect by virtue of this assumption, and it is worth considering whether they can also explain the results from the present experiment. In particular it is worth considering the predictions made by these theories when the salience of A and B is different to that of X. Computer simulations based on the equation provided by Rescorla and Wagner (1972) revealed that the relative validity effect will be found when A and B are either of greater or lower salience than X. In this respect, the results reported here are compatible with predictions from the theory. The effect was found both when X was a pattern (low salience) and A and B were colours (high salience), and when these relations were reversed. In addition, the simulations revealed that responding to X will be stronger when it is more salient than A and B, rather than less salient. This is true for both the true discrimination and the pseudo discrimination. In the present experiment, X was more salient than A and B in the Colour-True and Colour-Pseudo groups and, accordingly, responding in these groups should have been higher than in the Pattern-True and Pattern-Pseudo groups, respectively. Inspection of Figure 5.2 shows that this prediction was confirmed for the true discriminations, but for the pseudo discriminations the opposite outcome was found. These differences were not statistically significant, however, which makes it difficult to draw any firm conclusions from them. Furthermore, the difference between the results for the two pseudo-discrimination groups may have been a consequence of the difference between the response rates that were recorded during the final sessions of training. Inspection of Figure 5.1 reveals that by the end of training responding was more rapid in the Pseudo-Pattern than the Pseudo-Colour group.

On balance, therefore, the present results are inconsistent with the proposal (e.g. Mackintosh, 1975) that the relative validity effect results from a loss of associability by the target cue in the true discrimination group. In addition to their relevance for our understanding of this cue-competition effect, these results shed light on a second aspect of Mackintosh's (1975) theory. Mackintosh (1975) suggested the possibility that the associability of a CS might not only determine the rate at which its associative strength changes, but also—in a direct manner—the magnitude of the CR it evokes. On this view, a stimulus with high associability should, even if only moderately correlated with reinforcement, tend to elicit a strong CR. In most circumstances, stimuli putatively high in associability are well correlated with reinforcement, thereby making it impossible to dissociate the contributions of associability and associative strength to the strength of conditioned responding. Experiment 8 provided, in combination with the experiments in Chapter 4, a rare instance in which such dissociation could be achieved. Contrary to Mackintosh's (1975) proposal, the results showed that a stimulus with high associability can, at least under certain circumstances, be incapable of supporting substantial conditioned responding.

Chapter 6

6. General Discussion

In this thesis, I have tested two influential accounts of stimulus selection phenomena: the Comparator Hypothesis (Miller & Matzel, 1988; Denniston et al., 2001) and the attentional theory of Mackintosh (1975). To this effect, I have compared a number of novel predictions derived from each of these classes of theory with those derived from Standard Associative Theory (e.g. the Rescorla-Wagner, 1972, model). Because Chapters 2 and 3 were dedicated to the analysis of the Comparator Hypothesis, whereas Chapters 4 and 5 were occupied with the analysis of the theory of Mackintosh (1975), I next summarise the results pertaining to each theory separately.

6.1. Analysis of the Comparator Hypothesis (Schachtman & Miller, 1985; Miller & Matzel, 1988; Denniston et al., 2001, Savastano et al., 2003)

In Chapter 2, experiments tested two predictions made by the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988), which conflict with those that can be derived from the Rescorla-Wagner (1972) model. The first of these predictions concerns the influence that an inhibitor B may have on the response-eliciting properties of an excitor X after training with an X+ AX- discrimination. Experiment 1 tested this prediction by embedding the two cues in a true discrimination AX+ BX- and measuring the extent to which X overshadows A. The magnitude of the CR in the presence of A was compared with that in a feature-positive control group in which B was absent (AX+ X-).

The second prediction tested in this chapter concerns the outcome of single-phase blocking $A^+ AX^+$ with extended training. According to the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988), the extent to which A is able to reduce responding to X at test following single-stage blocking should tend to even out as training proceeds with respect to an overshadowing control were X is presented the same number of times. This is because A and X should gradually approach the same asymptotic associative strengths in both groups, given their continuous reinforcement schedule. Once the balance of associative strengths formed by A and X is matched across the groups, then A will serve as an equally effective comparator stimulus in both groups and no blocking will be observed. In fact, strictly speaking, the opposite effect of blocking is anticipated by the theory, because the degree to which A can act as a comparator for X is also determined by their within-compound association, which should be weaker in the blocking group by virtue of intermixing A^+ and AX^+ trials. In whatever form, this prediction of the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel 1988) was disconfirmed in Experiment 2: extensive training served to enhance, not abolish, single-stage blocking.

Further evidence against the account of cue competition advanced by the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) was provided in Chapter 3. In this case, the prediction tested concerned the effects that adding a cue X to a previously established excitator A has on the response-eliciting properties of the latter. According to the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988), the added cue X will form associations with reinforcement and with stimulus A, and therefore should in time come to act as a comparator for A. The extent to which the addition of X will reduce responding to A

is of course an empirical question, but it is certain that on this account the presence of X should at best be neutral, if not detrimental, to the magnitude of the CR evoked by A. Contrary to this prediction, Experiment 3 showed an increase in responding during A, by comparison with the control cue B, that was contingent upon the addition of X. Moreover, Experiment 4 provided some evidence consistent with the view that this enhancement in the response-evoking properties of A was mediated by its within-compound association with X. These results stand in stark contrast with previous studies showing recovery from overshadowing (Matzel, Schachtman, & Miller, 1985; Dickinson & Charnock, 1985; Kaufman & Bolles, 1981) and blocking (Blaisdell, Gunther, & Miller, 1999), in which extinction of the comparator stimulus typically leads to an increase in the strength of the CR evoked by the target cue.

One way in which advocates of the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) could attempt to accommodate the present results is by arguing that within-compound associations can, under certain circumstances, facilitate rather than disrupt the expression of a target cue's association with reinforcement. In fact, Stout and Miller (2007) have recently acknowledged this possibility when dealing with learning effects which seem to depend upon the facilitating influence of within-compound associations, such as second-order conditioning and sensory preconditioning. However, they limit the circumstances in which within-compound associations may play a role contrary to that envisaged in the comparator process to the early conditioning trials. The large amount of training with A and B both before and after the addition of X, therefore, detracts from applying this reasoning to the experiments in Chapter 3.

Before discussing some future directions in the analysis of the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988), it is worth recalling

that all the predictions tested in Chapters 3 and 4 are equally problematic for the extended version of the theory (Denniston et al., 2001; Savastano et al., 2003).

6.2. Future directions in the analysis of the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988; Denniston et al., 2001)

The experiments contained in Chapters 2 and 3 are of theoretical importance because they reveal some basic flaws in the fundamental tenets of the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988). These flaws refer to 1) the fate of blocking over asymptotic training, 2) the influence that an inhibitor may have on the excitator that fostered its inhibition and 3) the influence that adding a novel cue may have upon the excitatory status of a pretrained CS. In the future, the theory can be further assessed by testing a number of predictions in conflict with those derived from the Rescorla-Wagner (1972) model. For instance, one aspect of the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) that deserves attention is its fully elemental treatment of associative learning. Consider, for instance, the acquisition of a positive patterning discrimination A- B- AB+ and that of a negative patterning discrimination A+ B+ AB-. According to this account, A and B are partially reinforced in both discriminations, and will therefore acquire the same amount of associative strength. Moreover, since A and B are presented by themselves and in compound the same number of times in both discriminations, their within-compound association will be equally strong. Without any further assumption, therefore, the theory seems constrained to predict that the comparator process will be equally detrimental to the expression of the associations that A and B will respectively form with the US. It follows from this account that A and B should evoke

as strong a CR during training with a positive-patterning as during training with a negative-patterning discrimination. This is, of course, equivalent to saying that neither of these discriminations should ever be solved.

This drawback, common to all strictly elemental theories of learning, is addressed by the Rescorla-Wagner (1972) model by postulating the existence of “unique cues” that represent the presentation of two or more cues in compound (e.g. Rescorla, Grau & Durlach, 1985). Unique cues are allowed to gain associative strength in the same way as any regular stimulus. Thus, training with a positive-patterning discrimination will lead the unique cue formed by the compound presentation of A and B to eventually gain all the positive associative strength available on AB+ trials. This is because the elements A and B will tend to lose associative strength when presented in extinction by themselves. In this manner the model is able to predict the solution of a positive-patterning discrimination. By symmetry of reasoning, the unique cue “AB” formed on AB- trials during training with a negative-patterning discrimination will acquire a sufficient amount of inhibition to counter the summed excitation caused by the conjoined presentation of A and B. This allows the model to predict the extinction of the CR that the presentation of the excitors A and B would otherwise evoke.

Endowing the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) with the assumption of unique cues appears to be of no avail as far as the solution of positive- and negative-patterning discriminations goes. Consider, for argument’s sake, the case of a positive-patterning discrimination. If allowance is made for the presence of a unique cue X on AB+ trials, the extended version of the theory (Denniston et al., 2001; Savastano et al., 2003) appears better equipped to deal with the complex interactions arising from the insertion of this second comparator

stimulus. On A- trials, for example, cue A will retrieve, in addition to a directly-activated representation of the US, two indirectly-activated US-representation by virtue of its within-compound association with of each of B and the unique cue X, its comparator stimuli. However, since the B and X cues will additionally serve as second-order comparator stimuli for each other, they will tend to neutralise one another's detrimental influence on the expression of the A-US association. Application of the same reasoning when cue B is the target stimulus will lead to the conclusion that, on this account, responding should never cease on A- and B- trials.

Negative-patterning seems equally insoluble for an Extended Comparator Hypothesis (Denniston et al., 2001, Savastano et al., 2003) that included the unique cue X as a potential comparator. Because the compound AB is never reinforced in this case, X will acquire a minimal amount of associative strength and will therefore have little impact in the development of the discrimination. Because B will serve as the sole comparator for A on A+ trials and vice versa, the theory predicts that responding on A+ and B+ trials will be only as high as their respective comparator processes allow. Paradoxically –due to the absence of second-order comparators—, responding on A+ and B+ trials are actually predicted to be lower than that observed on the A- and B- trials of a positive-patterning discrimination.

In the face of such criticisms, the advocate of the Comparator Hypothesis might argue that positive- and negative-patterning, like other examples of configural learning, lie outside the range of phenomena that the theory was designed to account for. Indeed, Miller and associates have typically described the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) as an alternative account of stimulus-selection phenomena (e.g. Rescorla, 1968, Wagner et al., 1968) to the more traditional learning-deficit views. It seems imperative, however, that any theory of

associative learning attempting to characterise the way in which CSs interact with each other should incorporate a mechanism of configural learning. After all, instances of configural learning have been known since the dawn of research on conditioning (e.g. Pavlov, 1927; Woodbury, 1943).

Yet configural-learning phenomena do not uniquely illustrate the failure of the Comparator Hypothesis (Miller & Schachtman, 1985; Miller & Matzel, 1988) to account for the development of certain types of discrimination. It is not obvious, for example, how the theory could predict that subjects will stop responding in the presence of the nonreinforced compound AX during the solution of a feature-negative discrimination of the form A+ AX-. As a result of this training, the partially-reinforced cue A should on this account acquire a moderate amount of associative strength, whereas X should acquire negligible associative strength. It should be recalled that the comparator process is proposed to take place at the time of testing – as opposed to during training—a target CS; and, crucially, in the absence of the comparator stimulus (Miller & Matzel, 1988; Schachtman, Brown, Gordon, Catterson & Miller, 1987). Assuming therefore that no comparator process takes place on AX-trials, it seems clear that a simple response rule based on the combined associative strengths of A and X will never lead to the complete extinction of responding on those trials. To conclude, problems arise for the theory from its failure to specify an adequate set of rules to govern responding in the presence of compound stimuli during training.

6.3. Analysis of the attentional theory of Mackintosh (1975)

Chapters 4 and 5 of this thesis were dedicated to the analysis of Mackintosh's (1975) theory of attention. In Chapter 4, a novel behavioural technique designed to reveal associability changes in true-discrimination learning was tested in pigeons. Colours and patterns were used in three experiments as relevant and irrelevant stimuli, appropriately counterbalanced. In support of Mackintosh's (1975) theory, Experiments 5 and 6 revealed between-groups associability changes with the patterns but not the colours. Since the stimuli used in Experiments 5 and 6 were contained in squares which stood side by side on the centre of the monitors, the possibility was explored that some kind of receptor-orienting act might be responsible for the associability changes observed. To test this hypothesis, Experiment 7 compared the effect obtained when the stimuli from the colour and pattern dimensions were presented side by side, as in Experiments 5 and 6, with that obtained when they were integrated on the same location on the monitors. Experiment 7 confirmed the existence of what appeared as associability changes to patterns when the dimensions were presented separately, but revealed no such effect when they were integrated. These results, therefore, are important because they demonstrate the need to rule out peripheral explanations to associability changes before accepting support for attentional theories of discrimination. While a peripheral account based on orienting responses is compatible with nonattentional theories in the continuity tradition, it "lies outside the scope" of the theory of attention proposed by Mackintosh (1975, p. 296). These results, moreover, expand on previous reports with autoshaping procedures indicating that pigeons may solve visual discriminations by means of receptor-orienting acts (e.g. Jenkins & Sainsbury, 1969; Jenkins, 1973).

One clear outcome from all of the experiments contained in Chapter 4 is that true-discrimination training does not result in differences of associability of any kind when colours are relevant with respect to when they are irrelevant. Chapter 5 explored this result further by addressing a question of significant theoretical interest: whether associability changes can provide a complete account of relative validity in pigeons (e.g. Wasserman, 1974). In Experiment 8, two relative-validity preparations in pigeons were compared using a procedure similar to that used in Chapter 4. Relative validity was found both when the patterns and the colours were the target stimulus. These results are inconsistent with the view (e.g. Mackintosh, 1975) that relative validity is produced by differences in the associability of the target cue between the true discrimination and pseudodiscrimination. Thus, although —peripheral— associability changes may occur during true discrimination learning in pigeons, they are not the sole or even the main mechanism driving the occurrence of a relative validity effect in this species.

6.4. Future directions in the analysis of Mackintosh's (1975) theory

The experiments contained in Chapter 4 provide the first demonstration of associability changes yielded by the novel behavioural technique presented in this thesis. The spatial integration of the stimulus dimensions conducted in Experiment 7, moreover, shed light on the peripheral locus of the mechanism driving these changes. This failure to obtain associability changes with integrated stimuli joins those of Couvillon et al. (1976) and Hall and Channell (1985) using an IDS-EDS method in pigeons. To date, only one experiment with pigeons stands out as having reported associability changes using integrated stimuli from two visual dimensions

(Mackintosh & Little, 1969). It is nonetheless possible, for the present case, that centrally-generated associability changes of the kind envisaged by Mackintosh's (1975) theory would have surfaced had the test discrimination in the second stage been more demanding. This is equivalent to recognising that perhaps the readiness with which the integrated groups solved their transfer discriminations in Experiment 7 may have concealed actual differences in associability.

One way to test this prediction is to superimpose a further test discrimination based on previously nonreinforced compounds on that thus far employed. Thus, after receiving training in the first stage with AX+ BX- CY+ DY- AW+ BW- CZ+ DZ-, subjects could be confronted with a BZ- BW+ DZ+ test discrimination in addition to the AY+ AX- CY- discrimination. If associability effects occur, then the component discriminations based on previously relevant stimuli, BZ- DZ+ and AY+ CY-, should be solved more readily than those based on previously irrelevant stimuli, BZ- BW+ and AY+ AX-. This strategy presumes that, in keeping with Mackintosh's (1975) theory, the best predictors of both reinforcement (A and C) and nonreinforcement (B and D) will secure higher degrees of associability than relatively poor predictors (X, Y, W and Z). Thus, as well as maximising the chances to observe centrally-generated associability changes, this variant of the basic design provides a further test for the theory. Whatever fruit the application of this technique may bear in pigeons, it is of great theoretical interest to replicate this work in mammals. Some speculation exists after all on whether the ability to modulate attention divides mammals from birds (Macphail, 1982). Provided that centrally-generated associability changes can indeed be demonstrated, the method introduced here offers an unprecedented opportunity to explore the role of attention in discrimination learning and cue competition. A few examples will illustrate this point.

As mentioned in the Chapter 1, the method of blocking of unblocking (Mackintosh & Turner, 1971) has uniquely afforded us some evidence for a decrement in the conditionability to a blocked CS. A few blocking trials impaired an added CS from becoming an inhibitor when the US was subsequently omitted (Mackintosh & Turner, 1971), and from gaining excitation when the magnitude of the US was subsequently increased (Mackintosh & Turner, 1971) or its identity was altogether changed (Kruschke & Blair, 2000). An alternative way of measuring changes in associability in blocking is to assess the readiness with which the blocked CS can serve in the solution of a transfer discrimination. Tables 6.1 and 6.2 show two experimental designs intended to do just that. The first of these designs provides a comparison of the associability of the blocked cue with that of the blocking cue. After receiving two blocking treatments in Stages 1 and 2, subjects are transferred to a test discrimination equivalent in logic to that previously used for the study of associability changes in true discriminations (Chapter 4). If blocking leads to a difference in the

Table 6.1.

Stage 1	Stage 2	Test
A+ C+	AB+ CD+	AD + AB - CD-

Table 6.2.

Stage 1	Stage 2	Test
A+ C+	AB+ CD+ EF+ GH+	BF+ FD- BH-

associability of the blocking and the blocked cue, then the component test discrimination based on the formerly blocking cues (AD+ CD-) should proceed more readily than that based on the formerly blocked cue (AD+ AB-). The second design affords a more direct test of the account of blocking provided by the theory of Mackintosh (1975). This is achieved by comparing the associability of a blocked cue with that of an overshadowed cue. Blocking is defined, after all, by the failure of a cue to gain control over behaviour in a blocking treatment *by comparison with an overshadowing control treatment* ⁽²⁾. Obvious as this point may seem, it remained unaddressed (for methodological reasons) by the aforementioned studies of blocking of unblocking (Mackintosh & Turner, 1971; Kruschke & Blair, 2000). Therefore, if associability changes are to provide an explanation for the blocking effect, then the component test discrimination based on the previously blocked cues (BF+ FD-) should be harder than that based on the previously overshadowed cues (BH-).

A further use to which the present technique can be put is the investigation of the effects of partial reinforcement on the associability of a cue. Pearce, Kaye and Hall (1982) showed evidence that partial reinforcement maintains the associability of a cue at a relatively high level. In the first stage of their experiment, one group of rats received presentations of a tone consistently followed by the delivery of a mild shock, whereas for another group the tone was partially reinforced. For a control group, the shock was consistently signalled by the presentation of a light. Following this training, all rats were transferred to a second stage in which the same tone was consistently followed by the delivery of a stronger shock. Although the group pretrained with a partial-reinforcement schedule was somewhat more impaired than the control group pretrained with the light, this latent inhibition effect was by no means as large as that observed between the group that had received the tone

continuously reinforced and the control group. These results are consistent with the notion of associability put forward in the Pearce-Hall (1980) model. According to this account, the associability of a stimulus should remain high only as long as the animal is learning about the consequences of that event. Once the stimulus has established itself as an accurate predictor of its consequences, no further learning is needed and its associability will therefore decrease.

In view of the contradictory rules governing associability changes proposed by the Pearce-Hall (1980) model and Mackintosh's (1975) theory, a number of authors have suggested the possibility that the two theories might be using the term "associability" to refer to different processes (Pearce, George & Redhead, 1998; Le Pelley, 2004). Le Pelley (2004), for instance, has proposed that a Mackintosh-type (1975) of associability mechanism would be most useful at the outset of the conditioning experience, by enabling the animal to identify the best predictors of the presence and absence of reinforcement, as well as the poor predictors. Once good predictors are teased apart from poor ones, a Pearce-Hall-type (1980) of associability mechanism would then determine exactly how much processing power is needed for learning about each cue to take place. Unfortunately, this cannot be more than a conjecture at this point since, beyond the aforementioned evidence on the effects of predictive accuracy on simple conditioning, no equivalent evidence exists on the effects of predictive accuracy on discrimination learning. One way to test this assumption is to use the technique under discussion to compare the readiness with which a partially- reinforced cue and a continuously-reinforced cue are able to serve in the solution of a transfer discrimination. This may be achieved through the design shown in Table 6.3, in which two pairs of cues are trained on a continuously- and a partially-reinforced schedule, respectively. At test, the four cues are recombined into

two component test discriminations in the usual manner. If partial reinforcement maintains the associability of stimuli at a higher level than continuous reinforcement, then the component test discrimination based on the formerly partially-reinforced cues should be easier than that based on the formerly continuously-reinforced cues.

Table 6.3.

Training	Test
A+ B+ C+/- D+/-	AC+ AD- CB-

A final example of the way in which this technique can further our understanding of the processes driving associability changes comes from the counterintuitive results reported by Williams, Mehta, Poworoznyk, Orihel, George and Pearce (2002). In their basic design, rats received training with a negative-patterning discrimination with an additional, background cue present on all trials: AX+ BX+ ABX-. At test, the background cue X was found to evoke a supernormal CR by comparison with another cue Y trained independently. These results actually follow the predictions of the Rescorla-Wagner (1972) model and the configural theory of Pearce (1987, 1994).

According to the Rescorla-Wagner (1972) model, more frequent pairings of X with reinforcement relative to A and B should leave this stimulus as the fittest competitor for associative strength of the three. On nonreinforced trials, moreover, A and B—but not X—will gain inhibitory strength, allowing the discrimination to be solved. The inhibition of A and B will in turn propel X towards further increments of

associative strength, thus becoming a superexcitor. With continued training, the associative strength of X will tend to an asymptote at 2λ , whereas the associative strengths of A and B will tend to $-\lambda$. Notice that to generate this prediction one needs not assume the existence of “unique cues” on compound trials; common elements — i.e. X— afford the model yet another way in which configural discriminations can be solved.

The results lie outside the scope of Mackintosh’s (1975) theory. Its strict elemental account of conditioning precludes the model from handling any discrimination that relies on a configural solution. But quite apart from this failure to predict the acquisition of even the simplest negative-patterning discrimination, the results from Williams et al. (2002) take on a special significance for any attentional theory of discrimination learning. This is because they provide a rather unique example of a stimulus (X) that, being totally irrelevant to the solution of the discrimination it forms part of, may still acquire substantial —indeed, supernormal— amounts of associative strength. Two possibilities present themselves; each of the utmost theoretical importance. First, that X possesses high associability, as its prominent position as a predictor of reinforcement would suggest. This would imply that, contrary to established knowledge (e.g. George & Pearce, 1999), a factor other than relevance drives associability gains in discrimination learning. Second, that X possesses low associability, saving the account according to which irrelevant stimuli are ignored. The importance of the latter possibility is that it would suggest a dissociation between the associative status of a cue and its associability, posing no less of a challenge for theories of attention. The behavioural technique under discussion provides a way to decide between these alternatives by measuring the

associability of the supernormal cue X in a direct fashion. Table 6.4 shows the details of the experimental design.

Table 6.4.

Training	Test
AX+ BX+ ABX- CY+ DY+ CDY-	AX+ CX- AY-

In the first stage of training, subjects receive training with two negative-patterning discriminations, A+ B+ AB- and C+ D+ CD-, to which two background cues, respectively X and Y, have been added. If the results from Williams et al. (2002) are replicated, then this schedule should result in X and Y becoming superexcitatory stimuli, with A, B, C and D becoming inhibitors. The two discriminations in this stage should leave A, B, C and D as the relevant stimuli and X and Y as irrelevant. Following this training, two component test discriminations, AX+ CX- and AX+ AY-, are administered. If associability changes depend on the relevance of a cue to the solution of the discrimination in which is embedded, then the AX+ CX- discrimination should be solved more readily than the AX+ AY- discrimination. If, however, the associability of a cue is directly related to its associative status, but independent of its relevance, then the opposite should be found.

As for the results of Experiment 8, which indicate that associability changes are not a necessary condition for the relative validity effect to occur, an obvious question is prompted. Quite apart from the fact that some within-experiment evidence for the absence of associability changes in the target cue appears desirable, one may

ask what the nature of the mechanism driving this effect is. Although the results from Experiment 8 are consistent with the account advanced by the Rescorla-Wagner (1972) model, other explanations for the relative validity effect than those contemplated by the theories here discussed are available. Some of these explanations are suggested by the fact that in a typical relative validity experiment the effects of partial-reinforcement training are compared with those of discrimination training. For instance, a higher level of uncertainty about trial outcomes during pseudo- than true-discrimination training might bring into play extraneous variables such as frustration. Available evidences antedating the discovery of relative validity demonstrate that one of the effects of frustration may be that of “energising” the CR (e.g. Amsel & Roussel, 1952; Kimble, 1961, p. 309). A related point is of course the partial-reinforcement extinction effect (e.g. Roberts, Bullock & Bitterman, 1963; Rescorla, 1999), which in all probability may have contributed to waxing the effect in a number of demonstrations administering repeated tests with X, including Wagner et al’s (1968) original experiments. In a similar vein, pseudodiscriminations have been suggested to teach animals to respond indiscriminately given the insoluble nature of the problem (e.g. Overmier & Wielkiewicz, 1983). If it is assumed that flatter generalisation gradients around AX and BX will develop during pseudo- than true discrimination training, then a faster rate of responding in the presence of X following the former training is somewhat expected. To the extent that this is correct, it is unclear whether testing element X after pseudodiscrimination training can provide a sensitive measure of its associative strength.

Although it could be argued that these drawbacks should mainly apply to between-subjects, relative-validity experiments, it may well be that within-subjects designs do not escape some general related consequence of pseudodiscrimination

training, and evidence to this effect has been reported in pigeons (Rescorla, 1999). Issues such as these should be addressed in the future, lest the relative validity effect has been unduly construed as reflecting cue competition.

6.5. General Conclusions

The overall concern of this thesis has been to throw some light on the mechanisms responsible for cue competition. To this end, a number of predictions made by two influential theoretical accounts, the Comparator Hypothesis (Miller & Matzel, 1988; Denniston et al., 2001) and the attentional theory of Mackintosh (1975), have been assessed. Overall, the results presented in this thesis have found little support for the account of cue competition advanced by the Comparator Hypothesis. Clearly, more work is needed in order to determine which aspects of the theory hold a grain of truth and deserve to be salvaged, and which are plainly wrong. No less problematic for the attentional theory of Mackintosh (1975) were the results here reported. Not only did they fail to find evidence of “true” associability changes in pigeons but, crucially, produced evidence of relative validity in the absence of such changes. Although the nature of the mechanism underpinning relative validity and cue competition in general remains obscure, the results here reported are as a whole in accordance with Standard Associative Theory [e.g. Rescorla-Wagner (1972) model].

Notes

(1) It is worth interjecting here that the Extended Comparator Hypothesis (Denniston et al., 2001; Savastano et al., 2003) is able to account for the related phenomenon of supernormal conditioning, in which the presence of a formerly trained inhibitor enhances conditioning to a novel CS (Urushihara et al., 2005). This facilitating influence of an inhibitor upon an excitator other than that which fostered its inhibition can be derived from the operation of second-order comparator processes. In the first instance, a cue is driven, through a comparator process with a much more powerful excitator, to exhibit the behaviour typical of inhibition. When the cue is subsequently paired with the novel CS, it will contribute to the new comparator process with a negatively activated US-representation. As a result, the expression of the association between the novel CS and the US will be enhanced. No such second-order comparator mechanism is possible, of course, when the target CS being compared with the inhibitor is the very excitator driving its inhibition, as is the case in Experiment 1.

(2) Here lies, in fact, the merit of Kamin's (1968, 1969) seminal experiments. Previous demonstrations that a stimulus fails to condition when trained in the presence of another cue formerly trained go back to as early as 1906 (Palladin, in Razran, 1965). Unfortunately, none of the experiments preceding Kamin's included an overshadowing control group, in the absence of which it is unwarranted to attribute the response deficit of X to pretraining with A.

References

References

- Amsel, A., & Roussel, J. (1952). Motivational properties of frustration: I. Effect on a running response of the addition of frustration to the motivational complex. *Journal of Experimental Psychology*, 43(5), 363–366.
- Arcediano, F., Escobar, M., & Matute, H. (2001). Reversal from blocking in humans as a result of posttraining extinction of the blocking stimulus. *Animal Learning & Behavior*, 29, 354–366.
- Arcediano, F., Escobar, M., & Miller, R. R. (2004). Is stimulus competition an acquisition deficit or a performance deficit? *Psychonomic Bulletin & Review*, 11, 1105–1110.
- Arcediano, F., Matute, H., & Miller, R.R. (1997). Blocking of Pavlovian conditioning in humans. *Learning and Motivation*, 28, 188–199.
- Azorlosa, J. L., & Cicala, G. A. (1988). Increased conditioning in rats to a blocked CS after the first compound trial. *Bulletin of the Psychonomic Society*, 26, 254–257.
- Baker, A. G., Murphy, R. A., & Vallée-Tourangeau, F. (1996). Associative and normative models of causal induction: Reacting to versus understanding cause. In D.R.Shanks, K.J.Holyoak & D.L.Medin (Eds.), *The Psychology of Learning and Motivation*, 34, 1–45. San Diego, CA: Academic Press.
- Balaz, M.A., Gutsin, P., Cacheiro, H., & Miller, R.R. (1982). Blocking as a retrieval failure: Reactivation of associations to a blocked stimulus. *Quarterly Journal of Experimental Psychology*, 34B, 99–113.
- Batsell, W. R. (1997). Retention of context blocking in taste-aversion learning. *Physiology and Behavior*, 61, 437–446.
- Bitterman, M. E., & Coate, W. B. (1950). Some new experiments on the nature of discrimination learning in the rat. *Journal of Comparative and Physiological Psychology*, 43, 198–210.
- Blaisdell, A. P., Gunther, L. G., & Miller, R. R. (1999). Recovery from blocking achieved by extinguishing the blocking CS. *Animal Learning & Behavior*, 27, 63–76.

Blaisdell, A. P., & Miller, R. R. (2001). Conditioned inhibition produced by extinction-mediated recovery from the relative stimulus validity effect: A test of acquisition and performance models of empirical retrospective revaluation. *Journal of Experimental Psychology: Animal Behavior Processes*, 27, 48–58.

Blum, R., & Blum, J. (1949). Factual issues in the "continuity" controversy. *Psychological Review*, 56, 33–50.

Bower, G., Trabasso, T. (1963). Reversals prior to solution in concept identification. *Journal of Experimental Psychology*. 66, 409–18

Bradley, J. I. (1961). Pre-solution trials in discrimination learning. *Psychological Reports*, 8, 155–158.

Bush, R. R., & Mosteller, F. (1951). A model for stimulus generalization and discrimination. *Psychological Review*, 58, 413–423.

Cole, R. P., Barnet, R. C., & Miller, R. R. (1995). Effect of relative stimulus validity: Learning or performance deficit? *Journal of Experimental Psychology: Animal Behavior Processes*, 21, 293–303.

Cole, R. P., Denniston, J. C., & Miller, R. R. (1996). Reminder-induced attenuation of the effect of relative stimulus validity. *Animal Learning & Behavior*, 24, 256–265.

Cole, R. P., Gunther, L. M., & Miller, R. R. (1997). Spontaneous recovery from the effect of relative stimulus validity. *Learning and Motivation*, 28, 1–19.

Coleman, S. R. & Gormezano, I. (1971). Classical conditioning of the rabbit's (*Oryctolagus cuniculus*) nictitating membrane response under symmetrical CS–US interval shifts, *Journal of Comparative and Physiological Psychology*, 77, 447–455.

Couvillon, P. A., Tennant, W. A., & Bitterman, M. E. (1976). Intradimensional vs. extradimensional transfer in the discriminative learning of goldfish and pigeons. *Animal Learning & Behaviour*, 4, 197–203.

Denniston, J. C., Savastano, H. I., & Miller, R. R. (2001). The extended comparator hypothesis: Learning by contiguity, responding by relative strength. In R. R. Mowrer & S. B. Klein (Eds.), *Handbook of contemporary learning theories*, 65–117. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Dickinson, A. (1980). *Contemporary Animal Learning Theories*. Cambridge: Cambridge University Press.

Dickinson, A., Burke, J. (1996). Within-compound associations mediate the retrospective revaluation of causality judgements. *Quarterly Journal of Experimental Psychology*, 49B, 60–80

Dickinson, A., & Charnock, D. J. (1985). Contingency effects with maintained instrumental reinforcement. *Quarterly Journal of Experimental Psychology*, 37B, 397-416.

Dollard, J. and Miller, N. E. (1950). *Personality and psychotherapy*, McGraw-Hill, New York.

Durlach, P. J. & Rescorla, R. A. (1980). Potentiation rather than overshadowing in flavor-aversion learning: an analysis in terms of within-compound associations, *Journal of Experimental Psychology: Animal Behavior Processes*, 6, 175–187.

Estes, W. K. (1950). Toward a statistical theory of learning. *Psychological Review*, 57, 94-107.

Estes, W. K. (1955). Statistical theory of distributional phenomena in learning. *Psychological Review*, 62, 369-377.

Estes, W. K. (1959). Component and pattern models with Markovian interpretation. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory* (pp. 239-263). Stanford, CA: Stanford University Press.

Estes, W. K. (1969). Outline of a theory of punishment. In *Punishment and aversive behavior* (ed. B.A. Campbell and R.M. Church), 57-82. Appleton-Century-Crofts, New York, NY.

Galef, B. G., Jr & Osborne, B. (1978). Novel taste facilitation of the association of visual cues with toxicosis in rats. *Journal of Comparative and Physiological Psychology*, 92, 907-16.

Gatling, F. P. (1951). A study of the continuity of the learning process as measured by habit reversal in the rat. *Journal of Comparative & Physiological Psychology*, 44, 78-83.

George, D. N., & Pearce, J. M. (1999). Acquired distinctiveness is controlled by stimulus relevance not correlation with reward. *Journal of Experimental Psychology: Animal Behavior Processes*, 25, 363–373.

Guthrie, E. R. (1935). *The Psychology of Learning*. New York: Harper.

Hall, G., & Channell, S. (1985). A comparison of intradimensional and extradimensional shift learning in pigeons. *Behavioral Processes*, 10, 285–295.

Hall, G., Mackintosh, N. J., Goodall, G., & Dal Martello, M. (1977). Loss of control by a less valid or by a less salient stimulus compounded with a better predictor of reinforcement. *Learning and Motivation*, 8, 145–158.

Hall, G. (1991). *Perceptual and associative learning*. Oxford: Oxford University Press.

Hall, G. (1996). Learning about associatively activated stimulus representations: Implications for acquired equivalence and perceptual learning, *Animal Learning & Behavior*, 24(3), 233–255.

Hara, K., & Warren, J. M. (1961). Stimulus additivity and dominance in discrimination performance by cats. *Journal of Comparative and Physiological Psychology*, 54, 86–90.

Hebb, D. O. (1949). *The organization of behavior*, John Wiley, New York.

Holland, P. C. (1999). Overshadowing and blocking as acquisition deficits: No recovery after extinction of overshadowing or blocking cues. *Quarterly Journal of Experimental Psychology*, 52B, 307–333.

Holland, P. C., & Forbes, D. T. (1982). Representation-mediated extinction of conditioned flavor aversions. *Learning and Motivation*, 13, 454–471.

Honey, R. C., & Hall, G. (1989). Acquired equivalence and distinctiveness of cues. *Journal of Experimental Psychology: Animal Behavior Processes*, 16, 178–184.

Hughes, C. L., & North A. J. (1959). Effect of introducing a partial correlation between a critical cue and a previously irrelevant cue. *Journal of Comparative & Physiological Psychology*, 52, 126–128.

Hull, C. L. (1943). *Principles of behavior*. New York: D. Appleton-Century.

Hull, C. L. (1950). Simple qualitative discrimination learning. *Psychological Review*, 57, 303-313.

Hull, C. L. (1952). *A behavior system*. New Haven: Yale University Press.

Humphreys, L. G. (1939). The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions. *Journal of Experimental Psychology*, 25, 141-158.

Jenkins, H. M., & Sainsbury, R. S. (1969). The development of stimulus control through differential reinforcement. In N. J. Mackintosh & W. K. Honig (Eds.), *Fundamental issues in associative learning*. Halifax, Canada: Dalhousie University Press.

Jenkins, H. M. (1973). Noticing and responding in a discrimination based on a distinguishing element. *Learning and Motivation*, 4, 115-137.

Kamin, L. J. (1965). Temporal and intensity characteristics of the conditioned stimulus. In W. F. Prokasy (Ed.), *Classical conditioning: A symposium*. New York: Appleton Century-Crofts.

Kamin, L. J. (1968). "Attention-like" processes in classical conditioning. In M. R. Jones (Ed.), *Miami Symposium on the Prediction of Behavior, 1967: Aversive stimulation*, 9-31. Coral Gables, FL: University of Miami Press.

Kamin, L. J. (1969). Predictability, surprise, attention and conditioning. In B. A. Campbell & R.M. Church (Eds.), *Punishment and aversive behavior*. New York: Appleton-Century-Crofts.

Kaufman, M.A., and Bolles, R.C. (1981). A nonassociative aspect of overshadowing. *Bulletin of the Psychonomic Society*, 18, 318-320.

Kimble, G. A. (1961). *Hilgard and Marquis' Conditioning and Learning*. Appleton-Century-Crofts, Inc., New York.

Krechevsky, I. (1932). Hypotheses in rats. *Psychological Review*, 39, 516-532.

Krechevsky, I. (1938). A study of the continuity of the problem solving process. *Psychological Review*, 45, 107-133.

Kruschke, J. K. (1992). ALCOVE: an exemplar based connectionist model of category learning. *Psychological Review*, 99, 22-44.

Kruschke, J. K., & Blair, N. J. (2000). Blocking and backward blocking involve learned inattention. *Psychonomic Bulletin & Review*, 7, 636-645.

LaBerge, D. (1995). *Attentional processing*. Harvard University Press, Cambridge.

Larkin, M. J., Aitken, M. R. F., & Dickinson, A. (1998). Retrospective revaluation of causal judgments under positive and negative contingencies. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 24, 1331-1352.

Lashley, K. S. (1929). *Brain mechanisms and intelligence: A quantitative study of injuries to the brain*. Chicago: University of Chicago Press. Reprinted by Dover, New York, 1963.

Lashley, K. S. (1938). The mechanism of vision: XV. Preliminary studies of the rat's capacity for detail vision. *Journal of General Psychology*, 18, 123-193.

Lashley, K. S. (1942). An examination of the continuity theory as applied to discriminative learning. *Journal of General Psychology*, 26, 241-265.

Lawrence, D. H. (1949). Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *Journal of Experimental Psychology*, 39, 770-784.

Lawrence, D. H. (1950). Acquired distinctiveness of cues: II. Selective association in a constant stimulus situation. *Journal of Experimental Psychology*, 40, 175-188.

Lawrence, D. H. (1952) The transfer of a discrimination along a continuum. *Journal of Comparative and Physiological Psychology*, 45, 511-516.

Lawrence, D. H. (1963). The nature of a stimulus: some relationships between learning and perception. In *Psychology: a study of a science*, 5, (ed. S. Koch), 179-212. McGraw-Hill, New York.

Le Pelley, M. E. (2004). The role of associative history in models of associative learning: A selective review and a hybrid model. *Quarterly Journal of Experimental Psychology*, 57B, 193-243.

Logan, F. A. (1954). A note on stimulus intensity dynamism (V). *Psychological Review*, 61, 77-80.

Lovejoy, E. (1965). An attention theory of discrimination learning. *Journal of Mathematical Psychology*, 2, 342-362.

Lovejoy, E. (1968). *Attention in discrimination learning*. San Francisco: Holden-Day.

Lubow, R. E. (1973). Latent inhibition. *Psychological Bulletin*, 79, 398-407.

Lubow, R. E., & Moore, A. U. (1959). Latent inhibition: The effect of nonreinforced preexposure to the conditioned stimulus. *Journal of Comparative and Physiological Psychology*, 52, 415-419.

Luongo, A. (1976). Stimulus selection in discriminative taste-aversion learning in the rat. *Animal Learning & Behavior*, 4, 225-230.

Mackintosh, N. J. (1965a). The effect of attention on the slope of generalization gradients. *British Journal of Psychology*, 56, 87-93.

Mackintosh, N. J. (1965b). Selective attention in animal discrimination learning. *Psychological Bulletin*, 64, 124-150.

Mackintosh, N. J. (1973). Stimulus selection: Learning to ignore stimuli that predict no change in reinforcement. In R. A. Hinde & J. S. Hinde (Eds.), *Constraints on learning*, 75-96. London: Academic Press.

Mackintosh, N. J. (1974). *The psychology of animal learning*. San Diego, CA: Academic Press.

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

Mackintosh, N. J. (1976). Overshadowing and stimulus intensity. *Animal Learning & Behavior*, 4, 186–192.

Mackintosh, N. J. (1983). *Conditioning and Associative Learning* (Oxford University Press, New York).

Mackintosh, N. J., & Holgate, V. (1967). Effects of several pretraining procedures on brightness probability learning. *Perceptual and Motor Skills*, 25, 629–637.

Mackintosh, N. J., & Little, L. (1969). Intradimensional and extradimensional shift learning by pigeons. *Psychonomic Science*, 14, 5–6.

Mackintosh, N. J., & Turner, C. (1971). Blocking as a function of novelty of CS and predictability of UCS. *Quarterly Journal of Experimental Psychology*, 23, 359–366.

Mackintosh, N. J., & Honig, W. K. (1970). Blocking and attentional enhancement in pigeons. *Journal of Comparative & Physiological Psychology*, 73, 78–85.

Mahut, H. (1954). The effect of stimulus position on visual discrimination by the rat. *Canadian Journal of Psychology*, 8(3), 130–138.

Markman, A. B. (1989). LMS rules and the inverse base-rate effect: Comment on Gluck and Bower (1988). *Journal of Experimental Psychology: General*, 118, 417–421.

Matzel, L., Schachtman, T. & Miller R. (1985). Recovery of an overshadowed association achieved by extinction of the overshadowing stimulus. *Learning and Motivation*, 16, 398–412.

McCulloch, T. L., & Pratt, J. G. (1934). A study of the presolution period in weight discrimination by white rats. *Journal of Comparative Psychology*, 18, 271–290.

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

Miller, N. E. (1948). Theory and experiment relating psychoanalytic displacements to stimulus-response generalization. *Journal of Abnormal and Social Psychology*, 43, 155–178.

Miller, N. E. & Dollard, J. (1941). *Social learning and imitation*. Yale University Press.

Miller, R. R., & Schachtman, T. R. (1985). Conditioning context as an associative baseline: Implications for response generation and the nature of conditioned inhibition. In R. R. Miller & N. E. Spear (Eds.), *Information processing in animals: Conditioned inhibition*, 51–88. Hillsdale, NJ: Erlbaum.

Miller, R. R., & Matzel, L. D. (1988). The comparator hypothesis: A response rule for the expression of associations. In G. H. Bower (Ed.), *The psychology of learning and motivation*, 22, 51–92. San Diego, CA: Academic Press.

Miller, J. S., Jagielo, J. A., & Spear, N. (1993). The influence of retention interval on the US pre-exposure effect: Changes in blocking over time. *Learning and Motivation*, 24, 376–394.

Miller, R. R., Barnet, R. C., & Grahame, N. J. (1995). Assessment of the Rescorla-Wagner model. *Psychological Bulletin*, 117, 363–386.

Mumma, R., & Warren, J. M. (1968). Two-cue discrimination learning by cats. *Journal of Comparative & Physiological Psychology*, 66, 116-122.

Murphy, R. A., Baker, A.G. & Fouquet, N. (2001a). Relative validity effects with either one or two more valid cues in Pavlovian and instrumental conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, 27, 59-67.

Murphy, R. A., Baker, A.G., & Fouquet, N. (2001b). Relative validity of contextual and discrete cues. *Journal of Experimental Psychology: Animal Behavior Processes*, 27, 137-152.

Neimark, E. D., & Estes, W. K. (1967). *Stimulus sampling theory*. San Francisco: Holden-Day.

Overmier, J., & Wielkiewicz, R. (1983). On unpredictability as a causal factor in “Learned Helplessness”. *Learning and Motivation*, 14, 324-337.

Pavlov, I. P. (1927). *Conditioned reflexes* (trans. by G. V. Anrep). Oxford: Oxford Univ. Press.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, 87, 532–552.

Pearce, J. M., Kaye, H., & Hall, G. (1982). Predictive accuracy and stimulus associability: Development of a model for Pavlovian conditioning. In M. Commons, R. Herrnstein, & A. R. Wagner (Eds.), *Quantitative analysis of behaviour*, 241–255. Cambridge, MA: Ballinger.

Pearce, J. M. (1987). A model for stimulus generalization in Pavlovian conditioning. *Psychological Review*, 94, 61–73.

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

Pearce, J. M., George, D. N., & Redhead, E. S. (1998). The role of attention in the solution of conditional discriminations. In N. A. Schmajuk & P. C. Holland (Eds.), *Occasion setting: Associative learning and cognition in animals*. Washington, DC: American Psychological Association.

Pearce, J. M. & Redhead, E. S. (1999). Pavlovian supernormal conditioning with a change in appetitive reinforcer. *Animal Learning and Behavior*, 27, 369–378.

Pearce, J. M. (2002). Evaluation and extension of a connectionist theory of configural learning. *Animal Learning & Behavior*, 30, 73–95.

Pearce, J. M., George, D. N., & Aydin, A. (2002). Summation: Further assessment of a configural theory. *Quarterly Journal of Experimental Psychology*, 55B, 61–73.

Perkins, C. C., Jr. (1953). The relation between conditioned stimulus intensity and response strength. *Journal of Experimental Psychology*, 46, 225–231.

Pineño, O., Urushihara, K., & Miller, R. R. (2005). Spontaneous recovery from forward and backward blocking. *Journal of Experimental Psychology: Animal Behavior Processes*, 31, 172–183.

Pubols, B. H. (1956). The facilitation of visual and spatial discrimination by overlearning. *Journal of Comparative Physiology and Psychology*, 49, 243–248.

Pullen, M. R., & Turney, T. H. (1977). Response modes in simultaneous and successive visual discriminations. *Animal Learning & Behaviour*, 5, 73–77.

Razran, G. (1965). Russian physiologists' psychology and American experimental psychology. *Psychological Bulletin*, 63, 42-64.

Reid, L. S. (1953). Development of non-continuity behavior through continuity learning. *Journal of Experimental Psychology*, 46, 107-112.

Rescorla, R. A. (1968). Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative Physiology and Psychology*, 67, 504-509.

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

Rescorla, R. A. (1976). Second-order conditioning of Pavlovian conditioned inhibition, *Learning and Motivation*, 7, 161-172.

Rescorla, R. A., & Cunningham, C. L. (1978). Within-compound flavor associations. *Journal of Experimental Psychology: Animal Behavior Processes*, 4, 267-275.

Rescorla, R. A., & Freberg, L. (1978). The extinction of within-compound flavor associations. *Learning and Motivation*, 9, 411-427.

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

Rescorla, R. A., & Durlach, P. J. (1981). Within-event learning in Pavlovian conditioning. In N. E. Spear & R. R. Miller (Eds.), *Information processing in animals: Memory mechanisms*, 81-111. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Rescorla, R. A., Grau, J. W., & Durlach, P. J. (1985). Analysis of the unique cue in configural discriminations. *Journal of Experimental Psychology: Animal Behaviour Processes*, 11, 356-366.

Rescorla, R. A. (1997). Response-inhibition in extinction. *Quarterly Journal of Experimental Psychology*, 50B, 238-252.

Rescorla, R. A. (1999). Within-subject partial reinforcement extinction effect in autoshaping. *The Quarterly Journal of Experimental Psychology*, 52B (1), 75-87.

Rescorla, R. A. (2001). Unequal associative changes when excitors and neutral stimuli are conditioned in compound. *Quarterly Journal of Experimental Psychology*, 54B, 53-68.

Rescorla, R. A. (2002). Effect of following an excitatory-inhibitory compound with an intermediate reinforcer. *Journal of Experimental Psychology: Animal Behavior Processes*, 28, 163-174.

Restle, F. (1955). A theory of discrimination learning. *Psychological Review*, 62, 11-19.

Revusky, S. (1971). The role of interference in association over delay. In W. K. Honig & P. H. R. James (Eds.), *Animal memory*, 155-213. San Diego, CA: Academic Press.

Roberts, W. A., Bullock, D. H., & Bitterman, M. E. (1963). Resistance to extinction in the pigeon after partially reinforced instrumental training under discrete-trials conditions. *American Journal of Psychology*, 76, 353-365.

Rodgers, J. P., & Thomas, D. R. (1982). Task specificity in nonspecific transfer and in extradimensional stimulus generalization in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, 8, 301-312.

Roitblat, H. L. (1987). *Introduction to comparative cognition*. Freeman, New York.

Ross, S. M., & Ross, L. E. (1971). Comparison of trace and delay classical eyelid conditioning as a function of interstimulus interval. *Journal of Experimental Psychology*, 91, 165-167.

Savastano, H. I., Arcediano, F., Stout, S. C., & Miller, R. R. (2003). Interaction between preexposure and overshadowing: Further analysis of the extended comparator hypothesis. *Quarterly Journal of Experimental Psychology*, 56(B), 371-395.

Schachtman, T. R., Brown, A. M., Gordon, E. L., Catterson, D. A., & Miller, R. R. (1987). Mechanisms underlying retarded emergence of conditioned responding

following inhibitory training: Evidence for the comparator hypothesis. *Journal of Experimental Psychology: Animal Behavior Processes*, 13, 310-322.

Schachtman, T. R., Gee, J. L., Kasprow, W. J., & Miller, R. R. (1983). Reminder-induced recovery from blocking as a function of the number of compound trials. *Learning and Motivation*, 14, 154-164.

Shanks, D. R. (1985). Forward and backward blocking in human contingency judgement. *Quarterly Journal of Experimental Psychology*, 37B, 1-21.

Shanks, D. R., & Dickinson, A. (1987). Associative accounts of causality judgment. *Psychology of Learning and Motivation*, 21, 229-261.

Shepp, B. E., & Eimas, P. D. (1964). Intradimensional and extradimensional shifts in the rat. *Journal of Comparative and Physiological Psychology*, 57, 357-361.

Siegel, S. (1967). Overtraining and transfer processes. *Journal of Comparative and Physiological Psychology*, 64, 471-477.

Speers, M. A., Gillan, D. J. and Rescorla, R. A. (1980). Within-compound associations in a variety of compound conditioning procedures, *Learning and Motivation*, 11, 135-149.

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

Spence, K. W. (1937a). Analysis of the formation of visual discrimination habits in the chimpanzee. *Journal of Comparative Psychology*, 23, 77-100.

Spence, K. W. (1937b). The differential response in animals to stimuli varying within a single dimension. *Psychological Review*, 44, 430-444.

Spence, K. W. (1940). Continuous versus non-continuous interpretations of discrimination learning. *Psychological Review*, 47, 271-288.

Spence, K. W. (1945). An experimental test of the continuity and non-continuity theories of learning. *Journal of Experimental Psychology*, 35, 253-66.

Spence, K. W. (1951). Theoretical interpretations of learning. In *Comparative Psychology*, 239-91. Edited by C. P. Stone. New York: Prentice-Hall.

Spence, K. W. (1952). The nature of the response in discrimination learning. *Psychological Review*, 59(1), 89-93.

Spence, K. W. (1960). *Behavior Theory and Learning*. Englewood Cliffs, New Jersey: Prentice-Hall.

Stout, S. C., Arcediano, F., Escobar, M., & Miller, R. R. (2003). Overshadowing as a function of trial number: Dynamics of first- and second-order comparator effects. *Learning & Behavior*, 31, 85-97.

Stout, S. C., & Miller, R. R. (2007). Sometimes-competing retrieval (SOCR): a formalization of the Comparator Hypothesis. *Psychological Review*, 114(3), 759-783.

Suret, M. B. S., & McLaren, I. P. L. (2003). Representation and discrimination on an artificial dimension. *Quarterly Journal of Experimental Psychology*, 56B, 30-42.

Sutherland, N. S. (1959). Stimulus analyzing mechanisms. In *Proceedings of a Symposium on the Mechanization of Thought Processes*, 2. London: H.M.S.O.

Sutherland, N. S., & Andelman, L. (1967). Learning with one and two cues. *Psychonomic Science*, 15, 253-254.

Sutherland, N. S., & Mackintosh, N. J. (1971). *Mechanisms of animal discrimination learning*. New York: Academic Press.

Tassoni, C. J. (1995). The least mean squares network with information coding: A model of cue learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 21, 193-204.

Tennant, W. & Bitterman, M. E. (1973). Some comparisons of intra- and extradimensional transfer in discriminative learning of goldfish. *Journal of Comparative and Physiological Psychology*, 83, 134-139.

Tennant, W. A. & Bitterman, M. E. (1975). Blocking and overshadowing in two species of fish. *Journal of Experimental Psychology: Animal Behavior Processes*, 1, 22-29.

Thomas, D. R. (1970). Stimulus selection, attention, and related matters. In J. H. Reynierse (Ed.), *Current issues in animal learning*. Lincoln: University of Nebraska Press.

Thompson, R. F., & Spencer, W. A. (1966). Habituation: A model phenomenon for the study of neuronal substrates of behaviour. *Psychological Review*, 173, 16-43.

Tolman, E. C. & Brunswik, E. (1935). The organism and the causal texture of the environment. *Psychological Review*. 42(1), 43-77.

Turrisi, F. D., Shepp, B., & Eimas, P. D. (1969). Intra- and extra-dimensional shifts with constant and variable-irrelevant dimensions in the rat. *Psychonomic Science*, 14, 19-20.

Udell, H., & Rescorla, R. A. (1979). Conditioning of simultaneous and successive common elements in a discrimination and pseudo discrimination. *Bulletin of the Psychonomic Society*, 14, 453-456.

Urcelay, G. P., & Miller, R. R. (2006). Counteraction between overshadowing and degraded contingency treatments: Support for the extended comparator hypothesis. *Journal of Experimental Psychology: Animal Behavior Processes*, 32, 21-32.

Urushihara, K., Wheeler, D. S., Pineño, O., & Miller, R. R. (2005). An extended comparator hypothesis account of superconditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, 31, 184-198.

Van Hamme, L. J., & Wasserman, E. A. (1994). Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. *Learning and Motivation*, 25, 127-151.

Wagner, A. R. (1969a). Incidental stimuli and discrimination learning. In R. M. Gilbert and N. S. Sutherland (Eds.), *Animal Discrimination Learning*. London: Academic Press. Pp. 83-111.

Wagner, A. R. (1969b). Stimulus validity and stimulus selection in associative learning. In N. J. Mackintosh & W. K. Honig (Eds.), *Fundamental issues in associative learning*. Halifax, Canada: Dalhousie University Press.

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

Wagner, A. R., Logan, F. A., Haberlandt, K., & Price, T. (1968). Stimulus selection in animal discrimination learning. *Journal of Experimental Psychology*, 76, 171-180.

Wagner, A. R., & Rescorla, R. A. (1972). Inhibition in Pavlovian conditioning: Application of a theory. In R. A. Boakes & M. S. Halliday (Eds.), *Inhibition and learning*, 301-336. London: Academic Press.

Ward-Robinson, J., & Hall, G. (1996). Backward sensory preconditioning. *Journal of Experimental Psychology: Animal Behavioral Processes*, 22, 395-404.

Wasserman, E. A. (1974). Stimulus-reinforcer predictiveness and selective discrimination learning in pigeons. *Journal of Experimental Psychology*, 103, 284-297.

Wheeler, D. S., & Miller, R. R. (2005). Recovery from blocking between outcomes. *Journal of Experimental Psychology: Animal Behavior Processes*, 31, 467-476.

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

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

Wyckoff, L. B. (1952). The role of observing responses in discrimination learning: Part 1, *Psychological Review*, 59, 431-442.

Zeaman, D., & House, B. J. (1963). The role of attention in retardate discrimination learning. In N. R. Ellis (Ed.), *Handbook of mental deficiency: Psychological theory and research*, 159-223. New York: McGraw-Hill.

