http://waikato.researchgateway.ac.nz/

Research Commons at the University of Waikato

Copyright Statement:

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

The thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author’s right to be identified as the author of the thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author’s permission before publishing any material from the thesis.
PEAK SHIFT IN REMEMBERING

A thesis
submitted in partial fulfilment
of the requirements for the Degree
of
Master of Social Sciences
at the
University of Waikato
by
ANDROS HOAN

University of Waikato
2003
ABSTRACT

If remembering is discriminative behaviour along the dimension of time and if, as Sargisson and White (2001) argued, generalisation around a peak can occur in this behaviour, then the peak shift which has been shown in discrimination along so many other stimulus dimensions, might also occur in remembering. To examine this hypothesis, 6 hens were trained in a delayed matching-to-sample procedure at delays of 2 and 4 s. The probability of reinforcement for correct responses was initially 0.9 at both delays until performance stabilised. A generalisation probe was then carried out by inserting unreinforced trials at delays of 0, 1, 1.5, 2.5, 3, 3.5, 4.5, 5 and 6 s in a session amongst normal training delay trials. The generalisation functions had a slight peak around 2 s. After further training, a second generalisation probe showed a slightly declining function. The probability of reinforcement at the 2 s delay was then dropped to 0.1, so that in the terms of the classical generalisation/peak shift paradigm, 2-s delay trials became S− and 4-s delay trials became S+. A third generalisation probe then was conducted. This resulted in a flat function from 0 s to 3 s, and a large, clear peak in discriminative performance at 4.5 s over all hens. After more of the same differential reinforcement training, a fourth generalisation probe showed a broad curve peaking at 3 s, with minima at 1 s and 6 s and a global maximum at 0 s. Another training condition was then run, with the probability of reinforcement at the 2-s delay dropped to 0, to see if increasing the "aversiveness" of S− would again result in a peak shift. A fifth generalisation probe was then conducted. This showed a sharp decline in discriminability at shorter delays, a dip around 2 s, and a very small area shift beyond 4 s, but no clear peak shift. This was interpreted as being due to overlearning, with the consequences of remembering at S− no longer significantly affecting performance at S+. A final training condition was then run, with S− moved from 2 s to 3 s with zero probability of reinforcement, and for only a short period, to prevent overlearning. It was predicted that this would cause peak shift to re-occur. A sixth generalisation probe was then conducted. This found a further decline in discriminability at shorter delays, a shift in the dip from 2 s to 3 s, and a large, clear peak at 4.5 s. This demonstration of peak shift in a remembering process would not have been predicted by any traditional theory of memory, but strongly supports the conception of remembering as discriminative behaviour along the stimulus dimension of time.
ACKNOWLEDGEMENTS

I wish to give special thanks to Dr Mary Foster, without whose support at some crucial times, I would have abandoned the university system, and this thesis would never have been written. Thanks also to Dr Bill Temple for some memorable insights. Thanks to Dr James McEwan, whose graduate module in behavioural perception provided the rare intellectual stimulus which led eventually to the idea for this research. Finally I want to thank Jenny Chandler, who built the equipment and introduced me to MED programming, and my fellow graduate students at the Animal Behaviour and Welfare Research Centre, especially Karen Smith, who helped to make it happen.
## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>iii</td>
</tr>
<tr>
<td>Contents</td>
<td>iv</td>
</tr>
<tr>
<td>List of Figures</td>
<td>vi</td>
</tr>
<tr>
<td>List of Tables</td>
<td>viii</td>
</tr>
<tr>
<td>List of contents of Appendix D: CD-R disc</td>
<td>ix</td>
</tr>
<tr>
<td>General Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Experiment 1</td>
<td>12</td>
</tr>
<tr>
<td>Method</td>
<td>12</td>
</tr>
<tr>
<td>Results</td>
<td>16</td>
</tr>
<tr>
<td>Discussion</td>
<td>22</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>24</td>
</tr>
<tr>
<td>Method</td>
<td>24</td>
</tr>
<tr>
<td>Results</td>
<td>24</td>
</tr>
<tr>
<td>Discussion</td>
<td>29</td>
</tr>
<tr>
<td>Experiment 3</td>
<td>30</td>
</tr>
<tr>
<td>Method</td>
<td>30</td>
</tr>
<tr>
<td>Results</td>
<td>31</td>
</tr>
<tr>
<td>Discussion</td>
<td>37</td>
</tr>
<tr>
<td>Experiment 4</td>
<td>40</td>
</tr>
<tr>
<td>Method</td>
<td>40</td>
</tr>
<tr>
<td>Results</td>
<td>40</td>
</tr>
<tr>
<td>Discussion</td>
<td>45</td>
</tr>
<tr>
<td>General results</td>
<td>47</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>General Discussion</td>
<td>53</td>
</tr>
<tr>
<td>References</td>
<td>80</td>
</tr>
<tr>
<td>Appendix A: Summary of the conditions and contingencies in the study</td>
<td>98</td>
</tr>
<tr>
<td>Appendix B: The cumulative data from training sessions</td>
<td>99</td>
</tr>
<tr>
<td>Appendix C: The cumulative data from generalisation probe sessions</td>
<td>111</td>
</tr>
<tr>
<td>Appendix D: CD-R disc</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Inside back cover)</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Fig.</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Conceptual origins of the present study</td>
<td>3</td>
</tr>
<tr>
<td>2-1</td>
<td>Time series graph of Log $d$: TR2-GP3</td>
<td>17</td>
</tr>
<tr>
<td>2-2</td>
<td>Time series graph of Latency to choice: TR2-GP3</td>
<td>17</td>
</tr>
<tr>
<td>3-1</td>
<td>Generalisation probes 1, 2 and 3: Log $d$ versus delay</td>
<td>18</td>
</tr>
<tr>
<td>3-2</td>
<td>Generalisation probes 1, 2 and 3: Percentage correct versus delay</td>
<td>19</td>
</tr>
<tr>
<td>3-3</td>
<td>Generalisation probes 1, 2 and 3: Latency to choice versus delay</td>
<td>20</td>
</tr>
<tr>
<td>4-1</td>
<td>Generalisation probe 4: Log $d$ versus delay</td>
<td>25</td>
</tr>
<tr>
<td>4-2</td>
<td>Generalisation probe 4: Percentage correct versus delay</td>
<td>26</td>
</tr>
<tr>
<td>4-3</td>
<td>Generalisation probe 4: Latency to choice versus delay</td>
<td>27</td>
</tr>
<tr>
<td>5-1</td>
<td>Time series graph of Log $d$: PS3-GP5</td>
<td>32</td>
</tr>
<tr>
<td>5-2</td>
<td>Time series graph of Latency to choice: PS3-GP5</td>
<td>32</td>
</tr>
<tr>
<td>6-1</td>
<td>Generalisation probe 5: Log $d$ versus delay</td>
<td>33</td>
</tr>
<tr>
<td>6-2</td>
<td>Generalisation probe 5: Percentage correct versus delay</td>
<td>34</td>
</tr>
<tr>
<td>6-3</td>
<td>Generalisation probe 5: Latency to choice versus delay</td>
<td>35</td>
</tr>
<tr>
<td>7-1</td>
<td>Generalisation probe 6: Log $d$ versus delay</td>
<td>42</td>
</tr>
<tr>
<td>7-2</td>
<td>Generalisation probe 6: Percentage correct versus delay</td>
<td>43</td>
</tr>
<tr>
<td>7-3</td>
<td>Generalisation probe 6: Latency to choice versus delay</td>
<td>44</td>
</tr>
<tr>
<td>8-1</td>
<td>Time series graph of Log $d$: TR1-GP6 (whole study)</td>
<td>48</td>
</tr>
<tr>
<td>8-2</td>
<td>Time series graph of Latency to choice: TR1-GP6 (whole study)</td>
<td>48</td>
</tr>
<tr>
<td>9-1</td>
<td>Generalisation probes 2-6: Log $d$ versus delay</td>
<td>49</td>
</tr>
<tr>
<td>9-2</td>
<td>Generalisation probes 2-6: Percentage correct versus delay</td>
<td>50</td>
</tr>
<tr>
<td>9-3</td>
<td>Generalisation probes 2-6: Latency to choice versus delay</td>
<td>51</td>
</tr>
<tr>
<td>10-1</td>
<td>Generalisation probe 1: Log $d$ Sessions 1, 2 and 3, and mean.</td>
<td>59</td>
</tr>
<tr>
<td>10-2</td>
<td>Generalisation probe 2: Log $d$ Sessions 1, 2, and mean.</td>
<td>60</td>
</tr>
<tr>
<td>10-3</td>
<td>Generalisation probe 3: Log $d$ Sessions 1, 2, and mean.</td>
<td>61</td>
</tr>
</tbody>
</table>
Fig. 10-4. Generalisation probe 4: Log $d$ Sessions 1, 2, and mean. 62
Fig. 10-5. Generalisation probe 5: Log $d$ Sessions 1, 2, and mean. 63
Fig. 10-6. Generalisation probe 5: Log $d$ Sessions 1, 2, and mean. 64
Fig. 11. A possible conceptual future for DMTS remembering studies. 78
LIST OF TABLES

Table 1. List of selected studies in which peak shift was found 8
CONTENTS OF APPENDIX D: CD-R DISC

READMEFIRST.doc (Microsoft Word for Windows™ file)

Appendix A.doc (Microsoft Word for Windows™ file)……Summary of the conditions and contingencies in the study

Cumulative data (folder)……The cumulative data

Eventdata (folder)…………..The consecutive within-session event data

Excelcumdata (folder)……..The cumulative data analysed in Microsoft Excel™

MEDPCpgms (folder)………..The MEDSTATE NOTATION™ programs used to run the experiments

PEAKSHIFTR.ppt (Microsoft PowerPoint™ file)….PowerPoint™ Presentation of the key results of this study
Instead of "memory," we should say "remembering"; instead of "thought" we should say "thinking," instead of "sensation" we should say "seeing, hearing," etc. But like other learned branches, psychology is prone to transform its verbs into nouns. Then what happens? We forget that our nouns are merely substitutes for verbs, and go hunting for the things denoted by the nouns; but there are no such things, there are only the activities that we started with, seeing, remembering, and so on…(Woodworth, 1921, cited in Catania, 1979, p. 307)

"It all depends on where you look for your causes."

(W. Temple, personal communication, March 2000)

The systematic quantitative study of forgetting is generally considered to have begun with Ebbinghaus (1885/1964). Since then a vast literature, and a bewildering number of theories of memory have arisen. The question, "how do we remember?" is addressed by perhaps four broad classes of theories, the first two of which clearly dominate the recent memory literature:

(a) Physiological trace theories, going back at least to Sechenov (1863/1965), which assume as given that memory is something built up in the brain. Even when they manipulate external contingencies, workers in this tradition are not satisfied that they have found a cause until an internal correlate has been identified or hypothesised. See McGaugh (2000) for recent review.

(b) Structuralist "information processing" (IP) theories, which attempt to answer the question by modelling humans and/or other animals as computers (Atkinson & Shiffrin, 1968). In this paradigm, an observed outcome is considered to be explained by a model, if hypothetically, receiving similar inputs as the subjects did under similar circumstances, it would produce similar output. As Roediger III (1979) pointed out, people have long used analogies to the technology of their time, from the wax tablet to the computer, to explain remembering. Not surprisingly, IP theories have arisen mostly since the advent of computers in the 1950's, and have become extremely complicated.

(c) Ecological remembering approaches (Bartlett, 1932), which consider remembering in its social context. Writers in this tradition tend to regard physiological approaches, and some or all experimental approaches, as sterile, and not reflecting the richness of the experience of remembering, or its function(s) in real lives. See Neisser (1982) and Neisser and Winograd (1995) for recent expositions.
(d) Radical remembering theories, which reject the notion that memory is a "thing" stored, either as a copy or some other representation in a physiological or mental filing cabinet, and then retrieved. Instead, researchers focus on the act or behaviour of remembering (Branch, 1977; Ben Zeev, 1986; Palmer, 1991; Wilcox & Katz, 1981). The present thesis arises from this perspective, cross-fertilised with ideas from the stimulus discrimination and generalisation literature, which will therefore be briefly discussed.

The narrower question, "why do we forget?", has been predominantly addressed by two types of theories:

(a) Trace decay theory—essentially this is the idea that perception creates and stores a copy, or representation of the thing perceived. This copy/representation is assumed to start at its maximal value, and then to decline monotonically, according to a physiologically more or less pre-ordained curve, similar to that shown in Figure 1(a). Originally regarded as a physiological construct by Sechenov (1863/1965) and Pavlov (1927/1960), the memory trace was later adapted as a psychological construct, (e.g. Hebb, 1949/1961; Hull, 1952/1964; Wickelgren, 1974) but has never lost its pseudo-physiological connotations. See Gomulicki (1953), for a historical background. Trace decay is a primitive, inadequate, but remarkably resilient idea.

(b) Interference theories - a class of attempts to address the manifest inadequacies of simple trace decay theory by considering the effects of environmental contingencies on forgetting. McGeoch (1932) viewed these effects in terms of competing responses, but nowadays they are usually interpreted in terms of various hypothesised internal mediating constructs affecting the storage or retrieval of memory traces (Postman & Underwood, 1973). See Anderson and Neely (1996) for a recent overview.

Regardless of the proportions of decay or interference posited, advocates of both views nearly always agree that forgetting follows monotonically declining functions. Rubin and Wenzel (1996) analysed 210 sets of published data on forgetting functions, and concluded that there were four candidate functions – all monotonically declining, which fit the data about equally well. Even behaviourally oriented writers such Staddon, Chelaru and Higa (2002, p. 74) (incorrectly citing the year as 1999) have boldly endorsed Rubin and Wenzel's (1996) results as representing, "The empirical forgetting function". Thus it can fairly be said that the currently dominant psychological view of the course of forgetting functions over periods of up to about a minute, is that of a monotonically declining function such as that shown in Figure 1(a). Figures 1 (b), (c), and (d) illustrate the evolution of the idea for the present study, from the conventional view, via a radically different interpretation of the nature of memory, to a prediction of a phenomenon never before demonstrated in remembering.
Conventional "memory decay" curve - reinterpreted by White (1985) as discrimination along the dimension of time with respect to the stimulus at time zero.

Generalisation along the dimension of time around a peak in this discriminative behaviour - caused by reinforcing correct remembering at time delay T (Sargisson & White, 2001).

One possible hypothetical result of training at two delays, T1 and T2.

If we train at two delays, but reinforce poorly at one, (S¯), and richly at the other, (S+), will the peak in discriminative performance shift to a delay on the far side of S+ away from S¯?

**Figure 1.** Conceptual origins of the present study.
White (1985, 1991) proposed a theory of direct remembering, which treats remembering as an act of discrimination along the stimulus dimension of time. White's theory arises from the radical remembering perspective in general, but in particular, from a fusion of the direct perceptual ideas of Gibson (1977, 1979), and Skinner's (1938/1991, 1974) operant conditioning. To remember is to respond to a stimulus, some time after it was presented, i.e., at a distance in time away from its presentation. Forgetting over time, in the Direct Remembering (DR) view, is due to the stimulus being a greater distance away in time, just as a landmark becomes more distant as we walk away from it, till eventually it can no longer be seen.

Furthermore, DR theory argues that remembering and forgetting, like other behaviours, can be modified by their consequences. It thus provides a theoretically coherent experimental paradigm for the functional analysis of remembering behaviour. DR theory differs from traditional interference theories which, while rejecting a simple decay model, still treat memories as things which are stored and then retrieved. The "interference" in interference theories interferes with these processes. In DR studies, the accuracy of remembering is measured at various times, or delay points, after the presentation of some stimulus, and plotted graphically. However, lines drawn through these points, or any mathematical function fitted to them, constitute a relational behavioural description, not an intrinsic property of an internal event or construct. DR theory views remembering performance at any one of the infinite number of delay points along the temporal dimension as a function of contingencies, such as the consequences of remembering, applying at that particular delay, independently of those at other delays. This directed research toward investigating the effects of systematically modifying contingencies at various delays.

The experimental methodology generally used in this type of research (both with humans and animals), is that of delayed matching-to-sample (DMTS). DMTS has a long history (D'Amato, 1973), but the seminal modern experimental studies were done by Blough (1959), and Berryman, Cumming, and Nevin (1963). In a typical DMTS procedure, a sample stimulus (e.g. a red light) is presented to the subject for a defined length of time, and then withdrawn. After a further amount of time (the delay), comparison stimuli are presented (e.g. a red light and a green light), and the subject is required to choose one of these. If the comparison stimulus chosen matches the earlier sample, (in this case, the red light), then reinforcement is given, and the subject can learn to choose correctly upon subsequent trials. The extent to which subjects choose correctly, which usually declines rapidly as the delay interval is lengthened (Wixted, 1989), may be taken as a measure of how well they remembered the sample stimulus.
Medin (1976, p. 115) argued that, "Animal research may help to disentangle those aspects of human memory that depend uniquely on language from those that do not". Zentall (1973, p. 126), also pointed out, that in DMTS, "the fact that an animal can perform at a high level accuracy on control trials involving only the sample and comparison stimuli presentation indicates that the task 'instructions' have been learned" in a way comparable to humans who have been given explicit instructions. Laming and Scheiwiller (1985) argued that in many studies of memory, with both humans and animals, similar functions have been obtained under similar contingencies. The present author proceeds from the view that it is appropriate to assume, on the grounds of neutrality, parsimony, and the desirability of conceptual coherence in psychology, that similar laws may well apply to human as apply to other animal behaviour, unless or until in a particular domain, the evidence shows otherwise.

The preferred measure of remembering strength in White's Direct Remembering theory, is \( \log d \), a bias-free measure of discriminability (Davison & Tustin, 1978; McCarthy, 1983; McCarthy & White, 1987), derived from the generalised matching law. In a situation where a subject has only two choices, and each time they choose, the only possible outcomes are that the choice is either correct, or incorrect, then \( \log d \) is half of the logarithm (to base 10) of, the ratio of correct to incorrect responses made to the one stimulus multiplied by the ratio of correct to incorrect responses made to the other stimulus. Or, as Adamson, Foster, and McEwan (2000, p. 154) put it, \( \log d \), "is the logarithm of the geometric mean of the ratios of correct to incorrect responding on trials with the two different sample stimuli". As applied to the present study, in which the choice must be made between a red stimulus and a green stimulus, the equation for \( \log d \) is:

\[
\log d = 0.5 \log \left( \frac{\text{Correct red}}{\text{Incorrect red}} \cdot \frac{\text{Correct green}}{\text{Incorrect green}} \right)
\]

\( \log d \) is easier to interpret at very high or very low discriminability than proportion correct, which is bounded at 0.5 (a chance level of performance) and 1.0 (perfect performance). Sargisson and White (2001), extended the use of \( \log d \) to describe remembering performance in terms not only of temporal distance from time zero, but also in terms of generalisation of discriminative performance along the stimulus dimension of time, as follows:

\[
\log d = a \cdot \exp(-b \cdot t) + c \cdot \exp(-d \cdot |t - T|)
\]
Equation (2) describes a function which looks more or less like Figure 1(b). The first term defines the temporal distance component (the part between the Y-axis and the peak), while the second term defines the generalisation component. Where $T$ is the training delay (the time delay from the origin at which correct remembering was reinforced), $t$ is any given time (i.e., along the stimulus dimension of time) at which the value of the function is being defined.

**Stimulus discrimination**

Behaviour, including making choices, occurs in an environment. If an aspect of that environment provides an occasion, or setting, in which a subject's behaviour will have different consequences than would follow the same behaviour in a different setting, then the setting allows for a conditional choice, or discrimination. Stimulus discrimination has been studied widely with animals, along a range of stimulus dimensions, such as coloured lights, sounds, etc. When values of the training stimulus vary along a dimension, such as colour, this is termed intradimensional training.

**Generalisation**

The concept of stimulus generalisation was articulated by Pavlov (1927/1960), but its systematic experimental study was established by Guttman and Kalish (1956), and reviewed by Honig and Urcuioli (1981). When an organism is trained to respond to a particular stimulus, e.g. by being reinforced for that particular response, some responses will also occur to stimuli on either side of the target stimulus. "A stimulus generalisation gradient is the function obtained when the total number of responses to each of the stimulus values presented during the generalisation test are plotted against the dimension of generalisation" (Rilling, 1977, p. 423). Stimulus generalisation gradients are often used to study the degree of stimulus control achieved, peak performance being normally found at the training stimulus.

**Peak shift**

Hanson (1959), conducted intradimensional discriminative training at two points along a stimulus continuum, namely wavelength of light. Responses to one colour were reinforced. This was called a positive stimulus ($S^+$). Responses to another colour were never reinforced (i.e., extinction was in effect). This was called the negative stimulus ($S^-$). When generalisation tests were carried out at a range of wavelengths around the two trained wavelengths, it was found that highest rate of responding occurred not to the stimulus where correct choices were reinforced ($S^+$), but to a stimulus displaced away from $S^-$, on the far side of $S^+$. This was called peak shift. Overviews of peak shift research have been done by Purtle (1973), Rilling (1977), and Honig and Urcuioli.
(1981). As Cheng, Spetch and Johnston (1997) noted, individual differences exist, and some individuals fail to show the effect (see Purtle, 1973). Peak shift has also been found in classical conditioning studies (e.g. Moore, 1972), but those will not be further considered in this thesis.

As Table 1 shows, the peak shift phenomenon has subsequently been found to occur, both with a wide range of species, and along many very diverse stimulus dimensions. If Sargisson and White (2001) are correct in interpreting their results as generalisation of discriminative performance around a peak at a delay where correct responding was reinforced, then it may be that peak shift might occur in discrimination along the stimulus dimension of time, that is, in remembering, in a manner analogous to the peak shifts found in response rates along other stimulus dimensions. Indeed if White's DR theory is correct in treating discrimination along the temporal dimension as being in principle no different from discrimination along other dimensions, then peak shift ought to occur also in remembering.

Figure 1 (c) illustrates how generalisation of discriminative performance might occur around two peaks at two time delays, T1 and T2, if correct responding were reinforced at those delays. Figure 1 (d) then illustrates how peak shift in remembering might occur. If correct responding at time delay T1 were reinforced poorly or not at all, but correct responding at time delay T2 were reinforced richly, then trials at T1 could function as an S−, while trials at T2 could function as an S+, just like the S− and S+ stimuli along other dimensions such as light wavelength. Hence, during generalisation testing, the highest remembering accuracy might occur at a delay longer than the S+ delay, and at which correct responding would never have been reinforced. The purpose of the present study was to see if such a peak shift would occur in remembering.

Reasons for the particular design of the study

No previous studies existed on peak shift in remembering. Therefore, in order to gain guidance from the existing literature on peak shift in response rates, it was necessary to assume, at least as a working hypothesis, that discriminative performance along the temporal dimension might follow similar patterns under similar contingencies to those previously found with response rates along other dimensions.

The possibility of trying to obtain peak shift toward the y-axis was considered, but rejected for two reasons. First, it would be theoretically much more interesting to find a peak shift in remembering away from the y-axis. Second, it would be difficult to show that any improved performance closer to the y-axis constituted a peak shift rather than the ordinarily expected better remembering at shorter time delays. For any given data
<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Date</th>
<th>Subjects</th>
<th>Stimulus dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ames &amp; Yarczower</td>
<td>1965</td>
<td>Goldfish</td>
<td>Light wavelength</td>
</tr>
<tr>
<td>Baron</td>
<td>1973</td>
<td>Humans</td>
<td>Sound frequency</td>
</tr>
<tr>
<td>Bloomfield</td>
<td>1967</td>
<td>Pigeons</td>
<td>Angularity (line-tilt)</td>
</tr>
<tr>
<td>Blough</td>
<td>1973</td>
<td>Pigeons</td>
<td>Light wavelength</td>
</tr>
<tr>
<td>Bushnell &amp; Weiss</td>
<td>1980</td>
<td>Pigeons</td>
<td>Light wavelength</td>
</tr>
<tr>
<td>Capehart &amp; Pease</td>
<td>1968</td>
<td>Humans</td>
<td>Weight lifted</td>
</tr>
<tr>
<td>Cheng, Spetch &amp; Johnston</td>
<td>1997</td>
<td>Pigeons</td>
<td>Spatial location</td>
</tr>
<tr>
<td>Doll &amp; Thomas</td>
<td>1967</td>
<td>Humans</td>
<td>Light wavelength</td>
</tr>
<tr>
<td>Dougherty &amp; Lewis</td>
<td>1991</td>
<td>Horses</td>
<td>Circle size</td>
</tr>
<tr>
<td>Dukhayyil &amp; Lyons</td>
<td>1973</td>
<td>Pigeons</td>
<td>Light wavelength</td>
</tr>
<tr>
<td>Dysart, Marx, McLean &amp; Nelson</td>
<td>1974</td>
<td>Pigeons</td>
<td>Light wavelength</td>
</tr>
<tr>
<td>Ernst, Engberg &amp; Thomas</td>
<td>1971</td>
<td>Pigeons</td>
<td>Light intensity</td>
</tr>
<tr>
<td>Galizio &amp; Baron</td>
<td>1979</td>
<td>Humans</td>
<td>Sound frequency</td>
</tr>
<tr>
<td>Gerry</td>
<td>1971</td>
<td>Pigeons</td>
<td>Sound frequency</td>
</tr>
<tr>
<td>Grusec</td>
<td>1968</td>
<td>Pigeons</td>
<td>Light wavelength</td>
</tr>
<tr>
<td>Guttmann</td>
<td>1959</td>
<td>Pigeons</td>
<td>Light wavelength</td>
</tr>
<tr>
<td>Hanson</td>
<td>1959</td>
<td>Pigeons</td>
<td>Light wavelength</td>
</tr>
<tr>
<td>Hedges, Dickinson &amp; Modigliani</td>
<td>1983</td>
<td>Humans</td>
<td>Arm movements</td>
</tr>
<tr>
<td>Hendry, Switalski, &amp; Yarczower</td>
<td>1969</td>
<td>Rats</td>
<td>Sound click-rate</td>
</tr>
<tr>
<td>Honig &amp; Stewart</td>
<td>1993</td>
<td>Pigeons</td>
<td>Relative numerosity</td>
</tr>
<tr>
<td>Klipec, Akins &amp; Koerner</td>
<td>1979</td>
<td>Pigeons</td>
<td>Light wavelength</td>
</tr>
<tr>
<td>Klipec, Lindblom &amp; Lindblom</td>
<td>1979</td>
<td>Rats</td>
<td>Angularity (line-tilt)</td>
</tr>
<tr>
<td>Lyons, Ferraro, Lyons, Sullivan &amp; Downey</td>
<td>1973</td>
<td>Rats</td>
<td>Floor inclination</td>
</tr>
<tr>
<td>Lyons, Klipec &amp; Steinsultz</td>
<td>1973</td>
<td>Rats</td>
<td>Angularity (line-tilt)</td>
</tr>
<tr>
<td>MacKinnon</td>
<td>1972</td>
<td>Humans</td>
<td>Weight lifted</td>
</tr>
<tr>
<td>Mariner &amp; Thomas</td>
<td>1969</td>
<td>Pigeons</td>
<td>Light wavelength</td>
</tr>
<tr>
<td>Marsh</td>
<td>1972</td>
<td>Pigeons</td>
<td>Light wavelength</td>
</tr>
<tr>
<td>McCoy &amp; Lange</td>
<td>1969</td>
<td>Monkeys</td>
<td>Gravity</td>
</tr>
<tr>
<td>Newlin, Rodgers &amp; Thomas</td>
<td>1979</td>
<td>Humans</td>
<td>Light intensity</td>
</tr>
<tr>
<td>Nicholson &amp; Gray</td>
<td>1971</td>
<td>Humans</td>
<td>Angularity (line-tilt)</td>
</tr>
<tr>
<td>Authors</td>
<td>Year</td>
<td>Species</td>
<td>Stimulus</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>------</td>
<td>-------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Nicholson &amp; Gray</td>
<td>1972</td>
<td>Humans</td>
<td>Angularity (line-tilt)</td>
</tr>
<tr>
<td>Pierrel &amp; Sherman</td>
<td>1960</td>
<td>Rats</td>
<td>Sound intensity</td>
</tr>
<tr>
<td>Pierrel &amp; Sherman</td>
<td>1962</td>
<td>Rats</td>
<td>Sound intensity</td>
</tr>
<tr>
<td>Riccio, Urda &amp; Thomas</td>
<td>1966</td>
<td>Pigeons</td>
<td>Floor inclination</td>
</tr>
<tr>
<td>Rudolph &amp; Honig</td>
<td>1972</td>
<td>Chickens</td>
<td>Light wavelength</td>
</tr>
<tr>
<td>Rudolph, Van Houten &amp; Maddox</td>
<td>1971</td>
<td>Pigeons</td>
<td>Light wavelength</td>
</tr>
<tr>
<td>Sloane</td>
<td>1964</td>
<td>Pigeons</td>
<td>Light flicker rate</td>
</tr>
<tr>
<td>Spetch &amp; Cheng</td>
<td>1998</td>
<td>Pigeons</td>
<td>Light duration</td>
</tr>
<tr>
<td>Stevenson</td>
<td>1966</td>
<td>Pigeons</td>
<td>Light wavelength</td>
</tr>
<tr>
<td>Terrace</td>
<td>1964</td>
<td>Pigeons</td>
<td>Light wavelength</td>
</tr>
<tr>
<td>Terrace</td>
<td>1966</td>
<td>Pigeons</td>
<td>Light wavelength</td>
</tr>
<tr>
<td>Terrace</td>
<td>1968</td>
<td>Pigeons</td>
<td>Light wavelength</td>
</tr>
<tr>
<td>Thomas &amp; Burr</td>
<td>1969</td>
<td>Pigeons</td>
<td>Light wavelength</td>
</tr>
<tr>
<td>Thomas, Mood, Morrison &amp; Wiertelak</td>
<td>1991</td>
<td>Humans</td>
<td>Light intensity, &amp; Angularity (line-tilt)</td>
</tr>
<tr>
<td>Thomas, Ost &amp; Thomas</td>
<td>1960</td>
<td>Pigeons</td>
<td>Light wavelength</td>
</tr>
<tr>
<td>Thomas &amp; Setzer</td>
<td>1972</td>
<td>Rats, Guinea pigs</td>
<td>Sound intensity</td>
</tr>
<tr>
<td>Thomas, Svinicki &amp; Vogt</td>
<td>1973</td>
<td>Humans</td>
<td>Light intensity</td>
</tr>
<tr>
<td>Weinberg</td>
<td>1973</td>
<td>Rats</td>
<td>Sound click-rate</td>
</tr>
<tr>
<td>Wheatley &amp; Thomas</td>
<td>1974</td>
<td>Pigeons</td>
<td>Light wavelength</td>
</tr>
<tr>
<td>Wildemann &amp; Holland</td>
<td>1971</td>
<td>Pigeons</td>
<td>Circle size</td>
</tr>
<tr>
<td>Wills &amp; Mackintosh</td>
<td>1998</td>
<td>Pigeons, Humans</td>
<td>Symbolic arrays</td>
</tr>
<tr>
<td>Winton &amp; Beale</td>
<td>1971</td>
<td>Pigeons</td>
<td>Angularity (line-tilt)</td>
</tr>
<tr>
<td>Yarczower, Dickson &amp; Gollub</td>
<td>1966</td>
<td>Pigeons</td>
<td>Light wavelength</td>
</tr>
<tr>
<td>Yarczower, Gollub &amp; Dickson</td>
<td>1969</td>
<td>Pigeons</td>
<td>Light wavelength</td>
</tr>
<tr>
<td>Zentall, Collins &amp; Hearst</td>
<td>1971</td>
<td>Pigeons</td>
<td>Angularity (line-tilt)</td>
</tr>
</tbody>
</table>

**Note.** This table is not an exhaustive list of peak shift studies. It deliberately excludes those studies which investigated but did not find a peak shift. It also does not include classical conditioning studies.

set, a sceptic could simply fit a traditional "decay curve" higher up than otherwise, with the alleged peak lying above this curve, to account for the data. Convincingly defeating this criticism would require both a clear peak shifted from S+ toward the y-axis, and an excellent database showing at the half-second resolution level exactly where the curve
would have been expected to lie in the absence of the differential reinforcement contingency. This information did not yet exist, and was not feasible to obtain within the time and other resources available. It was therefore decided to use trials at a shorter delay as S−, and trials at a longer delay as S+.

Smith (1997), showed that domestic hens with no prior matching to sample training, could learn to choose, under a DMTS procedure, at a delay of 4 s. In Sargisson and White's (2001) study, with training delays of 2, 4, and 6 s, the clearest visible peak in performance during generalisation testing, occurred at the 4-s delay. Using the 4-s delay as the S+ stimulus point therefore seemed to offer the best chance of observing a shift of the expected peak in remembering accuracy at that delay.

White and Cooney (1996), using reinforcement probabilities for correct responses ranging from 0.1 to 0.9, showed that pigeons' DMTS remembering performance at delays of 0.1 s and 4 s was independent. This suggested that a 4-s gap between the two training delays would be too large for the present study, as occurrence of peak shift implies dependence of performance at one delay on the consequences of remembering at the other delay. The close proximity to the origin of an S− at the 0.1-s delay, would also make it difficult both to distinguish between performance at that delay and initial discriminability at time zero, and to discern the shape of the discriminative performance function around S−. Previous work with hens (e.g. Etheredge, 1997), has shown that hens can respond differentially at delays of 2 and 4 s after DMTS training. Comparability with Sargisson and White's delays, linearly spaced 2 s apart, was also desirable. It was therefore appropriate to use trials at the 2-s delay as the S−.

While Hanson (1959) used a variable interval (VI) schedule at S+, and non-reinforcement at S−, Gutman (1959) showed that a lower rate of reinforcement at S− (as opposed to extinction used by Hanson) compared to that at S+ still produced a peak shift. This would help to maintain responding during generalisation probe sessions. Wheatley and Thomas (1974) obtained a peak shift from a VI12/VI24 reinforcement schedule. These results, together with those of White and Cooney (1996), suggested that a reinforcement schedule whereby correct responses to S− are reinforced with probability 0.1, while correct responses to S+ are reinforced with probability 0.9, should be able to produce peak shift in DMTS remembering performance.

Generalisation testing in previous DMTS studies has usually been done either at linearly or logarithmically spaced multiples of 2-s delays. However, it was felt in this case, that a 2-s gap might be too big to allow peak shift to be detected. Theoretical explanation of either the occurrence or not of peak shift, would also be facilitated by a
more closely defined discriminative performance function. Testing intervals of 0.5 s were therefore considered to be appropriate.

Testing only at delays from 2 s to 6 s, would minimise the number of delays per series and so maximise the number trials per session. However, as Cheng, Spetch and Johnston (1997) advised, "the potential for range effects should be considered in all studies of peak shift" (p. 470). In order to exclude the possibility of an adaptation-level effect (Helson, 1947), causing any peak shift found, the generalisation test delay set needs to be arrayed symmetrically around the 2-s and 4-s training delays (assuming linear timing, after Sargisson and White, 2001). Testing at delays from 0 to 6 s would meet this requirement.

Generalisation testing is often done in extinction, after Guttman and Kalish (1956). However, the standard session length in this lab is 40 min, and extinction was undesirable, as a baseline performance measure was needed and so ongoing responding was required after the testing. Continuing to reinforce correct responses at the training delays during generalisation probe sessions, while not reinforcing correct responses at the novel test delays tends to produce sharper generalisation gradients (Pierrel, 1958). Nevertheless, Gerry (1971) maintained responding at S+ with a VI 60 s schedule during generalisation testing, but still found a peak shift. Sargisson and White (2001) maintained responding during generalisation testing, by adding 3 reinforced trials at the training delay to each block of 9 trials. It therefore seemed reasonable to include 3 reinforced trials per block at both the S− and S+ delays in the present study.

There is mixed evidence as to effect of the number of test sessions, and of the delay between training and testing, on the likelihood of finding peak shift. Pierrel and Sherman (1960) used one generalisation test session daily for 7 days, and found peak shift occurring only on the first four tests. Wildemann and Holland (1971) found peak shift occurring only on the first three tests. Thomas, Ost and Thomas (1960) administered generalisation tests 24 hours, 1 week or 3 weeks after the end of training, and found that all produced similar gradients. In contrast, Perkins and Weyant (1958), found flatter gradients after 1 week delay to testing. Taking all the above factors into account, it was decided to run three test sessions on successive days for the first generalisation probe condition (GP1).

For ease of presentation, this study has been divided into four consecutive experiments, each with a brief discussion explaining the reasons for conducting the succeeding condition. A general overview of the results will then be presented, to draw attention to some features which are not so clear from the separate experiment results.
Experiment 1

Method

Subjects

The subjects were six Shaver-Starcross domestic hens (*Gallus domesticus*), numbered 21 to 26, aged 2 years and 4 months at the start of the study. They had previous experience of operant key pecking on basic reinforcement schedules, but no matching-to-sample or delayed matching-to-sample experience. Housing was in individual cages with free water access, in a ventilated room, lit for 12 hours per day with two 100-W light bulbs. The hens were weighed daily before experimental sessions, and maintained at 80% (± 10%) of free-feeding bodyweights, usually by the wheat reinforcements obtained in session, but when necessary, by post-feeding in the home cage with commercial laying pellets. Grit and vitamins were supplied weekly.

Apparatus

The experiment was run in a chamber made of 20 mm thick woodchip particleboard, with internal dimensions of 415 mm deep, 575 mm wide, and 535 mm high. A 22-mm high removable metal tray with 3.5-mm thick wire mesh woven in 32-mm squares, covered the floor.

Three hinged 3 mm thick translucent plastic keys, diameter 30.5 mm, centres 105 mm apart horizontally and 407 mm up from the floor on the right hand wall, served as both stimulus and response operand. They were backlightable Red or Green by LED ('MARL' brand SX6 Midget Flange 28-V dc LED, from Farnell Electronics). The dominant wavelengths (as measured from the subjects' position, with a Minolta Chroma Meter CS-100) were 612 Nm for a key lit Red and 573 Nm (actually yellow-green) for a key lit Green. Luminance for both was 25 cd/m². A force of at least 0.2 N and travel of 2.5-3 mm at the lower end was required to operate the microswitch behind the key. A piezoelectric beeper (PKB5-A30 from RS Components) was mounted on the outside, below and to the left of the keys (centred 255 mm from the floor and 82 mm in from the back wall).

A food hopper could be raised to a 20 mm x 30 mm hole in the floor of a recess located centrally below the keys (120 mm from the floor at its bottom edge, 70 mm wide x 100 mm high x 75 mm deep), with a 24-V/1-W light bulb on its far wall.

An 80 mm covered fan mounted in the top of the opposite wall from the key panel, provided ventilation and a small amount of masking noise. There was no
houselight, but the fan and slight cracks in the box and around the key housings allowed a small amount of daylight into the white-painted box.

The equipment was remotely controlled, and data recorded by MEDSTATE NOTATION™ programs (see Appendix D) running on MED-PC™ Version 2.0 for MS-DOS® software, on a 486 IBM®-compatible personal computer, via a MED ASSOCIATES interface and a 28-V dc power supply. Key summary data for each session were also manually recorded.

Procedure

General

Sessions were run once per day, generally six days per week. A session was terminated after no more than 2400 s, but where possible, upon completion of the first trial after 2370 s, to facilitate training by minimising interruptions in mid-trial.

The operant unit was a peck of sufficient force and travel to actuate a key microswitch. Every such peck to a lit key produced a brief (approximately 0.05 s) beep as a discriminative stimulus. Pecks to unlit keys were recorded but produced no consequences for the subjects. Reinforcement consisted of 2.5 s access to wheat.

DMTS Training condition 1 (TR1)

A trial began with the centre key lighting up either Red or Green (selected pseudorandomly, with the limitation that the same colour could not occur more than three times in a row, according to a numerical series adapted from Gellermann, 1933). The subject was then required to peck the centre key 5 times. Upon the fifth peck, the key light went out. After a delay interval of either 2 s or 4 s (selected pseudorandomly as above), the side keys were lit, either Red on the left and Green on the right, or vice versa (selected pseudorandomly as above). A peck on the side key which was lit the same colour as the centre key had been, was defined as a correct choice. A peck on the other side key was defined as an incorrect choice. A trial remained in progress until a choice of either the left or right side key was made, whereupon the key lights immediately turned off, and one of three possible consequences occurred for 2.5 s.

An incorrect choice always resulted in 2.5 s of blackout. A correct choice always resulted in 2.5 s illumination of the food hopper light as a discriminative stimulus for correct responding. Simultaneously, reinforcement occurred with probability 0.1 for 2-s delay trials, but with probability 0.9 for 4-s delay trials. Thus about 90% of the time for 2-s delay trials and 10% of the time for 4-s delay trials correct choices resulted in
hopper-light illumination only. A 10-s inter-trial interval (ITI) followed each consequence.

Under these contingencies a number of problems developed. One subject, 24, responded so little that learning seemed unlikely. All the others displayed extreme bias for either the left or right key. Latencies to choice fell rapidly at first, but then increased again. By the tenth session responding declined to near zero for four of the six subjects. It was possible that responding might eventually have increased, but the time available for the study was limited and the extinction was undesirable. Zentall, Collins and Hearst (1971) found that a previously positive S− still caused peak shift of similar magnitude. The main difference compared to that of a stimulus which had only ever been negative, was that the effect was more easily reversed. Therefore the probability of reinforcement for correct responses (Pr(rft|Corr)) at 2 s was increased to 0.9 until the DMTS task at two delays was learned and performance stabilised. This took 38 sessions. Stability of log $d$ was determined graphically, so that with the help of 8-period moving average trendlines, no trend was evident. A generalisation probe was then conducted.

**Generalisation probe 1 (GP1)**

Three sessions were run on three successive days. Eleven probe delays were used: 0, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, and 6.0 s. All correct 2-s and 4-s responses were reinforced. Responses at other delays were never reinforced. To ensure enough reinforcements to maintain responding, two extra 2-s and two extra 4-s trials were added, making a total of 15 trials per "series". Within each series, the sequence of delays was determined by random selection without replacement. The sessions again ended after 2370-2400 s, so the final series in a session was not necessarily completed. The effect of having one fewer trial per session at certain delays was found to be small, especially after averaging over sessions. It was also considered to be less undesirable than losing data by using only completed series or variably extending session length to allow subjects to complete series.

For each session, the mean value of log $d$ at each delay was calculated according to Equation (1), using as inputs the mean scores over all six subjects for Green correct, Green incorrect, Red correct and Red incorrect choices. The mean log $d$ per session was then plotted. It was expected that at this level there would be no 100% correct performance and hence no zero incorrect score, thus avoiding the need to adjust the result by adding 0.5 to each cell as per Hautus (1995). However, to facilitate comparisons with individual subjects' log $d$'s which did require Hautus' adjustment, the
mean result was eventually also adjusted. After the end of the study, the mean log $d$ calculated instead as the arithmetic mean of the six subjects’ individual Hautus-adjusted log $d$’s was also graphed. The differences between the two methods were very small, and in no case changed the basic shape or interpretation of the graphs (see Appendix D).

Close inspection of the results of GP1 revealed that 100% correct performance was associated with substantially higher values of log $d$ at 2 s and 4 s delays than at other delays. It was then realised that all three 2-s delay trials and all three 4-s delay trials per series had been used in the calculation of the value of log $d$ (Hautus adjusted), inflating its value at those two delays compared to the surrounding probe delays. After the end of the study it was realised that dividing each of the input values for Green correct, Green incorrect, Red correct and Red incorrect choices at 2-s and 4-s trials by three, would give close, unbiased estimates of the values that would have resulted if the program had only counted 1 in 3 in the first place. However, at the time it was decided that another generalisation probe was necessary, and hence another training condition was also run.

Training condition 2 (TR2)
This consisted of 38 more sessions with Pr(rft|Corr) = 0.9 at 2 and 4s, according to the same criteria as condition TR1. Another generalisation probe was then conducted.

Generalisation probe 2 (GP2)
All details for GP2 were the same as for GP1 except that: a) The program was modified to count only the first of every three 2-s and 4-s trials for the purpose of calculating log $d$, and b) The third session in GP1 was deemed redundant. Therefore, to minimise the risk of the subjects learning generalisation probes as separate conditions, only two sessions were run for GP2.

Training condition 3 (TR3)
This was a return for 4 sessions to the same conditions as TR2. Due to time pressure, the full stability criterion was dispensed with.

Peak shift training condition 1 (PS1)
All details for PS1 were the same as for TR2 except that the probability of reinforcement for correct responses on 2-s delay trials was reduced from 0.9 to 0.1. The condition ran for 46 sessions. Generalisation probe number 3 was then conducted.
**Generalisation probe 3 (GP3)**

All details for GP3 were the same as for GP2.

**Results**

The cumulative raw data for all the training conditions in the study (TR1, TR2, PS1, TR3, PS2, PS4, PS5, and PS6) are presented both in printed form in Appendix B, and electronically in Appendix D (the attached CD-R disc). The cumulative raw data for all the generalisation probe conditions in the study (GP1, GP2, GP3, GP4, GP5, and GP6) are presented both in printed form in Appendix C, and electronically in Appendix D. The consecutive within-session event data for the entire study are presented electronically in Appendix D.

The results for the initial training condition, TR1, are not presented here, but will be presented later in this thesis as part of a graphical overview of the whole study. Figure 2-1 presents time-series data for the mean discriminative performance over all 6 hens (Hautus adjusted log $d$, calculated as described in the procedure for GP1), at 2 and 4-s delays, for conditions TR2, GP2, PS1 and GP3. Figure 2-2 presents time-series data for the mean latency to choice over all 6 hens, at 2 and 4-s delays, for conditions TR2, GP2, PS1 and GP3. The individual subjects' time series data for both latency and discriminative performance were more variable than the mean data, but showed the same patterns. They are therefore not presented in this results section, but may be viewed electronically in the folder Excelcumdata in Appendix D. Figure 3-1 presents the individual and mean data for discriminative performance (Hautus adjusted log $d$, calculated as described in the procedure for GP1) at all test delays, for the generalisation probe conditions, GP1, GP2 and GP3. Figure 3-2 presents the individual and mean data for proportion correct at all test delays, for conditions GP1, GP2 and GP3. Figure 3-3 presents the individual and mean data for latency to choice at all test delays, for conditions GP1, GP2 and GP3. The data in the above-mentioned figures are described below.

In GP1 (Figure 3-1, open circles), the mean forgetting function was a broad curve with a global peak in discriminative performance at 2 s. No subject performed best at 0 s, or displayed a monotonically declining function. The individual peaks, after training at delays of 2 and 4 s, occurred at delays ranging from 1.5 to 4.5 s. Thus, while the
Figure 2-1. Discriminability (Hautus adjusted log \( d \)) at 2 and 4 s delays, for conditions TR2 to GP3 inclusive - mean across all Hens 21-26. Note. log \( d \) values in generalisation probes cannot be directly compared to those in training sessions, as there were 7 trials at each probe delay versus about 50 in training.

Figure 2-2. Mean latency to choice at 2 and 4 s delays, for conditions TR2 to GP3 inclusive - mean across all Hens 21-26.
Figure 3-1. The mean discriminability (Hautus adjusted log $d$) per session, for the three sessions of GP1, two sessions of GP2 and two sessions of GP3, for Hens 21-26, and the mean across all six hens.
Figure 3-2. The mean proportion correct per session, for the three sessions of GP1, two sessions of GP2 and two sessions of GP3, for Hens 21-26, and the mean across all six hens.
Figure 3-3. The mean latency to choice per session, for the three sessions of GP1, two sessions of GP2 and two sessions of GP3, for Hens 21-26, and the mean across all six hens.
individual results were obviously more variable than the mean, none showed a basically different pattern from it. Discriminative performance at delays on either side of the 2-s delay was more like a flattened normal curve than a negative exponential decline. The single-point sub-peak was at the 4.5-s delay rather than at the 4-s training delay. The mean proportions correct (Figure 3-2) followed a similar pattern, ranging from 0.79 at the 0-s delay and 0.78 at the 6-s delay to 0.90 at the 2-s delay. The mean latencies to choice (Figure 3-3) declined from a maximum of 1.6 s at the 0-s delay to 1.2 s at the 2-s delay and then formed a roughly flat function out to a value of 1.2 s at the 6-s delay.

In TR2 (Figure 2-1), mean discriminative performance over the last twenty sessions at the 2-s delay was about 12% higher than at 4 s. There was however, a rough correspondence between discriminative performance levels at the two delays, in that they tended to rise and fall together. Similarly, latencies to choice (Figure 2-2) at the 2 and 4-s delays tended to rise and fall together, generally within a narrow range from 1 to 1.5 s. Mean latency over the last twenty sessions at the 2-s delay was about 5 % higher than at the 4-s delay, but this was mostly caused by an outlier in session 87.

In GP2 (Figure 3-1, solid grey circles), mean discriminative performance compared to GP1 remained virtually the same at delays from 2 to 4 s, and at 5 and 6 s. It rose at 0 and 1 s, but fell at 1.5 and 4.5 s, so that the former sub-peak at 4.5 s became the global minimum. The individual functions were again not consistently different from the mean. Overall, the changes in the mean pattern of discriminative performance from GP1 to GP2 were neither large nor systematic. The mean proportions correct (Figure 3-2) ranged from 0.75 at the 4.5-s delay to 0.90 at the 1-s delay. Mean latency to choice (Figure 3-3) formed a nearly flat function, virtually the same as that in GP1, averaging 1.3 s. Subject 25 however showed a marked increase in latencies.

During PS1 (Figure 2-1), mean discriminative performance over the last twenty sessions at the 2-s delay (where reinforcement probability for correct responses was reduced from 0.9 to 0.1), fell by about 23% compared to TR2. However, at the 4-s delay, where reinforcement probability for correct responses remained the same, at 0.9, mean discriminative performance over the last twenty sessions also fell, by about 10%. Mean latencies to choice (Figure 2-2) increased during PS1, by about 65% (from 1.4 s to 2.3 s) at the 2-s delay, and by about 15% (from 1.3 to 1.5 s) at the 4-s delay. Even with the diverging trend, latencies at the 2 and 4-s delays still tended to rise and fall together, though not as consistently as during TR2.

In GP3 (Figure 3-1, solid black circles), mean discriminative performance showed a clear peak at the 4.5-s delay, increasing by 113% compared to GP2 and 20%
compared to GP1. Discriminative performance at 0 s fell slightly compared to GP2, returning to the same level as in GP1. Discriminative performance at 6 s remained the same in all three probes. Discriminative performance fell in the 1-s to 3-s delay range, so that overall the function was asymmetrically shaped, with a gently rising gradient leading to the 4.5-s peak and then a sharp fall. The mean proportions correct (Figure 3-2) displayed a similar pattern, with minima of 0.75 at the 1.5-s and 5-s delays, while rising to a maximum of 0.92 (up 6% compared to GP1 and 24% compared to GP2) at the 4.5-s delay. Mean latency to choice in GP3 (Figure 3-3), increased compared to GP2 at delays from 0 to 3.5 s, by an average of about 45% (from 1.3 to 1.9 s). Subject 25’s latencies increased sharply in the 1 to 3 s delay range, with a peak of 3.7 s at the 2-s delay. Two subjects departed systematically from the mean pattern. Subject 26’s latencies increased only very slightly but evenly across the range of delays, while 23’s latencies increased most at the longer delays of 3 to 6 s.

Discussion

The mean remembering function in GP1 differed in two respects from the "twin-peak" outcome hypothesised in Figure 1(c). Firstly, performance at delays on either side of the 2-s delay was more like a flattened normal curve than a negative exponential decline. Secondly, the single-point sub-peak was at 4.5 s rather than at the 4-s training delay. It is not clear at this stage whether this is significant.

During PS1, (see Figure 2-1), mean discriminative performance at the 2-s delay (where reinforcement probability for correct responses was reduced from 0.9 to 0.1), fell by about 23% compared to TR2. The result in GP1 was basically consistent with the results of Smith (1997) and Sargisson and White (2001), who found maximum mean discriminative performance at a single delay, (2, 4 or 6 s) after training at that delay.

Despite the fact that the probabilities of reinforcement were the same in GP2 as they had been in GP1, the mean discriminative performance function in GP2 looked rather different. The reasons for this are unknown. GP3 showed a peak in the mean discriminative performance function, similar to the one hypothesised in Figure 1(d), the first evidence of peak shift in remembering.

The fact that any such pattern occurred at all, as opposed to a monotonically declining curve, supports the concept of generalisation of remembering performance (Sargisson & White, 2001). It also confirms that the phenomenon of generalisation of discriminative performance along the temporal dimension, may be usefully measured at
the 0.5-s resolution level. The fact that the testing delays were mostly only 0.5 s apart makes it difficult to justify fitting any function whose form differs from that of the data.

The fact that the mean latency to choice at the 2-s delay increased during PS1, prior to the condition in which peakshift occurred, suggested that latency might be a predictor of peakshift. The present study makes a basic analogy between generalisation gradients of discriminative performance and those of response rates. One way of considering the possible precursors to peak shift, would to examine, "discriminability contrast" (analogous to the negative induction, or "behavioral contrast" (Reynolds, 1961), found in response rates). "Latency contrast" could also be used to describe diverging trends, although the theoretical meaning of this would be at best speculative. Actually, positive latency induction, or "contrast" (Skinner, 1938/1991) occurred, i.e., latency increased at both delays during PS1. Discriminability contrast also did not occur (as in PS1 log $d$ went down at the 4-s delay as well as at the 2-s delay).

After a peak shift had been found in remembering, in GP3, it was unclear how best to proceed with the experiment. One option was suggested by the work of Terrace (1966), who found that after further periods of training, each followed by more testing, the magnitude of peak shift gradually reduced in successive tests. This raised the question of whether the same would happen in the present study. It was therefore decided to keep running the experiment to see if peak shift would still occur during generalisation testing after extended training with differential reinforcement, or if it would reduce or vanish.
Experiment 2

As discussed, the purpose of this experiment was to see if peak shift would still occur after extended training with differential reinforcement, or if it would reduce or vanish.

Method

Subjects and apparatus
These were the same as for Experiment 1.

Procedure

Peak shift training condition 2 (PS2)
All details for PS2 were the same as for PS1 in Experiment 1, that is DMTS was in effect with 2 and 4 s delays, with Pr(rft|Corr) = 0.1 at 2 s and 0.9 at 4 s. The condition ran for 28 sessions. Generalisation probe number 4 was then conducted.

Generalisation probe 4 (GP4)
All details for GP4 were the same as for GP3 in Experiment 1.

Results

The results for PS2 are not presented graphically here, but will be presented later in this thesis as part of a graphical overview of the whole study. Over the last twenty sessions of PS2, mean discriminative performance over all 6 subjects (Hautus adjusted log \(d\), calculated as described in the procedure for GP1) at the 2-s delay, was 4% lower than over the last twenty sessions in PS1. However, at the 4-s delay, mean discriminative performance over the last twenty sessions was 8 % higher than in PS1. Mean latency to choice over all 6 subjects, decreased by 8 % at the 2-s delay and 2% at the 4-s delay during the last twenty sessions of PS2 compared to PS1.

Figure 4-1 presents the individual and mean data for discriminative performance (Hautus adjusted log \(d\), calculated as described in the procedure for GP1) at all test delays, for GP4 compared to GP3. Figure 4-2 presents the individual and mean data for proportion correct at all test delays, for GP4 compared to GP3. Figure 4-3 presents the individual and mean data for latency to choice at all test delays, for GP4 compared to GP3. The data in the above-mentioned figures are described below.
Figure 4-1. The mean discriminability (Hautus adjusted log $d$) per session, for the two sessions of GP3 and two sessions of GP4, for Hens 21-26, and the mean across all six hens.
Figure 4-2. The mean proportion correct per session, for the two sessions of GP3 and two sessions of GP4, for Hens 21-26, and the mean across all six hens.
Figure 4-3. The mean latency to choice per session, for the two sessions of GP3 and two sessions of GP4, for Hens 21-26, and the mean across all six hens.
In GP4 (Figure 4-1, solid black circles), mean discriminative performance was 9% higher than in GP3 (Figure 4-1, open circles) at the 2-s delay, while remaining the same as in GP3 at the 4-s delay. However, the mean discriminative performance function at delays around the two training delay points changed considerably, mostly forming a broad curve with a local maximum at the 3.0-s delay, and minima at the 1-s and 6-s delays. The global maximum occurred at the 0-s delay, where mean discriminative performance rose by 56% compared to GP3. Mean discriminative performance fell by 34% from its former peak at the 4.5-s delay. The individual discriminative performance functions were not as smooth as the mean function. 4 subjects (21, 22, 23, and 26) broadly conformed to the mean pattern, in that either a local or global maximum in discriminative performance occurred at the 0-s delay, followed by a dip to a local or global minimum at the 1.0, 1.5, or 2-s delay, followed by a rough curve with a peak in discriminative performance at a delay from 3 to 5 s. Subject 24's discriminative performance was arguably consistent with the mean pattern, but was extremely variable across delays, displaying two further dips at the 2 and 4-s delays. Subject 25's discriminative performance function was systematically different from the mean pattern, being roughly flat at delays from 0 to 3.5 s, with a clear peak at the 4-s delay.

The mean proportions correct (Figure 4-2, solid black circles) displayed a similar pattern to the mean discriminative performance, with minima of 0.73 at the 1-s and 6-s delays (down 6% compared to GP3), a local maximum of 0.88 (up 10% compared to GP3) at the 3-s delay, and a global maximum of 0.91 (up 16% compared to GP3) at the 0-s delay. The mean proportion correct was 0.81 (up 3% compared to GP3) at the 2-s delay, and 0.81 (down 5% compared to GP3) at the 4-s delay. The mean proportion correct at the 4.5-s delay was 0.81 (down 12% from its former peak in GP3).

Mean latency to choice in GP4 (Figure 4-3, solid black circles), formed a slightly curved function, declining from a maximum of 2.0 s at the 1-s delay, to a minimum of 1.1 s at the 4.5-s delay, and then rising again to 1.2 s at the 6-s delay. This generally reflected the individual patterns well, except for Subject 25, whose latencies were more variable across delays than the other subjects'. Compared to GP3 (Figure 4-3, open circles), the mean latencies in GP4 decreased at delays from 1.5 s to 6 s, by an average of about 18% (from 1.7 to 1.4 s). However, this relative mean decrease was caused mostly by Subjects 23 and 25, whose latencies fell sharply, to levels close to those they displayed in GP1 (Figure 3-3, open circles). Mean latency at the 0-s delay, reflecting all the individual results well, remained virtually the same as in GP3, at 1.9 s.
Discussion

Repeated intradimensional discrimination training has been found to result in steepening generalisation gradients around a single S+ (Farthing & Hearst, 1968; Hirota & Clarkson, 1973). If, as per Spence's (1937) theory, peak shift is due to the interaction of a gradient of excitation around S+ and a gradient of inhibition around S−, then the sharpening of both gradients would tend to produce a bimodal combined gradient, with peaks centred on S− and S+ (Kalish & Guttman, 1957), and to reduce or eliminate peak shift.

Dysart, Marx, McLean and Nelson (1974) on the contrary, found that three of their six subjects displayed a curve with a maximum at a stimulus point in-between S− and S+, which they described as, "some type of summation" (p. 468). In the present study, GP4 also resulted in such a curve, not consistent with steepening gradients over extended training, for mean discriminative performance. Zeiler (1969), suggested that there is no problem in repeatedly testing the same subjects, but clearly something happened here from GP3 to GP4, for a reason or reasons unknown.

Some specific changes in outcomes in relation to contingency changes were that: (a) discriminative performance fell at the 2-s delay compared to PS1 and rose at the 4-s delay, but this did not lead to peak shift in GP4, (b) within GP4, discriminative performance at the 2 and 4-s delay points was almost the same as in GP3, i.e., it rose 9% at the 2-s delay, but was still lower at the 2-s delay than at the 4-s delay. At the same time, however, the discriminative function around these two points changed considerably (Figure 4-1), and (c) discriminative performance at the 2-s delay fell 4% in PS2 compared to PS1, but rose 9% in GP4 compared to GP3, i.e., the change in the generalisation probe condition was in the opposite direction to that in the preceding training condition. It is, in short, difficult to discern a consistent pattern in these discriminative performance functions during the training condition, which would predict the occurrence of peak shift in the subsequent generalisation probe condition.

Both from the literature, and from the empirical data in the study thus far, there were contradictory and confusing indications. It was therefore decided to continue running the experiment, and once the mean discriminative performance had stabilised again, to do another generalisation probe.
Experiment 3

Initially, it was intended just to continue running the experiment under the same contingencies as in Experiment 2, and see how the generalisation gradients would evolve over repeated periods of training and testing. However, during this period, another theoretically interesting possibility occurred.

Grusec (1968) found that punishment with electric shock caused a bigger peak shift than extinction alone at S−. For ethical reasons it is against the practice of this lab to use electric shock, so the most aversive contingency that we can apply at S− is extinction.

The purpose of this experiment was thus to see if peak shift would re-occur if the probability of reinforcement at the S− delay of 2 s was reduced from 0.1 to 0.

Method

Subjects and apparatus

These were the same as for Experiment 1.

Procedure

Peak shift training condition 3 (PS3)

All details for PS3 were the same as for PS1 in Experiment 1, that is, DMTS at 2- and 4-s delays, with Pr(rft|Corr) = 0.1 at 2 s and 0.9 at 4 s. The condition ran for 17 sessions. No generalisation probe was then conducted. Instead, another training condition, with altered contingencies followed immediately.

Peak shift training condition 4 (PS4)

All details for PS4 were the same as for PS1, except that Pr(rft|Corr) at the 2-s delay was reduced to 0. The condition ran for 28 sessions. Generalisation probe number 5 was then conducted.

Generalisation probe 5 (GP5)

Details for GP5 were mostly the same as for GP2-GP4, except that Pr(rft|Corr) at the 2-s delay was reduced from 1 to 0, so that the probe condition would better reflect the contingencies of the preceding training condition.
Figure 5-1 presents time-series data for the mean discriminative performance over all 6 hens (Hautus adjusted log \( d \), calculated as described in the procedure for GP1), at 2 and 4-s delays, for conditions PS3, PS4 and GP5. Figure 2-2 presents time-series data for the mean latency to choice over all 6 hens, at 2 and 4-s delays, for conditions PS3, PS4 and GP5. The individual subjects' time series data for both latency and discriminative performance were more variable than the mean data, but showed the same patterns. They are therefore not presented in this results section, but may be viewed electronically in the folder Excelcumdata in Appendix D. Figure 6-1 presents the individual and mean data for discriminative performance (Hautus adjusted log \( d \), calculated as described in the procedure for GP1) at all test delays, for the generalisation probe condition GP5 compared to the previous probe condition, GP4. Figure 6-2 presents the individual and mean data for proportion correct at all test delays, for GP5 compared to GP4. Figure 6-3 presents the individual and mean data for latency to choice at all test delays, for GP5 compared to GP4. The data in the above-mentioned figures are described below.

During PS3 (Figure 5-1, sessions 175 to 191), mean discriminative performance (Hautus adjusted log \( d \), calculated as described in the procedure for GP1) over the last ten sessions at the 2-s delay (where reinforcement probability for correct responses remained at 0.1), was at about the same level as it had been over the last twenty sessions of both PS1 and PS2. However, at the 4-s delay (where reinforcement probability for correct responses remained at 0.9), mean discriminative performance over the last ten sessions of PS3 was at a level about 21% higher than over the last twenty sessions of PS1, and 13% higher than over the last twenty sessions of PS2. During PS4 (Figure 5-1, sessions 192 to 217), mean discriminative performance at the 2-s delay (where reinforcement probability for correct responses was reduced from 0.1 to 0) followed an overall downward trend, while at the 4-s delay (where reinforcement probability for correct responses remained at 0.9) it stayed on average at the same level. Over the last ten sessions of PS4 mean discriminative performance at the 2-s delay was on average at a level about 46% lower than it had been in TR2, and 33% lower than in PS1, PS2 and PS3. At the 4-s delay, mean discriminative performance over the last ten sessions of PS4 was roughly at the same level as in TR2 (down 5%) PS1 (up 2%) and PS2 (down 5%), but 16% lower than in PS3. Within these overall trends, a cyclical trend was
**Figure 5-1.** Discriminability (Hautus adjusted log $d'$) at 2 and 4 s delays, for conditions PS3 to GP5 inclusive - mean across all Hens 21-26.

Note. log $d'$ values in generalisation probes cannot be directly compared to those in training sessions, as there were 7 trials at each probe delay versus about 50 in training.

**Figure 5-2.** Mean latency to choice at 2 and 4 s delays, for conditions PS3 to GP5 inclusive - mean across all Hens 21-26.
Figure 6-1. The mean discriminability (Hautus adjusted log \( d \)) per session, for the two sessions of GP4 and two sessions of GP5, for Hens 21-26, and the mean across all six hens.
Figure 6-2. The mean proportion correct per session, for the two sessions of GP4 and two sessions of GP5, for Hens 21-26, and the mean across all six hens.
Figure 6-3. The mean latency to choice per session, for the two sessions of GP4 and two sessions of GP5, for Hens 21-26, and the mean across all six hens.
arguably discernible, with discriminative performance at the 2 and 4-s delays alternately converging and diverging. Thus, in session number 195, performance at the two delays was very close. It then diverged, but came together in session 201; diverged again but converged by session 204; diverged again but converged by session 210. It then diverged, and was possibly converging again when GP5 was carried out.

Mean latency to choice over all 6 subjects, at the 2-s delay, during the last ten sessions of PS3 was at an average level of about 1.9 s (15% lower than in PS1, and 8% lower than in PS2). Mean latency to choice over all 6 subjects, at the 4-s delay, during the last ten sessions of PS3 was at an average level of about 1.4 s (6% lower than in PS1, and 5% lower than in PS2). During PS4, the mean latencies to choice over all 6 subjects changed systematically at both delays. At the 2-s delay, the mean latency to choice varied considerably but increased sharply overall during PS4, to an average of about 4.5 s over the last ten sessions (up 98% compared to PS1, 114% compared to PS2 and 133% compared to PS3). At the 4-s delay, the mean latency to choice fell during PS4, to an average of about 1.2 s over the last ten sessions (down 21% compared to PS1, 20% compared to PS2 and 16% compared to PS3).

In GP5 (Figure 6-1, solid black circles), mean discriminative performance at the 2-s delay (where reinforcement probability for correct responses was reduced from 0.1 to 0), fell by 46% compared to GP4 (Figure 6-1, open circles). Mean discriminative performance also fell compared to GP4 at the 1.5-s (down 31%), 2.5-s (down 25%), and 3-s (down 22%) delays. At delays from 3.5 s to 6 s, mean discriminative performance followed the same pattern in GP5 as in GP4, but at a slightly higher level, increasing 5% at the 3.5-s delay, 2% at the 4-s delay (where reinforcement probability for correct responses remained at 1, as in previous generalisation probes), and 12% at the 4.5 and 5-s delays, then falling 9% compared to GP4 at the 6-s delay. Overall, the mean discriminative performance function displayed a global maximum at the 3.5-s delay, a local maximum at the 1-s delay, and local minima at the 2-s and 6-s delays. The global minimum occurred at the 0-s delay, where mean discriminative performance fell by 65% compared to GP4.

The mean proportions correct (Figure 6-2, solid black circles) displayed a pattern mostly similar to that of the log \(d\) values, with a global maximum of 0.87, the same as in GP4 (Figure 6-2, open circles), at the 3.5-s delay, local minima of 0.70 (down 13% compared to GP4) at the 1.5-s delay and 0.71 (down 4% compared to GP4) at the 6-s delay, a local peak of 0.74 (the same as in GP4) at the 1-s delay, and a global minimum of 0.66 (down 28% compared to GP4) at the 0-s delay.
Mean latency to choice over all six hens increased sharply in GP5 (Figure 6-3, solid black circles), compared to GP4 (Figure 6-3, open circles), at delays in the range from 0 s to 2.5 s. The maximum mean latency of 7.9 s (up 288% compared to GP4) occurred at the 1-s delay. The mean latency at the 2-s delay (where reinforcement probability for correct responses was reduced from 0.1 to 0) was 5.9 s (up 270% compared to GP4). The mean latencies at the 0-s, 1.5-s, and 2.5-s delays were 6.9 s, 6.3 s and 3.1 s (up 268%, 266% and 113% respectively compared to GP4). No subject displayed maximum latency to choice at the 2-s delay. The individual subjects’ maximum latencies occurred at delays of 0 s (Subject 21), 1 s (Subjects 23, 24 and 25) or 1.5 s (Subjects 22 and 26). While the magnitude of the increase in latency at delays from 0 to 2.5 s varied considerably between subjects (from an average of 52% for Subject 22 to an average of 572% for Subject 21), all the individual latency patterns were consistent with the mean pattern of increase in latencies at the shorter delays. However, Subject 23 also showed an increase in latency, averaging 45% compared to GP4, at delays in the range from 3 s to 6 s. Subject 21 showed an outlier from the pattern, with a latency of 7.8 s at the 3.5-s delay.

Discussion

The most notable result of GP5, was that there was not a clear peak shift in discriminative performance, but there was a small "area shift" from GP4 to GP5 (Rilling, 1977; Terrace, 1964). Log \( d \) and latency to choice in GP5 followed different patterns. The minimum discriminative performance was at the \( S^- \) delay of 2 s, while the maximum latency was at the 1-s delay, which coincided with a local peak in discriminative performance. None of the individual peaks or dips in discriminative performance coincided with dips or peaks in latency to choice. Despite the large increase in latencies at shorter delays, there was not a clear peak shift in discriminative performance toward longer delays. Taken together with similarly inconsistent results from Experiments 1 and 2, it seems reasonable to conclude that log \( d \) and latency to choice are functionally largely independent in DMTS. Also, the sharp increase in latencies during PS4 did not lead to a peak shift in discriminative performance in GP5.

While latency to choice did not seem to be predictive of peak shift in discriminative performance in GP5, there was arguably a latency peak shift. Maximum latency to choice occurred neither at the \( S^- \) delay of 2 s, nor at the 0-s delay as would be expected if subjects were categorising dichotomously long/short (Spetch & Cheng,
However, the latency function did not so much "shift" left or right in GP5 compared to GP4, it just showed a large increase in latency at the same delay of 1 s and delays around it. Positive latency contrast was not found in PS4, but neither was discriminability contrast. Instead, log $d$ decreased at both delays. Figure 5-1 shows the cyclical converging/diverging pattern which evolved over PS4. Reducing the probability of reinforcement from 0.1 to 0, led to a much bigger increase in latency than reducing the probability of reinforcement from 0.9 to 0.1 in PS1. This could be regarded as indicating a relatively higher degree of "aversiveness" for extinction.

A large fall in discriminative performance occurred at the 2-s delay and delays around it from 1 to 3 s, due to the reduction in the probability of reinforcement from 0.1 to 0, but very small effect on discriminative performance at delays of 3.5 s and longer suggests that the discriminative task had been learned too well for changes in contingencies at the S⁻ delay to any longer substantially affect performance at the S⁺ delay. In other words, the generalisation gradients around S⁻ and S⁺ (Farthing & Hearst, 1968; Hirota & Clarkson, 1973) may have been steepening. If peak shift is caused by the interaction of a gradient of excitation around S⁺ and a gradient of inhibition around S⁻, then steepening generalisation gradients would tend to reduce peak shift (Honig & Urcuioli, 1981). In White and Cooney's (1996) terms, discriminative performance at the two delays had become relatively independent. This raised the question of how it might be possible to cause peak shift to re-occur.

Many studies have found a bigger peak shift when S⁻ and S⁺ are closer together. (Baron, 1973; Hanson, 1959; Pierrel & Sherman, 1962; Thomas, 1962). The opposite effect was found by Hearst (1968) for the group with smallest difference between S⁻ and S⁺, i.e., peak shift increased as the separation between S⁻ and S⁺ increased (see also Thomas, Svinicki & Vogt, 1973; Thomas, Mood, Morrison & Wiertelak, 1991). It is possible that S⁻ and S⁺ were too close to be discriminated at the closer position.

From the consistent gradient between S⁻ and S⁺ in GP5, it was clear that the subjects in the present study could discriminate in 1-s intervals and that discriminative performance at the 3-s delay was about half way in between performance at the 2 and 4-s delays. This suggested that moving S⁻ from the 2-s position to the 3-s position should again result in a peak shift. It was possible that all six subjects might successfully discriminate in half-second intervals, so that moving S⁻ from 2 s to 3.5 s might produce an even bigger peak shift. However, a 0.5-s gap between delays might also have been too close to be successfully discriminated. While the above studies involved separate groups of subjects trained at various S⁻/S⁺ separations, there seemed every reason to
think that a renewed interaction between the putatively sharpened generalisation gradients around $S^-$ and $S^+$, within the same study, would produce the same effect, and cause a peak shift. Indeed, if the foregoing analysis is correct, then a peak shift ought to occur.
Experiment 4

The purpose of this experiment was to see if, as previously discussed, a short period of training with $S-$ moved closer to $S+$ than in the previous conditions, would cause peak shift to re-occur. It was predicted that moving $S-$ closer to $S+$ would produce a peak shift.

Method

Subjects and apparatus

These were the same as for Experiment 1.

Procedure

Peak shift training condition 6 (PS6)

All details for PS6 were the same as for PS4 in Experiment 3, except that the $S-$ delay was changed from 2 s to 3 s. That is DMTS was in effect with 3 and 4 s delays, with $Pr(rft|Corr) = 0$ at 3 s and 0.9 at 4 s. The condition ran for 7 sessions. Generalisation probe number 6 was then conducted.

Generalisation probe 6 (GP6)

All details for GP6 were the same as for GP5 in Experiment 3, except that the $S-$ delay (at which correct responses were never reinforced), was changed from 2 s to 3 s. The two extra unreinforced 2-s trials per series present in GP5, were replaced in GP6 by two extra unreinforced 3-s trials. As in GP1-5, all correct responses at the 4-s delay were reinforced.

Results

The results for PS6 are not presented graphically here, but will be presented later in this thesis as part of a graphical overview of the whole study. Over the 7 sessions of PS6, mean discriminative performance over all 6 subjects (Hautus adjusted log $d$, calculated as described in the procedure for GP1) at the 3-s delay, was 23% higher than it had been at the 2-s delay over the last seven sessions in PS4 and the four sessions of PS5. At the 4-s delay, mean discriminative performance during the seven sessions of PS6 was 19 % higher than during the last seven sessions in PS4 and 25 % higher than
during the four sessions of PS5. It also displayed a trend, increasing by about 15% over the seven session of PS6. During the seven sessions of PS6, mean latency to choice over all 6 subjects at the 3-s delay, was 3.2 s (28% lower than it had been at the 2-s delay over the last seven sessions in PS4 and 21% lower than at the 2-s delay during the four sessions of PS5). Mean latency to choice over all 6 subjects at the 4-s delay, was 1.3 s (8 % higher than during the last seven sessions in PS4 and 16 % higher than during the four sessions of PS5). Figure 7-1 presents the individual and mean data for discriminative performance (Hautus adjusted log \( d \), calculated as described in the procedure for GP1) at all test delays, for GP6 compared to GP5. Figure 7-2 presents the individual and mean data for proportion correct at all test delays, for GP6 compared to GP5. Figure 7-3 presents the individual and mean data for latency to choice at all test delays, for GP6 compared to GP5. The data in the above-mentioned figures are described below.

Overall, the mean discriminative performance function in GP6 (Figure 7-1, solid black circles) showed a clear peak at the 4.5-s delay, a local peak at the 2-s delay, a local minimum at the 3-s delay, and a global minimum at the 1-s delay. Mean discriminative performance at the 2-s, 4.5-s and 6-s delays, increased by 57%, 34% and 66% respectively, compared to GP5 (Figure 7-1, open circles). Mean discriminative performance at the 1-s and 3-s delays fell by 48% and 33% respectively compared to GP5. Mean discriminative performance at the 0-s delay fell by 4% compared to GP5. The individual discriminative performance functions for Subjects 22, 23, 25 and 26 were broadly consistent with the mean pattern, in that they showed global peaks at either the 4-s or 4.5-s delay, a general increase at delays from 4 s to 6 s, a dip at either the 2.5-s or 3-s delay, a local peak at the 1.5-s, 2-s or 2.5-s delay, and a local or global minimum at either the 1-s or 1.5-s delay. Discriminative performance at the 0, 1.0 and 1.5-s delays fell to chance levels for Subject 23, and well below chance levels for Subject 25. Subject 21's discriminative performance function showed a clear peak at the 4.5-s delay, but departed partially from the mean pattern by increasing at delays from 0 s to 2 s, to produce a v-shaped gradient with a clear minimum at the 3-s delay. Subject 24's discriminative performance function departed systematically from the mean pattern, showing not only a clear global peak at the 1.5-s delay, but a general fall at delays from 3 to 5 s. The mean proportions correct in GP6 (Figure 7-2, solid black circles) displayed a similar pattern to that of the log \( d \) values, with a global maximum of 0.94, an increase of 9% compared to GP5 (Figure 7-2, open circles), at the 4.5-s delay,
Figure 7-1. The mean discriminability (Hautus adjusted log $d^*$) per session for GP6 (after 7 sessions of $S^{-} =$ extinction at 3 s), compared to GP5 (after 26 sessions of $S^{-} =$ extinction at 2 s) for Hens 21-26, and the mean across all six hens. Note: for Hen 25, $d^*_{1.0} = -0.36.$
Figure 7-2. The mean proportion correct per session for GP6 (after 7 sessions of $S^-=\text{extinction at 3 s}$), compared to GP5 (after 26 sessions of $S^-\text{ extinction at 2 s}$) for Hens 21-26, and the mean across all six hens.
Figure 7-3. The mean latencies to choice per session for GP6 (after 7 sessions of $S^− =$ extinction at 3 s), compared to GP5 (after 26 sessions of $S^− =$ extinction at 2 s) for Hens 21-26, and the mean across all six hens. Note: for Hen 25, $t_{1.0} = 30.07$. 
a local minimum of 0.76 (down 5% compared to GP5) at the 3-s delay, a local peak of 0.78 (up 10% compared to GP4) at the 2-s delay, and a global minimum of 0.64 (down 14% compared to GP5) at the 1-s delay.

Mean latency to choice over all six subjects followed almost the same pattern in GP6 (Figure 7-3, solid black circles) as in GP5 (Figure 7-3, open circles), with a clear peak of 8.2 s (up 5%) at the 1-s delay. Mean latency at the 0-s delay was 4.5 s (down 35% compared to GP5). However, while all 6 subjects continued to show longer latencies at delays in 0-s to 2.5-s range than at delays in the 3-s to 6-s range, the stability of the mean latency pattern from GP5 to GP6 masked opposite movements by the individual subjects. Subjects 21 and 23 showed a sharp fall in latencies at delays from 0 to 2 s. Subject 26 showed a fall in latency mostly at the 1.5-s delay. Subject 25 showed a sharp increase in her peak latency at the 1-s delay, to 30.1 s. Subject 22's latency function shifted to the right and upwards at delays in the range from 0 s to 3 s, compared to GP5, to a clear peak at the 2-s delay. Subject 24's latency function shifted to the right compared to GP5, to a clear peak at the 2-s delay.

Discussion

As predicted, GP6 showed a clear peak shift in discriminative performance compared to GP5, supporting the argument made in the discussion of Experiment 3.

The mean accuracy of 0.94 at the 4.5-s delay in GP6, was the highest mean accuracy shown at any delay point in the whole study. The peak shift in remembering performance did not result merely from a fall in performance at shorter delays. It was also more than a shift to the right. Instead, the remembering function shifted upwards at longer delays (4 to 6 s), due to a change in contingencies a shorter delays (2 and 3 s) – a completely novel and revolutionary finding.

The effect on discriminative performance at the 4.5 and 5.0-s delays, of moving S– closer to S+ in GP6, was bigger than the effect of reducing probability of reinforcement from 0.1 to 0 in GP5. The generalisation gradient clearly steepened in the 3 to 4.5-s delay range in GP6, compared to the gradient in the 2 to 3.5-s delay range in GP5. This outcome was consistent with Spence's (1937) idea that a gradient of inhibition around S– was interacting with a gradient of excitation around S+ (Honig & Urcuioli, 1981).

GP6 showed evidence of negative peak shift (Guttman, 1965). Demonstrating negative peak shift has long been difficult (Grusec, 1968; Terrace, 1968). Hendry,
Switalski and Yarczower (1969), found a negative peak shift, but only by using electric shock as punishment at S−. As log \(d\), unlike rate of responding, can be less than zero, the present study could show a shift to a minimum at a point beyond S− away from S+ despite discriminative performance being very low at delays from 0 to 3 s.

Moving S− closer to S+ also meant that the training delays of 3 and 4 s were asymmetrically located in the 0 to 6 s testing range. Adaptation-level theory would suggest that if most of the testing delays are at the shorter end, then the peak shift to a weighted average position toward that region. However, the peak in discriminative performance shifted toward a longer delay in GP6. Subject 24 displayed a reverse peak shift in GP6 (Figure 7-1). Her discriminative performance both increased at shorter delays and decreased at longer delays.

Subjects 22, 24 & 25 showed latency peaks at a delay in between 0 s and S−. However, only for Subject 25 did a peak in latency coincide with minimum discriminative performance. Even there, both latency and discriminative performance fell at the 0-s delay. Thus once again, in this condition, there was no consistent relationship between latency and log \(d\).

The results for GP6 also need to be seen in the context of the progress of this study through its successive conditions, which may provide insights that are difficult to see from the results of just one condition.
General Results

To facilitate detailed presentation and discussion of the data arising from this study, it has been treated thus far as consisting of four separate experiments. However, it was actually a continual process, carried out with the same subjects, equipment, and general aims, over a period of eleven months. It is therefore appropriate now to give an integrative overview, and to draw attention to some features not so readily apparent from the separate experiment sections. Figure 8-1 presents time-series data for the mean discriminative performance over all 6 hens (Hautus adjusted log \( d \), calculated as described in the procedure for GP1), at 2, 3 and 4-s delays, for all the conditions in the study (TR1, GP1, TR2, GP2, TR3, PS1, GP3, PS2, GP4, PS3, PS4, GP5, PS6 and GP6). Figure 8-2 presents time-series data for the mean latency to choice over all 6 hens, at 2, 3 and 4-s delays, for all the conditions in the study. The individual subjects' time series data for both latency and discriminative performance were more variable than the mean data, but showed the same patterns. They are therefore not presented in this results section, but may be viewed electronically in the folder Excel Cumdata in Appendix D. Figure 9-1 presents the individual and mean data for discriminative performance (Hautus adjusted log \( d \), calculated as described in the procedure for GP1) at all test delays, for the generalisation probe conditions GP2, GP3, GP4, GP5 and GP6. Figure 9-2 presents the individual and mean data for proportion correct at all test delays, for conditions GP2, GP3, GP4, GP5 and GP6. Figure 9-3 presents the individual and mean data for latency to choice at all test delays, for conditions GP2, GP3, GP4, GP5 and GP6. Some features of the data in the above-mentioned figures are described below.

The mean log \( d \) values (Figure 8-1) during TR1 and TR2, were on average about 12% higher at the 2-s delay than at the 4-s delay, but they tended to rise and fall together. However, during PS1, mean discriminative performance at the 2-s delay both fell, and showed less, if any, tendency to covary with discriminative performance at the 4-s delay. During PS2, mean discriminative performance at the two delays tended to covary. During PS3 and PS4, mean discriminative performance at the two delays followed a cyclical opposing pattern.

Mean discriminative performance at the 2-s delay increased compared to the previous training condition, for the 5 to 6 sessions of the training condition following the generalisation probe sessions of GP2, GP3 and GP4, (namely, in TR3, PS2 and PS3). Mean discriminative performance at the 4-s delay increased compared to the previous training condition, for the 5 to 6 sessions of the training condition.
**Figure 8-1.** Log \( d \) (Hautus adjusted) at 2, 3 and 4-s delays - mean across all Hens 21-26, for all conditions.

Note. Log \( d \) was not calculated for sessions 1-21.

Note. Values of log \( d \) in generalisation probes cannot be directly compared to those in training periods, as there were only 7 trials per delay versus about 50 in training sessions.

**Figure 8-2.** Mean latency to choice at 2, 3 and 4-s delays - mean across all Hens 21-26, for all conditions.
Figure 9-1. The mean discriminability (Hautus adjusted log $d$) per session for GP6, compared to the mean of GP5, the mean of GP4, the mean of GP3 and the mean of GP2 for Hens 21-26, and the mean across all six hens. Note: for Hen 25, $\Delta_{1.0} = -0.36$. 
Figure 9-2. The mean proportion correct per session for GP6, compared to the mean of GP5, the mean of GP4, the mean of GP3 and the mean of GP2 for Hens 21-26, and the mean across all six hens.
Figure 9-3. The mean latency to choice per session for GP6, compared to the mean of GP5, the mean of GP4, the mean of GP3 and the mean of GP2 for Hens 21-26, and the mean across all six hens. Note: for Hen 25, $A_{1.0} = 30.07$. 
following the generalisation probe sessions of GP1, GP2, GP3 and GP4, (namely, in TR2, TR3, PS2 and PS3). This effect did not occur after GP5.

Mean latency to choice over all 6 subjects (Figure 8-2) was virtually the same at the 2-s and 4-s delays during TR1 and TR2. During PS1, PS2, and PS3, the mean latency at the 2-s delay was at a higher level than at the 4-s delay. However, across all these conditions, there was a strong tendency for the mean latencies at the two delays to covary. Even during PS4 some covariation was still evident. Mean latencies at the training delays (2 s and 4 s for GP1-GP5; 3 s and 4 s for GP6) during all generalisation probe sessions, were at levels very close to those in the preceding training condition.

Subject 21 consistently showed a dip, or v-shaped generalisation gradient in discriminative performance (Figure 9-1) centred at the S¯ delay in all the generalisation probes following differential reinforcement conditions (GP3-GP6). No other subject displayed such a pattern, and no subject consistently showed a generalisation gradient centered at the S+ delay.
Overall, this study provides compelling evidence of the occurrence of peak shift in remembering performance. Apart from GP2 (see Figure 2-1), the discriminative performance functions along the temporal dimension, obtained from generalisation testing, are all empirically consistent with the kind of effects shown with response rates in many stimulus discrimination and generalisation studies along other stimulus dimensions under similar contingencies. They are not consistent with any traditional theory of memory. While many detailed theoretical questions remain to be resolved, this study strongly supports in principle White's (1985, 1991) view of remembering as an act of discrimination along the stimulus dimension of time.

Of all the generalisation functions in the present study, only GP2 even remotely resembles some sort of traditional declining "memory" function. The GP1 "twin peak" function (Figure 3-1) is broadly consistent with theoretical extrapolation from the Sargisson and White (2001) "single peak" result, as shown in Figure 1(c). It remains to be seen in future studies whether the fact that the small single-point peak in GP1 is at 4.5 s, rather than at 4 s is significant. While the explanation for the functions found in GP2 and GP4 can only be speculative at present, they are not inconsistent with the discrimination interpretation in general. The curved function covering the two delays in GP4 could be explained in terms of either summation or adaptation level. The small positive area shift from GP4 to GP5 (Figure 6-1), while not a clear peak shift, was in the direction predicted.

The term "peak shift" has been used to describe results in which generalisation gradients following two-stimulus training were displaced, but without having first demonstrated a single gradient centred on S+ after training at that stimulus. However, as Winton and Beale (1971, p. 79) pointed out, "a problem with this is that it does not provide evidence of a shift in the peak from the S+. It is possible that the test stimulus responded to maximally after differential training would have been responded to maximally after non-differential training."

In the present study, subjects were initially trained at both 2 and 4 s, not at 4 s alone. Also, GP1 (see Figure 3-1) already showed a slight local peak at 4.5 s rather than at 4 s, as well as a bigger peak at 2 s. It could thus be argued that the peak performance at 4.5 s in GP3 (see Figure 3-1) was not actually a peak shift, as no peak had been previously demonstrated by the same subjects at the S+ delay of 4 s. Four arguments can be made to counter that; (a) the peak at 4.5 s in GP3 was higher than in GP1, (b) the
peak at 4.5 s in GP3 lies at the end of a rising gradient, whereas in GP1 it was only a single point standing out by a small amount, (c) by GP2 the minimum performance was at 4.5 s, and it is from this base that the large change to the GP3 peak should be seen, and (d) compared to many previous studies of remembering, which have found declining functions, a result as remarkably different as the function in GP3 can be seen as a shift of maximum discriminability to a longer delay from where it would have been expected. It would be desirable to replicate the present study, but with subjects initially trained at the S+ delay only, so that a peak might be shown there, which might then be shown to shift after addition of training at an S− delay. Nevertheless, the generalisation function in GP3 thus provides reasonable preliminary evidence peak shift in remembering. By GP6 (Figure 7-1) there was a clear peak shift, as well as an area shift, with respect to both GP4 (Figure 4-1) and GP5 (Figure 6-1). Caution is appropriate in directly comparing the absolute magnitude of the two peak shifts, due to the possible effects of training in previous conditions, but the occurrence of the effect, and its importance for theories of memory is not in doubt.

**Implications for some current theories of memory**

*Physiological trace theory*

Physiological trace theorists, guided as they are by the perseveration-consolidation hypothesis, (Müller & Pilzecker, 1900; cited in McGaugh, 2000) emphasise how performance is maintained over time, and posit an inverted wine-glass shaped short-term memory function, not consistent either with decay, or any function declining from an initial maximum, as almost universally favoured in psychological theories. One problem in relating the results of the present study to the physiological trace literature, is that the time-span over which short-term memory has been regarded as operating, has varied from seconds to hours (Gold & McGaugh, 1975). Regardless of the label for the relevant time span however, the dynamic changes in remembering functions shown in the present study, over delays from 0 to 6 s, especially the upward and rightward shift in GP6 (Figure 7-1) will need to be taken into account in any future physiological models of memory. Presumably these changes will have to be seen as new evidence of (inferred) brain plasticity (Kolb, 1998). They may also serve as a reminder of the interplay between organism and environment, and the limitations of seeking explanations of behaviour in essentially inexorable neural processes located in particular parts of the brain (Greenberg, 2002).
**Information processing theory**

Information processing theories tend to be so general and flexible, that they can be used to explain almost any outcome retrospectively (Wexler, 1978; Massaro, 1988), or be modified in some way to do so (Watkins, 1990). There seems little point therefore in discussing their relevance to peak shift in remembering. It suffices here to say that no information processing theory would have predicted the present results, but the discriminative hypothesis (White, 2002a) did.

**Interference theory of forgetting**

An intrinsic limitation of interference theories is that they can only usefully explain negative vertical variations from a theoretically maximal level (representing perfect remembering), or empirically maximal but declining reference curve (Adams, 1967). Examples of this would be an increased overall rate of forgetting or a lower initial discriminability between conditions. Under such a theory, an increase in accuracy at any delay can only be explained in terms of a decrease in a hypothetical interference inferred to have been present under the relevant ideal comparison conditions. Such an interpretation might just lead to some new insight, but would risk making the theory non-disconfirmable. In the present study, the chamber could perhaps have been made less interfering by soundproofing it and removing the ventilation fan. However, any within-trial retroactive interference (Zentall, 1973) due to the chamber environment would have been proportional to the delay, and hence if it were a major factor, should have led to a monotonically declining remembering function during generalisation testing. This clearly did not happen.

Between-trial proactive interference with learning in general, was minimised by using a fairly long ITI of 10 s (Nelson & Wasserman, 1978). The absence of reinforcement on a given trial might be seen as proactively interfering with performance on the following trial. Such interference might even build up over a series of trials within a session (Edhouse & White, 1988). However, the use of multiple generalisation test sessions, each with multiple probe series, and pseudorandom sequencing of delays within a series ensured that any such effects could not systematically influence the obtained remembering functions.

If the concept of interference were extended to apply specifically at each delay, a low rate or absence of reinforcement at certain delays might be interpreted as proactively interfering with performance on subsequent trials at those delays. Such a concept would be problematic, as it would necessarily also require a retrospective
component, given that a particular delay length could only be identified once it had ended. Performance should then have been a function of the difference between the maximum (100%) and actual reinforcement rate at each delay. Actually, performance was worse at 2 s than at 4 s in both GP3 and GP4 despite the fact that correct responses were always reinforced at both delays. Proactive interference from the contingencies of the preceding training conditions (PS1 and PS2 respectively), acting across the many trials of both generalisation test sessions, could be invoked, but that would make the concept so nebulous as to add nothing of actual explanatory value to the observed facts. Neither the rightward shift of the peaks in GP3 and GP6, nor the dynamic changes in the overall functions, can be explained by interference.

In short, interference cannot plausibly account for the selective "absence" of forgetting in the present study. The effects of contingencies of reinforcement do not exclude interference as a factor in forgetting, but do place limits on the ability of interference theories alone to explain forgetting.

Trace decay theory of forgetting

Trace theory has been used to try to account for the monotonically declining curve found in many memory studies (e.g. Wickelgren, 1973, 1974) and also specifically in DMTS remembering performance (Grant & Spetch, 1991; Roberts & Grant, 1976). It is however very difficult to see how any theory of a declining trace could explain the functions observed in the present study, with their peaks and dynamic shifts.

Up until the presentation of the colour comparison samples, a 2-s delay was indistinguishable from the first 2 s of a 4-s delay. Therefore no prospective coding (Grant & Roberts, 1973; Roberts & Grant, 1976) was possible in the present study. Some kind of retrospective coding (Roitblat, 1982) acting as a cue to forget (Grant, 1981) at the time of comparison stimulus presentation, could possibly account for a decline in performance at a particular delay, and thus for a shift downwards of the forgetting function. It cannot account for a shift upwards and to the right such as found in the present study.

Two possible devices might be tried for maintaining some kind of trace theory in the light of the present data: (a) That a trace arises at its maximum strength, but is selectively temporally inhibited by some as yet unknown mechanism at the shorter delays and held in suspended animation by a sort of hypothesised neural hot-air balloon, and carried along by a neural wind to express itself at the later time. How would this "inhibition" work? How would this timing mechanism work? Why would the function
first rise, then peak and fall, as opposed to a traditional decay curve starting at a point along the time axis? (b) that a trace starts at a lower value, but is then progressively temporally amplified along the functions shown in this study, to reach a maximum at a later time after which it then declines. How would this amplification work? What mechanism would control the amplification? How would the size of shift, height of peak and slope of gradient be controlled?

Both of the above approaches would discard the main advantage which psychological trace theory has hitherto had, namely, its superficial face-validity. They would instead formidably increase psychological trace theory's problems without resolving them. Attempts to explain remembering in terms of neural correlates will undoubtedly continue, and legitimately so within the science of physiology. Even if physiological correlates of the behavioural processes shown in the present study are eventually found however, they will be psychologically meaningful only in the context of a functional analysis of the behaviour concerned (Skinner, 1974). However, the present data constitute a clear empirical refutation of traditional psychological trace theory. The present study has achieved something which trace theory has never achieved in its long history, namely, prediction and then demonstration of a novel, lawful relationship between environmental contingencies and behaviour. It is difficult to justify any after the fact relabelling of the demonstrated behavioural functions as "traces" when a more parsimonious, productive and integrative alternative is available in the functional analysis of behaviour.

Rubin, Hinton, and Wenzel (1999) recommend that, "in order for time between presentation and testing to be unambiguous, each item should be presented only once" (p. 1162), that is, a "pure" measure of memory retention should be uninfluenced by a learning history. This is a reasonable way to limit the effect of uncontrolled variables, but it is important to remember that it excludes only the effect of the intra-study history for a specific stimulus. Actually, the subjects in any verbal remembering study have extensive generalised learning histories. It would be remarkable if their response even on any single trial were uninfluenced by that history. A critic might argue that the remembering functions in the present study arose only because the subjects were conditioned by a particular (in this case non-verbal) history. It would however then be logically inconsistent to not also attribute the "forgetting" functions seen in traditional memory studies to a conditioning history. History is not absent just because we do not intentionally manipulate it and hence ignore it. If remembering functions can be so dynamically changed by contingencies, surely that says something about the nature of
remembering which cannot be readily explained away by any attempt to defend the special status of the traditional decay curve. If anything, it would suggest a careful ecological re-examination of the nature of the contingencies that normally shape remembering.

**Discrimination and generalisation**

**Generalisation testing**

As shown in Figure 8-1, discriminative performance at both training delays increased compared to the previous training condition, for the 5 to 6 sessions of the training condition following the generalisation probe sessions of GP2, GP3 and GP4, (namely, in TR3, PS2 and PS3). This increased discriminative performance did not occur after GP1 or GP5. While the effect did not interfere with the objectives of the present study, it raises the question of whether a similar effect might also occur during subsequent generalisation testing sessions if more than two sessions were to be employed in future studies.

When discriminative performance data from the first and second sessions of GP1, GP2 and GP3 were originally examined (folder Excelcumdata in Appendix D) shortly after these conditions were run, no consistent changes in pattern from the first to the second session were noted. It was concluded from this that generalisation testing using two 40-min sessions on two separate days, did not result in any between-session, within-probe learning effect of a sufficient magnitude to undermine the procedure's validity. This conclusion was not then re-examined during the remainder of the study. However, during the writing of the present discussion, it was realised both (a) that any between-session consistencies in individual subjects' behaviour were important in themselves, and (b) that the dynamic changes shown in GP5 (Figure 6-1) and GP6 (Figure 7-1) suggested the possibility of systematic between-session changes within those conditions. It would also be interesting to see how such variations, if any, contributed to the mean across-session discriminative performance functions previously shown in Figures 3, 4, 6, and 7. Accordingly, Figure 10-1 presents the individual and mean data for discriminative performance (Hautus adjusted log \(d\), calculated as described in the procedure for GP1) at all test delays, for the first, second and third sessions of GP1, compared to the mean of all three, as previously shown. Figures 10-2 to 10-6 present the individual and mean data for discriminative performance (Hautus adjusted log \(d\), calculated as described in the procedure for GP1) at all test delays, for the first and second sessions compared to the mean of the two, as previously shown, for GP2 to GP6 respectively.
Figure 10-1. The discriminability (Hautus adjusted log $d$) per session, for the first, second and third sessions of GP1, and the mean of all three as previously shown, for Hens 21-26, and the mean across all six hens.

Note: for hen 21, $D_{5.0} = -0.42$.

Note: From Session 1, the 0-s delay data, for Hens 21, 24 and 25 were lost.
Figure 10-2. The discriminability (Hautus adjusted log $d$) per session, for the first session compared to the second session of GP2, and the mean of the two, as previously shown, for Hens 21-26, and the mean across all six hens.
Figure 10-3. The discriminability (Hautus adjusted log $d$) per session, for the first session compared to the second session of GP3, and the mean of the two, as previously shown, for Hens 21-26, and the mean across all six hens.
Figure 10-4. The discriminability (Hautus adjusted log $d$) per session, for the first session compared to the second session of GP4, and the mean of the two, as previously shown, for Hens 21-26, and the mean across all six hens.
Figure 10-5. The discriminability (Hautus adjusted log $d'$) per session, for the first session compared to the second session of GP5, and the mean of the two, as previously shown, for Hens 21-26, and the mean across all six hens.
Figure 10-6. The discriminability (Hautus adjusted log $d$) per session, for the first session compared to the second session of GP6, and the mean of the two, as previously shown, for Hens 21-26, and the mean across all six hens. Note: for Hen 25, $^{*O_{1.5}} = -0.46$, $^{*\Delta_0} = -0.61$, $^{*\Delta_{1.0}} = -0.53$. 
Figure 10-1 shows no consistent change in the discriminative performance functions within GP1 from the first session to the second, but by the third session, variability increased. Figures 10-2 to 10-6 show a series of remarkably consistent patterns from the first to the second session, not only at the mean level, but in the individual discriminative performance functions as well. Figure 10-4 shows a systematic change during GP4, toward a peak shift, for Subjects 23 and 24. Figure 10-5 shows systematic changes during GP5, toward a higher peak at S+, or deeper dip at S−, for Subjects 21, 23, and 25. Figure 10-6 shows systematic changes during GP6, toward a higher peak at S+, a deeper dip at S−, or both, for Subjects 21, 24, 25 and 26. This suggests that the generalisation testing process became less neutral as the study progressed. This is not a problem for a peak shift study, as peak shift is an inherently dynamic phenomenon. It illustrates, inter alia, the power of an individual graphical analysis to reveal patterns which a statistical analysis would have hidden.

The very fact that the generalisation testing method employed in the present study revealed discriminative performance functions very similar to those found in previous studies from post-discrimination testing of response rates, may be taken as an indirect validation. A further unplanned element of validation arose from the fact that the second session of GP2 was mistakenly run not 24 hours, but 48 hours after the first session. Despite this, the mean discriminative performance functions for the first and second sessions of GP2, shown in Figure 10-2, are very closely similar.

The mean numbers of trials per session, during GP1, GP2, GP3, GP4, GP5, and GP6 were 107, 106, 105, 106, 95, and 96 respectively. In other words, the reduction in probability of reinforcement for correct responses, from 0.9 to 0.1 at the 2-s delay, did not effect the mean level of responding during GP3 and GP4 compared to GP1 and GP2. Reducing the reinforcement probability at the 2-s delay further in GP5 and GP6, from 0.1 to 0, produced a 10% fall in the mean number of trials completed per session. This did not prevent the finding of peak shift in GP6, and at the highest mean level of accuracy in the study. Now that it has been shown that this level of reinforcement adequately maintains hens' responding during probe sessions, it would be desirable in future to have the probability of reinforcement at the training delays the same in probe sessions as during training, (e.g. in the case of the present study, 0.9 rather than 1 at the 4-s delay) or, like Sargisson and White (2001), to add extra reinforced trials at the training delays, but to have an unreinforced trial at each training delay in the generalisation test series.
Generalisation gradients of discriminative performance

Another feature of the present study, was the use of eleven different delays, mostly only 0.5 s apart, compared to 2 s or more. Having the probe delays so close together helped to define the discriminative performance function more fully than in previous studies. As Rilling (1977, p. 449) pointed out with regard to the response-rate generalisation literature, "sharp peaks in published gradients may just reflect a lack of data points in the vicinity of S+". Similarly, the peaks in the curves fitted by Sargisson and White (2001) to their data, may reflect not only the theoretical legacy of a concern with negative exponential functions, but also the methodological fact of testing at delays spaced at 2-s intervals from 0 to 10 s. If, in the present study, only the data points at the 0, 2, 4 and 6-s delays were to be considered, then GP1 and GP2 (Figure 3-1) would show a single peak and global maximum at the 2-s delay, GP4 (Figure 4-1) would show a global maximum at the 0-s delay and a single peak at the 4-s delay, while GP3 (Figure 3-1), GP5 (Figure 6-1), and GP6 (Figure 7-1) would show a single peak and global maximum at the 4-s delay. All could then be fitted with a double negative exponential curve according to Equation (2). The actual testing at 0.5-s intervals however, revealed a more complex and dynamic process. Closer spacing of test delays implies that either there must be a greater number of test points, and hence fewer trials at each test point during a session, or a further restriction of the range of the stimulus dimension tested.

Curve-fitting

The shape of the remembering function in GP3 (Figure 3-1) could perhaps be seen in terms of Sargisson and White's (2001) double negative exponential function, Equation (2), which would look roughly like Figure 1(b) with a peak at the 4.5-s delay. At least one more term would need to be added to the equation, to account statically for remembering performance at the second training delay. However, fitting such a function to the present data would not explain the location or height of the peak, or why peak shift occurred. In other words, passive after the fact curve-fitting cannot explain the dynamics of peak shift in remembering.

The shape of the functions in GP4-6 is radically different from any previously documented remembering function, and not like any combination of negative exponential or power functions (Rubin & Wenzel, 1996; Wixted & Ebbesen, 1997; Killeen, 2001). Instead, they look much more like the kind of functions found in studies of stimulus discrimination and generalisation along dimensions such as light wavelength (e.g. Blough, 1961, 1969).
Furthermore, the quantity of "percentage of variance accounted for", commonly given along with fitted curves, is logically and conceptually only meaningful if the curve does in fact represent the true underlying pattern in the data, and "errors" are randomly distributed around the curve.

In the present study, in Figures 10-1 to 10-6, the individual subjects display unique patterns, such as local peaks or dips in their behavioural functions, some of which either repeat almost perfectly across sessions, or extended sequences of which shift up, down, or sideways. This suggests that these individual patterns are real, or meaningful, even if we don't yet know what that meaning is. Averaging such data is particularly risky (Sidman, 1960). If, for example, each subject displayed a local peak and dip, but at a different place along the temporal dimension, the mean could simply be a horizontal straight line. Fitting any kind of curve, which would treat these consistent and therefore potentially discoverably lawful patterns as "errors", would merely hinder the process of discovery.

Wickens (1998) argued that the differences between a data set and a simple model which by visual inspection obviously does not fit the data, may help reveal a psychological process. However, it may also obscure one. Jenkins and Dallenbach (1924) found that sleep improved recall after 1, 2, 4 and 8 hour intervals, compared to normal activity. The actual data show slightly better remembering after 4 or 8 hours sleep than after 2 hours. However, Loftus and Loftus (1976) represent these data with two monotonically declining curves, reflecting their theoretical preconceptions, (a) about how the data ought to look, and (b) that unexpected data points represent "deviation" or "error" from this pattern, and may thus be legitimately ignored or buried by averaging. If such monotonically declining fitted curves are presented in textbooks as "facts", it becomes difficult for students, who will comprise the next generation of researchers, to look at those data differently.

Wickens (1998) went on to propose that a fitted retention function should be considered on three grounds, (a) an initial level which may be large but not infinite, (b) a final asymptote which should not be less than chance level of the measure employed (assuming that memory is a "thing" that is stored and therefore cannot decline to less than nothing), and (c) monotonicity of decline. Wickens then noted that, "Imposing this restriction does not deny the possibility that in some designs the amount of material recovered may reliably increase between certain time points. It does indicate that such hypermnesic increases are interesting because they are exceptions to the typical pattern" (p. 381). To the present author, the very occurrence of such interesting "exceptions",
whose number the present study has happily augmented, would suggest instead that the "process" has been fundamentally misunderstood.

Surely, if remembering performance can increase as a function of the research design, i.e., as a function of environmental contingencies, that suggests that the preoccupation with curve-fitting to describe a process assumed to be essentially understood, as storage and loss, is inappropriate. In a functional behavioural account, unhampered by the assumption that memory is a stored thing, a measure such as log $d$ may well be less than zero (Figure 7-1), and vary lawfully in this sub-zero region.

As Roberts and Pashler (2000) showed, even a fitted curve which coincides well with an expected outcome for a particular theory, provides very little evidence for that theory unless the aetiology of the model is valid, alternative explanations for the observed data are effectively ruled out, and the model actually predicts something not already known. In short, at this stage there seems to be no good reason to fit any kind of curve to the present data. A more promising approach would seem to be examination of the dynamics of the remembering process.

*Theories of peak shift*

Peak shift in response rates has been previously demonstrated along many different stimulus dimensions (see Table 1). The present study has extended the generality of the peak shift phenomenon both to the stimulus dimension of time, and to the response domain of discrimination, as measured by log $d$.

The opinion of Mariner and Thomas (1969, p. 764) that, "a peak shift cannot be reliably predicted from the discrimination performance immediately preceding the generalization test", has been borne out repeatedly in the present study. Trying to predict peak shift from prior performance at two delay points necessarily involves trying to guess the position of an unknown curve from only two of the points presumed to lie on that curve, when those two points can only possibly define a straight line. Mathematically, a minimum of three points are required to define a curve. Perhaps therefore, training at three time delay points might allow prediction of peak shift from prior discriminative performance at those delay points, as long as the subjects can successfully discriminate between the three points.

Three theoretical approaches to the problem are considered below.
**Spence's summation theory**

Spence (1937) proposed a theory to explain interactive effects of discriminations at different points along a stimulus continuum in terms of summation of independent gradients of inhibition and excitation. Spence's theory considers learning at each stimulus in absolute terms rather than in relation to learning at other stimuli. There is an inherent philosophical problem with Spence's account. On the one hand, Spence postulates the existence of a generalisation gradient of responding along the stimulus dimension, around a single trained stimulus, which necessarily implies that responding at points around the stimulus is dependent on, or related to, responding at the stimulus. On the other hand, responding at separate stimulus points is said to be independent, so that the gradients around different points may be vertically added or subtracted. These two propositions are, in principle, contradictory. Nevertheless, Spence's theory has survived the test of time in accounting for many results in stimulus generalisation either fully or partially.

In practice, there are three challenges for Spence's theory, (a) demonstrating the existence of a gradient of inhibition, (b) demonstrating existence of a gradient of excitation, and (c) the fact that summation can only at best account for peak shift but not for the increased height, even if both gradients exist and if they do interact (e.g. Hearst, 1968). An increased net gradient in the delay range from $S^- \rightarrow S^+$ may be accounted for, depending on the shapes of the inhibitory and excitatory gradients. As shown in Figure 9-1, from GP3 to GP5, Subject 21 produced a v-shaped function consistently centered on the $S^-$ delay of 2 s. In GP6, this v-function then shifted precisely to the new $S^-$ delay at 3 s. This could be interpreted as a gradient of inhibition around $S^-$, and hence as evidence for Spence's theory. However, none of the other subjects showed such a consistently clear effect, and no comparable gradient of excitation was apparent.

From GP5 to GP6, the mean gradient of the function between $S^-$ and $S^+$ steepened, and the height of the peak increased. While consistent with the results of many previous peak shift studies, this effect is contrary to Spence's summation account. Spence's theory predicts not only transposition (shift) of the maximum performance to a point beyond $S^+$ away from $S^-$, but also a shift of the minimum to a point beyond $S^-$ away from $S^+$. Thus summation predicts a function with a single peak on one side and a single dip on the other side.

In some studies, (e.g. Hanson, 1959) the testing range did not allow such a negative peak shift to be shown. Even if the testing range is adequate, there is a difficulty in that even in reinforcement-maintained generalisation testing, the test delays
are effectively all S−'s, so that responding at and near the intended S− may be too low
to show a clear negative peak. Guttman (1965) addressed this problem by starting from
a positive rate baseline, and found a negative peak shift, supporting Spence. As
discriminative performance, unlike rate of responding, can be less than zero, the present
study can show a shift to a minimum at a point beyond S− away from S+ despite
discriminative performance being very low in the short-delay region.

In GP5 there was a pattern of three points (at 1.0, 1.5 and 2.0 s) trending down to
a local minimum at S−, followed by three more points trending up to the peak. This
pattern was maintained, but shifted upwards and to the right in GP6. In both cases there
was thus a local minimum at S−, but at shorter delays outside this pattern, i.e. at 0 s in
GP5, and 0, 1.0 and 1.5 s in GP6, performance fell to an even lower global minimum.
This pattern of multiple peaks and dips arising from training at just two delays, is a
remarkable relational effect which cannot be explained by summation of simple
generalisation gradients around those delays.

Adaptation-level theory

Adaptation-level (Helson, 1947, 1959), is a kind of perceived average value
within a test series of stimuli, toward which a subject's responding shifts. Many studies
have examined the possible effects of adaptation-level on peak shift (e.g. Thomas,
Svinicki, & Vogt, 1973; Thomas, 1974, 1993; Thomas, Mood, Morrison, & Wiertelak,
1991). However, the present study was designed so that the testing series of delays,
from 0 s to 6 s was symmetrically arranged around the 2-s and 4-s training delays.
Assuming linear timing therefore, no peak shift due to adaptation-level would have been
expected. The arithmetic mean length of delay interval, in the middle of both the 2 and
4-s training delays, and of the 0-6-s test delay series, was at 3 s. It is possible that the
broad curve centered on the 3-s delay in GP4 (Figure 4-1), was a consequence of an
adaptation-level developing at that delay. However, other explanations, such as
summation of flattening generalisation gradients, are not ruled out. In both the cases in
which a clear peak shift occurred, GP3 and GP6, the shift was away from the 3-s delay,
beyond the 4-s delay, in other words, in the opposite direction from what would have
been expected from adaptation-level theory. Indeed, in GP6, because the S− delay was
moved from 2 s to 3 s, the testing delay series was no longer symmetrical about the
training delays, and the argument against adaptation-level, accordingly stronger.
Although in GP6 (Figure 7-1), Subject 24 displayed a reverse peak shift toward a
shorter delay, the peak was not at the 3-s delay (as would have been predicted by
adaptation-level theory), but at the 1.5-s delay. In short, to the limited extent that it could, the present study did not provide support for adaptation-level theory.

**Emotional responses to extinction**

Terrace (1963) suggested that an "emotional reaction" to extinction for errors contributes to peak shift. Terrace (1964) in fact claimed that experience with extinction at S− was necessary to produce peak shift. In the present study (PS1-PS5), responses at the S− delay, would have been equivalent to Terrace's "errors". As a trial remained in progress until a choice was made, the subjects had to respond in an S− trial so as to have the opportunity to have more S+ trials. In short, "errors" were unavoidable. Terrace (1968) claimed that behavioral contrast during discrimination training is a necessary condition for peak shift, "and that a sufficient condition for the occurrence of both phenomena is a reduction in the rate of responding to one of two alternating discriminative stimuli" (p. 737). The present study is not a test of errorless discrimination training, as the subjects did make errors. However, functions of latency to choice, both from training conditions and generalisation test conditions in the present study, were examined as possible indicators of "emotional" responses.

**Latency to choice**

The generalisation probe conditions revealed orderly contingency-dependent generalisation gradients of latency along the temporal dimension (Farmer, Schoenfeld, & Harris, 1966; Moody, Stebbins, & Iglauer, 1971). However, these gradients were not consistently related to discriminative performance (cf. Green, Terman, & Terman, 1979). During PS4, when extinction was introduced at S−, latencies to choice declined slightly on S+ trials, but increased dramatically on S− trials. This could be described as latency contrast, comparable to behavioral contrast in response rates (Bloomfield, 1969).

In Terrace's (1966, p. 617) terms these could be regarded as "emotional responses", with long latencies presumably indicating greater "frustration". The high latency level at S− was maintained during the following generalisation probe, GP5, but only a small area shift, and no peak shift occurred. This does not support extension of Terrace's (1968) claim that behavioral contrast is a necessary and sufficient condition for peak shift, to the realm of remembering. The mean generalisation gradient for latencies showed a peak at 1 s, with a steady decline to the original baseline latency by the 4-s delay. This could be seen as a negative peak shift for latency. However, this
function did not coincide with the more complex discriminability gradient, which had a local peak at 1s, where the global minimum should have been if discriminability peak shift were caused by the maximum "emotional" latency response at 1 s. Latencies continued to be very long during PS6, when S− was moved from 2 s to 3 s.

In GP6 the mean latency generalisation gradient was virtually identical to that in GP5, which would imply no change in the distribution of "emotionality" to account for the peak shift which occurred in discriminability. The individual latency gradients in GP6 were more variable. Subject 21 (the only one to produce a consistent, large "gradient of inhibition" around S−) showed a sharp fall in latencies at shorter delays in GP6, as did 23. Subject 26's latencies stayed almost the same. Both 22 and 24 showed a rightward shift in their latency gradients, peaking at the 2-s delay. However, they displayed opposite trends in discriminability peak shift. Only Subject 25 showed the pattern which would be expected under Terrace's interpretation, namely a large increase in latency to a peak coinciding with the point of minimum discriminability.

Overall, based on the use of latency functions as indicators of emotionality, the present study does not support the idea that an "emotional reaction" is a necessary condition for, or systematic predictor of peak shift in remembering.

Implications for Direct Remembering Theory

Overall, discriminative performance at the two delays in training conditions was broadly related to temporal distance from time zero and probability of reinforcement for correct responses at the particular delay, as would be predicted by White's (1985, 1991) Direct Remembering (DR) theory. Replication of the Smith (1997) and Sargisson and White (2001) studies which showed peaks in remembering performance at delays where remembering is reinforced would be desirable. Also, the external validity of the peak shift phenomenon could be enhanced by replicating the present study in its original form, namely with probability of reinforcement 0.1 at S− and 0.9 at S+. Nevertheless, the key concepts of discrimination along the temporal dimension, and generalisation of remembering performance (Sargisson & White, 2001) have both been strongly supported by the present study. However, the present study, while inspired by DR theory, has produced several results which cannot be accounted for by its current quantitative formulation.

Independence of performance at different delays
A central part of DR theory has been the idea that performance at a particular delay will be affected by the contingencies applying at that delay, independently of contingencies at other delays (White, 1991; White & Cooney, 1996; White & Wixted, 1999). Consistency with this independence was one reason for preferring an exponential to a power function to describe remembering performance.

The idea of independence was implicitly modified by Sargisson and White's (2001) finding of generalisation around a peak, as reinforcement at one delay had an effect on performance at delays on either side of it. Thus performance at the surrounding delays was dependent on the consequences of performance at the delay point where reinforcement occurred.

The present demonstration of peak shift has now added a further dimension of dependence, as maximum discriminative performance occurred at delays at which no reinforcement was ever given. Furthermore, in the present study, major shifts occurred in the remembering functions around the two training delays. For example, from GP3 to GP4 (see Figure 4-1) \( \log d \) at 2 and 4 s remained virtually the same, but the shape of the function changed markedly. In GP6 (see Figure 9-6), performance all the way from 0 to 6 s was affected by contingencies at 3 and 4s.

The time-series training data (Figure 8-1) also show elements of dependence in the form of positive and negative induction of discriminative performance. For example, in TR1 and TR2, the \( \log d \) values were high at both the 2 and 4-s delays, but about 12% higher at 2 s than at 4 s, as would be predicted by DR theory. Then, during PS2, when the probability of reinforcement at the 2-s delay was reduced from 0.9 to 0.1, discriminative performance fell at the 2-s delay, as expected. However, it also fell at the 4-s delay (positive induction), contrary to the expectation from DR theory that performance would remain the same, as the probability of reinforcement on previous trials at that delay, remained the same (White & Wixted, 1999).

During PS3, discriminative performance compared to the previous condition was higher at the 4-s delay while remaining the same at the 2-s delay (negative induction, or contrast), despite the fact that the contingencies at the two delays remained the same. Indeed, the probability of reinforcement at the 4-s delay was constant at 0.9, across all training conditions. Since the mean number of trials per session over the training conditions from PS1 to PS6 varied by no more than 3%, the absolute level of reinforcement at the 4-s delay was therefore almost constant as well. DR theory would thus predict a nearly constant level of discriminative performance over those conditions at the 4-s delay (White & Wixted, 1999). However, Figure 8-1 shows that the
discriminative performance at the 4-s delay varied by up to 33% between conditions from PS1 to PS6. This variation could not be attributed to the intervening slightly different contingency (probability of reinforcement 1.0) during preceding test conditions, as (a) the direction of shift was not constant after test conditions, and (b) two large changes in discriminative performance, from PS3 to PS4, and from PS5 to PS6 followed not a test condition, but only a change in contingencies at the other training delay. The absolute reinforcement rate can still be regarded as the main determinant of this ongoing discriminative performance, but there is also a substantial relational component.

It is therefore clear that the radical degree of independence of performance at different delays envisaged by White (1991) is not tenable. It was however, a theoretically very productive idea, and remains important. For example, the result for GP5 (Figure 6-1), in which lowering the probability of reinforcement at the 2-s delay to 0 resulted in a steeper generalisation gradient between the 2-s and 4-s delays, but little effect at or beyond the 4-s delay, suggests that a large degree of independence of performance can be shaped at two different delays.

The concept of the independence of initial discriminability, log $d_0$, from the rate of forgetting, $b$ (White, 1985), is still supported, as any given $b$ is consistent with varying degrees of dependence or independence of performance at its contributory delay points.

Figure 1(c) presented a hypothetical "twin peak" result of training with rich reinforcement at two delays. This idea received some support from GP1 (Figure 3-1), but not from GP2. Figure 1(d) presented a hypothetical result of training with rich reinforcement at one delay and poor reinforcement at another delay. This idea received strong support from GP3 (Figure 3-1) and GP6 (Figure 7-1), and in respect of the dip in the region of $S^-$ and the peak in the region of $S^+$, but without a peak shift, from GP4 (Figure 4-1) and GP5 (Figure 6-1). Only the hypothesised shape of the peak, a double negative exponential decline, was not well supported. The relatively flatter tops of the functions imply a greater degree of interdependence at delays very close to $S^+$.

Although no well-defined psychophysical scale exists for remembering performance, the fact that the dips in the functions were sharper than the tops cannot be explained in terms of a just-noticeable difference. How or why such a stimulus class (Skinner, 1938/1991) or an effective stimulus (Guttman, 1963) is formed around $S^+$ remains to be explained.
Questions for future research to explore, are at how many different delays, how far apart, after how much training, and at what reinforcement ratios, differential remembering performance can be produced for a given species. For example, does 2-delay training result in 2 peaks as in Figure 1(c), or a curve as in GP4 (Figure 4-1)? If there are two peaks, do they occur at the training delays, or is one displaced as is the small peak at the 4.5-s delay in GP1 (Figure 3-1)? If 2-delay training results in 2 peaks, does 3-delay training result in 3 peaks? Further research could also examine how many delay points can support separate peaks before these peaks merge into the declining curve seen from mixed-delay training. In the light of such data, it may turn out that psychological distance is not defined absolutely by temporal distance, as assumed by Sargisson and White (2001), after Shepard (1987), but interactively with the number of trained delays and reinforcement ratios.

**Diffusion and temporal discrimination**

White (2002b) explained forgetting as a diffusion process acting over time. That is, discrimination can be described in terms of distributions whose variance increases with increasing temporal distance from presentation of the stimulus (White & Wixted, 1999). This accounts well for traditional declining curves, and also for Sargisson and White's (2001) peak at a delay where correct choices were reinforced. It does not however, explain why maximum discriminative performance should occur at a delay at which choices were never reinforced. It also does not explain the various changes in discriminative performance at delays around the training delays, which were found in the present study. Perhaps however, White's (2002b) diffusion account could be extended to account for peak shift in terms of discrimination along two separate dimensions, time and, in the present case, colour, by the following reasoning.

If a creature remembered a particular stimulus equally well at all points in time for the rest of its life, that is, if it displayed a forgetting rate of zero, the remembering function would be a horizontal straight line. The creature would therefore not be discriminating between temporal intervals with respect to its response to that particular stimulus. It follows that some timing ability is a logically necessary condition for differential, and thus for declining or variable, remembering. The description given earlier of remembering as, "discrimination along the dimension of time" can be modified to, "remembering is discrimination along the stimulus dimension of time, of a stimulus other than time, in the present case colour".
As the discriminations along these two different dimensions are broadly concurrent, it is tempting to treat them both as independent and as truly simultaneous. As a matter of logical rigour however, it should be remembered that they may be lagged, with one primary and the other secondary, or that they may alternate in some way. In molecular experimental design terms, the difference is that at the beginning of each trial a colour sample is presented, and after a delay the subject can choose a matching colour sample, correctly or incorrectly. In contrast, no time sample is presented to which the subject can match a later time sample: There is no "correct" time or "incorrect" time within a sample. It is important to remember however, that the experimenter's definitions of "correct" and "incorrect" are purely artifactual. In an absolute sense, the choice of a red key following a red sample is no more "correct" than choice of a green key.

Furthermore, as Spetch and Rusak (1992) pointed out, time varies continuously, and the length of an interval can only be determined when it has ended. A colour on the other hand is present immediately and throughout the stimulus presentation. These distinctions however were developed in the context of studies in which time intervals were considered as secondary stimuli (i.e., some non-time stimulus was presented for a certain length of time), presented initially, and then after an interval presented again. In the present study, it could be said that the passage of a 2-s or 4-s delay could only be determined once the two or four seconds respectively had passed. However, up until the presentation of the colour comparison samples, a 2-s delay was always indistinguishable from the first 2 s of a 4-s delay. Given that there were only two delays, after the passage of the shorter, it could be determined according to some increasing probability function, that the longer was in effect, even before it had ended. From a molar perspective, over a series of trials, the subject's response to a key the same colour as that most recently seen was reinforced: in TR1-PS3 more so after a longer delay than after a shorter delay, in PS4-GP6 only after the longer delay.

It remains to be seen to what extent under what contingencies these time/non-time discriminations may be independent or interactive. It is possible that peak shift in remembering may be explained purely in terms of a peak shift in timing. Alternatively, the time and colour discriminations may be additive or multiplicative in some way.

If time is to be treated as a separately discriminated dimension in remembering, then consideration will need to be given to existing theories of timing, such as Scalar Expectancy theory (Gibbon, 1977), the Behavioral Theory of timing (Killeen & Fetterman, 1988), Learning to time (Machado, 1997), or Multiple-time-scale (MTS)
The occurrence of peak shift in remembering may require re-evaluation of existing timing theories, or development of new ones. For example, the present study provides dynamic evidence which clearly refutes the sweeping implicit assertion by Staddon, Chelaru and Higa (2002, p. 74), that "the empirical forgetting function" is necessarily the traditional monotonically declining one. MTS theory assumes the existence of such monotonically declining functions, elevated to the level of "traces", as one of its core constructs, and therefore cannot explain the present results. Staddon and Higa (1999b, p. 478) argue that, "the choose-short and choose-long effects are the strongest evidence in favour of a trace theory of time discrimination". It is possible that the choose-short effect, a tendency for subjects choose samples of shorter duration more accurately than samples of longer duration, over longer delays (Spetch & Wilkie, 1983), is indeed due to differential forgetting of two stimuli describable by monotonically declining functions. However, in the light of the present results, that would need to be shown empirically by means of generalisation probing. A symbolic DMTS procedure, with one symbol associated with a short stimulus and another symbol associated with a long stimulus could be used to track both in the same experiment. Such forgetting functions, if demonstrated, would then still need to be explained in terms of the contingencies in the study, not merely asserted axiomatically as "traces". As Skinner (1966, p. 216) said, "the experimental analysis of behavior dispenses with theories of that sort by proceeding to find out".

Future research could explore the effects on peak shift, of varying the distance between the two delay points by linearly spaced intervals and exponentially spaced intervals. Also, as discrimination along the temporal dimension has an absolute zero point at the 0-s delay, the effects of relative versus absolute stimulus relationships could be examined. For example, the distance from origin of both delays might be varied, with either a linearly or exponentially constant gap between them. Kim's (2002) finding that some hens partitioned time intervals arithmetically while others did so geometrically, is a reminder that there may be no one function or even dynamic displayed by all subjects.

One limitation of a diffusion approach is that a diffusion process would normally be expected to operate from some positive level down toward zero. This could not account for systematic patterns of performance at worse than chance levels (and hence log $d$ values below zero), such as that displayed by Subject 25 in GP6 (see Figure 9-6).
Even if such behaviour occurs only very locally, it has important theoretical implications which should not be glossed over by averaging or ignoring it. Another problem is that diffusion alone would be expected to operate around time delays where correct responding had been reinforced, and thus, like Spence's summation account, be unable to explain increased peak heights or some steepened generalisation gradients.

Whatever the specific processes may be, it may perhaps be helpful to visualise at least in general terms the possible dimensional relationships. Figure 11 indicates one possible version of the relationships discussed above, between remembering strength, time, and colour.

*Figure 11. A possible conceptual framework for future DMTS remembering studies.*

Previous applications of the concept of diffusion to remembering (Kinchla & Smyzer, 1967; Sargisson & White, 2001), have assumed a monotonic, linear increase in the variance of the discriminable distributions with increasing distance along the stimulus dimension, in this case, time. Thus the temporal dimension has played a passive role. In the framework shown in Figure 11 however, the temporal dimension plays an active role. Contingencies may change at any time, at various points along the dimension, thus changing the discriminative process, and outcomes. This framework can accommodate the possibility of either additive, or multiplicative interactions between the dimensions of time and colour, discussed above, and allow these to be readily visualised. The broad model does not dictate any one mechanism for controlling discriminative performance along the time or colour axes. This allows researchers the freedom to explore alternatives, while directing attention to the dynamic nature of the discriminative process.
Conclusion

In conclusion, the present study has demonstrated the occurrence of peak shift along the dimension of time, in a remembering process. This is a novel, counterintuitive finding, which would not have been predicted by any traditional theory of memory. It arose from and builds on a body of work treating remembering as an act of discrimination along the stimulus dimension of time, which might in principle be subject to contingencies in a way similar to discriminations which have been studied along many other stimulus dimensions. These results evidence the scientific fruitfulness of exploring remembering as discriminative operant behaviour along the stimulus dimension of time.
References


R. T. Davis (Eds.), *Processes of animal memory* (pp. 79-112). Hillsdale, NJ: Erlbaum.


### Appendix A

*Summary of Conditions in the study*

<table>
<thead>
<tr>
<th>Condition</th>
<th>Delay/Reinforcement probability</th>
<th>Dates</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experiment 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training condition no.1 (TR1)</td>
<td>2 s / 0.9 4 s / 0.9</td>
<td>23/5/01-25/7/01</td>
</tr>
<tr>
<td>Generalisation probe no.1 (GP1)</td>
<td>2 s / 1 4 s / 1</td>
<td>27/7/01-29/7/01</td>
</tr>
<tr>
<td>Training condition no.2 (TR2)</td>
<td>2 s / 0.9 4 s / 0.9</td>
<td>2/8/01-11/10/01</td>
</tr>
<tr>
<td>Generalisation probe no.2 (GP2)</td>
<td>2 s / 1 4 s / 1</td>
<td>12/10/01-14/10/01</td>
</tr>
<tr>
<td>Training condition no.3 (TR3)</td>
<td>2 s / 0.9 4 s / 0.9</td>
<td>15/10/01-19/10/01</td>
</tr>
<tr>
<td>Peak shift training no.1 (PS1)</td>
<td>2 s / 0.1 4 s / 0.9</td>
<td>20/10/01-14/12/01</td>
</tr>
<tr>
<td>Generalisation probe no.3 (GP3)</td>
<td>2 s / 1 4 s / 1</td>
<td>15/12/01-16/12/01</td>
</tr>
<tr>
<td><strong>Experiment 2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peak shift training no.2 (PS2)</td>
<td>2 s / 0.1 4 s / 0.9</td>
<td>17/12/01-1/2/02</td>
</tr>
<tr>
<td>Generalisation probe no.4 (GP4)</td>
<td>2 s / 1 4 s / 1</td>
<td>2/2/02-3/2/02</td>
</tr>
<tr>
<td><strong>Experiment 3</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peak shift training no.3 (PS3)</td>
<td>2 s / 0.1 4 s / 0.9</td>
<td>4/2/02-23/2/02</td>
</tr>
<tr>
<td>Peak shift training no.4 (PS4)</td>
<td>2 s / 0 4 s / 0.9</td>
<td>24/2/02-27/3/02</td>
</tr>
<tr>
<td>Generalisation probe no.5 (GP5)</td>
<td>4 s / 1</td>
<td>28/3/02-29/3/02</td>
</tr>
<tr>
<td>Peak shift training no.5 (PS5)</td>
<td>2 s / 0 4 s / 0.9</td>
<td>30/3/02-3/4/02</td>
</tr>
<tr>
<td><strong>Experiment 4</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peak shift training no.6 (PS6)</td>
<td>3 s / 0 4 s / 0.9</td>
<td>7/4/02-13/4/02</td>
</tr>
<tr>
<td>Generalisation probe no.6 (GP6)</td>
<td>4 s / 1</td>
<td>14/4/02-15/4/02</td>
</tr>
</tbody>
</table>
Appendix B

The cumulative data from training sessions

These data are reproduced here in print to ensure archival integrity. For convenience of re-analysis however, they are also included on the attached CD-R disc. The raw number arrays, grouped by condition, are in the folder "CumulativeData". The same information, in larger font, combined with the probe data, grouped by subject, and with column headings on each page for easy visual inspection, may be found in the sub-folder Word.dat.

Key to headings for the training data. One horizontal line shows the cumulative data for one daily session. The column numbers in the data array, (e0 - e30), are those assigned by the MEDSTATE NOTATION ™ programs MTSPEAK.MPC (for TR1-3 and PS2-5) and MTSPEAK3.MPC (for PS6). Abbreviated descriptive column headings are also shown as follows:
No. = Subject number
Date = Date in year/day/month format
(e0) LC = Correct responses on the Left key 2 s(3 s in PS6)
(e1) LI = Incorrect responses on the Left key 2 s(3 s in PS6)
(e2) RC = Correct responses on the Right key 2 s(3 s in PS6)
(e3) RI = Incorrect responses on the Right key 2 s(3 s in PS6)
(e4) LRFT = Number of Reinforcements obtained from Left key responses 2 s(3 s in PS6)
(e5) RRFT = Number of Reinforcements obtained from Right key responses 2 s(3 s in PS6)
(e6) LLC = Total latency to Left Correct choices in tenths of a second 2 s(3 s in PS6)
(e7) LLI = Total latency to Left Incorrect choices in tenths of a second 2 s(3 s in PS6)
(e8) LRC = Total latency to Right Correct in tenths of a second 2 s(3 s in PS6)
(e9) LRI = Total latency to Right Incorrect choices in tenths of a second 2 s(3 s in PS6)
(e10) LC = Correct responses on the Left key 4 s
(e11) LI = Incorrect responses on the Left key 4 s
(e12) RC = Correct responses on the Right key 4 s
(e13) RI = Incorrect responses on the Right key 4 s
(e14) LRFT = Number of Reinforcements obtained from Left key responses 4 s
(e15) RRFT = Number of Reinforcements obtained from Right key responses 4 s
(e16) LLC = Total latency to Left Correct choices in tenths of a second 4 s
(e17) LLI = Total latency to Left Incorrect choices in tenths of a second 4 s
(e18) LRC = Total latency to Right Correct in tenths of a second 4 s
(e19) LRI = Total latency to Right Incorrect choices in tenths of a second 4 s
(e20) PC2 = Percentage of Correct responses after 2-s delay
(e21) PC4 = Percentage of Correct responses after 4-s delay
(e22) TIME = Total length of session in seconds
(e23) 2RC = Number of Correct choices of a Red Key after a 2-s delay
(e24) 2RI = Number of Incorrect choices of a Red Key after a 2-s delay
(e25) 2GC = Number of Correct choices of a Green Key after a 2-s delay
(e26) 2GI = Number of Incorrect choices of a Green Key after a 2-s delay
(e27) 4RC = Number of Correct choices of a Red Key after a 4-s delay
(e28) 4RI = Number of Incorrect choices of a Red Key after a 4-s delay
(e29) 4GC = Number of Correct choices of a Green Key after a 4-s delay
(e30) 4GI = Number of Incorrect choices of a Green Key after a 4-s delay
Appendix C

The cumulative data from generalisation probe sessions

These data are reproduced here in print to ensure archival integrity. For convenience of re-analysis however, they are also included on the attached CD-R disc. The raw number arrays, grouped by condition, are in the folder "CumulativeData". The same information, combined with the training data, grouped by subject, and with column headings on each page for easy visual inspection, may be found in the sub-folder Word.dat.

Key to headings for the generalisation probe data. Data for each of the eleven probe delays are on a separate line. Thus the data for one daily session occupy eleven lines. The column numbers in the data array, (e0 - e16), are those assigned by the MEDSTATE NOTATION™ programs MTSPROBE.MPC (for GP1-GP5) and MTSPROB3.MPC (for GP6). Abbreviated descriptive column headings are also shown as follows:

<table>
<thead>
<tr>
<th>No.</th>
<th>Date</th>
<th>e0 SEC</th>
<th>e1 LC</th>
<th>e2 LI</th>
<th>e3 RC</th>
<th>e4 RI</th>
<th>e5 LRFT</th>
<th>e6 RRFT</th>
<th>e7 LLC</th>
<th>e8 LLI</th>
<th>e9 LRC</th>
<th>e10 LRI</th>
<th>e11 RC</th>
<th>e12 RI</th>
<th>e13 GC</th>
<th>e14 GI</th>
<th>e15 PC</th>
<th>e16 TIME</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Delay in tenths of a second (thus 10 = 1 s, 15 = 1.5 s ....60 = 6 s)</td>
<td>Correct responses on the Left key</td>
<td>Incorrect responses on the Left key</td>
<td>Correct responses on the Right key</td>
<td>Incorrect responses on the Right key</td>
<td>Number of Reinforcements obtained from Left key responses</td>
<td>Number of Reinforcements obtained from Right key responses</td>
<td>Total latency to Left Correct choices in tenths of a second</td>
<td>Total latency to Left Incorrect choices in tenths of a second</td>
<td>Total latency to Right Correct in tenths of a second</td>
<td>Total latency to Right Incorrect choices in tenths of a second</td>
<td>Number of Correct choices of a Red Key</td>
<td>Number of Incorrect choices of a Red Key</td>
<td>Number of Correct choices of a Green Key</td>
<td>Number of Incorrect choices of a Green Key</td>
<td>Percentage of Correct responses</td>
<td>Total length of session in seconds</td>
</tr>
</tbody>
</table>