The Role of Instructions and Intention in Learning

Jessica C. Lee

School of Psychology
Faculty of Science
University of Sydney

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

2016
Abstract

This thesis investigates how manipulating intention to learn (learning orientation) through verbal instructions affects learning in a range of putatively associative and implicit tasks. Within three different paradigms, learning orientation was manipulated so that learning was either incidental to, or aligned with (i.e. intentional) the aims of the task. The first series of experiments investigated sequence learning, as measured in the serial reaction time task. Sequence learning was found to result reliably under incidental conditions and was selectively improved by instructions promoting discovery of a relational rule describing a set of probabilistic contingencies. The second series of experiments used the prototype distortion task, where it has been claimed that implicit learning of a category of prototype-centered stimuli can occur automatically as a result of exposure. Using a visual search task as a means of incidental exposure, equivocal evidence for the implicit status of learning in the prototype distortion task was found, and instructions directing participants to memorize the stimuli resulted in greater evidence of learning the similarity structure of the category. Finally, the third series of experiments assessed generalization along stimulus dimensions following a difficult discrimination task. Instructions directing attention to a particular stimulus dimension promoted rule-based generalization and facilitated a dissociation in the pattern of generalization obtained as a result of reducing rule applicability on test. The results suggest that human learning is highly susceptible to learning orientation, which has implications for the way implicit learning should be viewed as a psychological construct. Theories of learning, whether single- or dual-process, need to better account for this seemingly pervasive role of learning orientation.
Acknowledgements

First and foremost, I would like to thank my supervisor, Evan Livesey, for his encouragement, patience and guidance throughout my candidature. Your insight and enthusiasm for research is inspiring, and I am fortunate to have had such a dedicated and brilliant mentor.

I am extremely lucky to have been surrounded by a group of bright and fun students and researchers throughout my PhD. They have supported or inspired me in some form through discussing their own research, providing me with feedback on mine, or allowing me to work on theirs (thanks Ben – even though I got fired regularly!). To the 520 (and beyond) crew, I have enjoyed every moment including Bayes Club, Muscle Time, Friday Thaiday, and living with the best housemates, Nicky and Hilary. Thank you especially to Dot for her intellectual and emotional support from the very beginning and through all the highs and lows.

I would like to acknowledge the input of several anonymous reviewers of my work, as well as Ian McLaren. Their feedback has allowed me to clarify and refine my ideas in my written work. I would also like to thank Andy Wills for allowing me to visit his lab in Plymouth during my candidature.

This thesis would not have been possible without support from my family. To my sister, Amanda, thank you for your assistance with proofreading and baking me cakes at all the right moments. To my parents, I am deeply grateful for the sacrifices you have made to afford me the opportunities I have enjoyed. Thank you to my Mum especially, who instilled in me the value of hard work, and for reminding me that education is a privilege she never had. Lastly, thank you to Tung for everything. You are the worst study buddy, but the best partner.
Statement of Originality

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Jessica Lee
Table of Contents

Abstract ................................................................................................................................. i
Acknowledgements .......................................................................................................... ii
Originality Statement ..................................................................................................... iii
Table of Contents ........................................................................................................ iv
List of Figures ................................................................................................................... xiii
List of Tables .................................................................................................................... x
List of Publications .......................................................................................................... xi
List of Conference Presentations ..................................................................................... xii

CHAPTER 1: General Introduction .................................................................................. 1
  1.1 The Nature of Associative Learning ......................................................................... 2
  1.2 Learning in Humans ................................................................................................. 6
  1.3 Dual-Process Theories ............................................................................................. 8
    1.3.1 Associative and Propositional Learning ......................................................... 9
    1.3.2 Implicit and Explicit Learning ....................................................................... 16
    1.3.3 Rules and Associations ............................................................................... 22
  1.4 Some Commonalities ............................................................................................... 24
    1.4.1 Learning without Awareness? ...................................................................... 24
    1.4.2 Learning without Thought? .......................................................................... 25
    1.4.3 Theoretical Questions .................................................................................. 26
  1.5 General Aims ........................................................................................................... 27
  1.6 Research Questions ................................................................................................. 29
  1.7 Outline of Chapters ................................................................................................. 30

CHAPTER 2: Sequence Learning ..................................................................................... 34
  2.1 Introduction ............................................................................................................. 34
  2.2 Experiment 1 .......................................................................................................... 46
    2.2.1 Method ........................................................................................................... 50
      2.2.1.1 Participants and Apparatus ................................................................... 50
      2.2.1.2 Procedure ........................................................................................... 51
    2.2.2 Results and Discussion .................................................................................. 52
  2.3 Experiment 2 .......................................................................................................... 57
    2.3.1 Method ........................................................................................................... 57
      2.3.1.1 Participants .......................................................................................... 57
      2.3.1.2 Procedure .......................................................................................... 58
    2.3.2 Results ........................................................................................................... 58
    2.3.3 Discussion ...................................................................................................... 60
  2.4 Experiment 3 .......................................................................................................... 61
    2.4.1 Method ........................................................................................................... 64
      2.4.1.1 Participants .......................................................................................... 64
CHAPTER 4: Post-Discrimination Generalization

4.1 Introduction ........................................................................... 151

4.2 Experiment 1 ........................................................................ 168
  4.2.1 Method ............................................................................ 170
    4.2.1.1 Participants ................................................................. 170
    4.2.1.2 Apparatus ................................................................... 170
    4.2.1.3 Stimuli ........................................................................ 171
    4.2.1.4 Procedure .................................................................... 173
  4.2.2 Results and Discussion ..................................................... 177
    4.2.2.1 Data Analysis ............................................................... 177
    4.2.2.2 Exclusion Criteria ....................................................... 178
    4.2.2.3 Questionnaire ............................................................. 178
    4.2.2.4 Training ............................................................... 180
    4.2.2.5 Category Judgements ................................................ 180
    4.2.2.6 Typicality Ratings ...................................................... 185
    4.2.2.7 Summary .................................................................... 185

4.3 Experiment 2 ........................................................................ 187
  4.3.1 Method ............................................................................ 189
    4.3.1.1 Participants ................................................................. 189
    4.3.1.2 Procedure .................................................................... 189
  4.3.2 Results and Discussion ..................................................... 190
    4.3.2.1 Exclusion Criteria ....................................................... 190
    4.3.2.2 Questionnaire ............................................................. 191
    4.3.2.3 Training ............................................................... 193
    4.3.2.4 Category Judgements ................................................ 194

3.3.1.2 Procedure ..................................................................... 123

3.3.2 Results and Discussion ..................................................... 125
  3.3.2.1 Signal Detection Analysis ............................................. 127
  3.3.2.2 Comparison to Experiment 1 ......................................... 128

3.4 Experiment 3 ........................................................................ 131
  3.4.1 Method ............................................................................ 132
    3.4.1.1 Participants ................................................................. 132
    3.4.1.2 Procedure .................................................................... 132
  3.4.2 Results and Discussion ..................................................... 133
    3.4.2.1 Memorize vs. Search .................................................. 133
    3.4.2.2 Search vs. Search-Terminate ....................................... 134
    3.4.2.3 Signal Detection Analysis ............................................. 135
    3.4.2.4 Comparison to Experiment 1 ......................................... 136

3.5 General Discussion .............................................................. 138
  3.5.1 Learning during Visual Search ......................................... 140
  3.5.2 The Effect of Encoding Conditions ................................... 144
  3.5.3 The Dissociation between Categorization and Recognition ... 147
  3.5.4 Conclusion ....................................................................... 149

3.2.1 Discrimination Generalization .......................................... 116

3.1.3.1 Procedure .................................................................... 91
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Schematic diagram of the three-choice serial reaction time (SRT) task</td>
<td>45</td>
</tr>
<tr>
<td>2.2</td>
<td>Coding scheme for subsequences</td>
<td>50</td>
</tr>
<tr>
<td>2.3</td>
<td>Sequential effects (RTs and errors) for each fourth-order subsequence in Experiment 1</td>
<td>54</td>
</tr>
<tr>
<td>2.4</td>
<td>Cueing effect (RT) for Hint and No Hint groups in Experiment 2</td>
<td>59</td>
</tr>
<tr>
<td>2.5</td>
<td>Cueing effect (RT) for Hint and No Hint groups in Experiment 3</td>
<td>66</td>
</tr>
<tr>
<td>2.6</td>
<td>Recognition and prediction test performance in Experiment 3</td>
<td>67</td>
</tr>
<tr>
<td>2.7</td>
<td>Cueing effect (RT) for Hint and No Hint groups in Experiment 4</td>
<td>73</td>
</tr>
<tr>
<td>2.8</td>
<td>Recognition and prediction test performance in Experiment 4</td>
<td>74</td>
</tr>
<tr>
<td>2.9</td>
<td>Cueing effect (RT) for Hint and No Hint groups in Experiment 5</td>
<td>78</td>
</tr>
<tr>
<td>2.10</td>
<td>Recognition and prediction test performance in Experiment 5</td>
<td>80</td>
</tr>
<tr>
<td>2.11</td>
<td>Cueing effect (RT) for two hint groups in Experiment 6</td>
<td>85</td>
</tr>
<tr>
<td>2.12</td>
<td>Sequence used in Jiménez, Méndez, &amp; Cleeremans (1996)</td>
<td>89</td>
</tr>
<tr>
<td>2.13</td>
<td>SRT task used in Sanchez &amp; Reber (2013). Reprinted from Cognition, 126, Sanchez, D. J. &amp; Reber, P. J., Explicit pretraining instruction does not improve implicit perceptual-motor sequence learning, pp. 341-351, Copyright (2013), with permission from Elsevier</td>
<td>90</td>
</tr>
<tr>
<td>3.1</td>
<td>Examples of circle and line stimuli used in Experiments 1-3</td>
<td>111</td>
</tr>
<tr>
<td>3.2</td>
<td>Prototypicality gradients for all four groups in Experiment 1</td>
<td>116</td>
</tr>
<tr>
<td>3.3</td>
<td>Examples of circle and line stimuli seen during the visual search task in Experiments 2 and 3</td>
<td>122</td>
</tr>
<tr>
<td>3.4</td>
<td>Prototypicality gradients in the Memorize and Search groups in Experiment 2</td>
<td>125</td>
</tr>
<tr>
<td>3.5</td>
<td>Results from the signal detection analysis in Experiment 2</td>
<td>128</td>
</tr>
<tr>
<td>3.6</td>
<td>Comparison of the Search group in Experiment 2 to the No Exposure-Familiarity group in Experiment 1</td>
<td>129</td>
</tr>
<tr>
<td>3.7</td>
<td>Prototypicality gradients in the Memorize, Search, and Search-Terminate groups in Experiment 3</td>
<td>134</td>
</tr>
<tr>
<td>3.8</td>
<td>Results from the signal detection analysis in Experiment 3</td>
<td>136</td>
</tr>
<tr>
<td>4.1</td>
<td>Simulations of rule- and similarity-based generalization</td>
<td>153</td>
</tr>
</tbody>
</table>
Figure 4.2  Examples of stimuli seen in the training phase of Experiments 1 and 2  

Figure 4.3  Schematic diagram of the test stimuli in Experiment 1 and 2A  

Figure 4.4  Schematic diagram of the test stimuli in Experiment 2B  

Figure 4.5  Examples of the test stimuli seen by the Inconsistent Group  

Figure 4.6  Results from the 3AFC self-report question in Experiment 1  

Figure 4.7  Results from the 2AFC rule-identification question in Experiment 1  

Figure 4.8  Generalization (category judgements and typicality ratings) on the more attended dimension in Experiment 1  

Figure 4.9  Generalization (category judgements and typicality ratings) on the less attended dimension in Experiment 1  

Figure 4.10  Results from the 3AFC self-report question in Experiment 2  

Figure 4.11  Results from the 2AFC rule-identification question in Experiment 2  

Figure 4.12  Generalization (category judgements and typicality ratings) on the attended dimension in Experiment 2  

Figure 4.13  Generalization (category judgements and typicality ratings) on the unattended dimension in Experiment 2  

Figure 4.14  Results from the sequential reanalysis combining categorization accuracy and typicality ratings from Experiments 1 and 2
List of Tables

Table 2.1  Subsequences at second-, third-, and fourth-order levels coded according to direction, repetitions and alternations of target direction, and target location.  49
Table 2.2  Frequencies and probabilities of subsequences at second-, third-, and fourth-order levels  61
Table 3.1  Parameters used to create the circle and line stimulus sets for Experiments 1-3  111
Table 4.1  Number of icons in each stimulus along the artificial dimension in Wills & Mackintosh (1998)  154
Table 4.2  Distortion values for test stimuli  166
Table 4.3  Parameters used to create the training and test stimuli  171
Table 4.4  Results from the 3AFC self-report question in Experiment 1  179
List of Publications

Experiments 1 and 3 (Chapter 2) of this thesis is published as:

Experiments 1 and 2 (Chapter 2) of this thesis are published as:

Experiments 1, 2, and 3 (Chapter 3) of this thesis are published as:
List of Conference Presentations


Lee, J. C., & Livesey, E. J. (2014, November). Generalization along an attended and unattended dimension following category learning in humans. Poster presented at the 55th Annual Meeting of the Psychonomic Society, Long Beach, California, USA.


Chapter 1: General Introduction

“Learning without thought is labor lost.”
-Confucius

Humans possess a remarkable ability to learn about and adapt to their environment. However, learning is not uniform across situations. Variables including the exact instructions given, the goals and requirements of the task, and the knowledge possessed by the individual about the contingencies and rules in the task can all determine how an individual engages with the task and therefore what is learned. Manipulating these task parameters can thus create situations in which learning might occur incidentally in the absence of reasoned thought, or situations in which learning is intentional and effortful. The degree to which an individual has the intention to learn will henceforth be referred to as learning orientation. In this thesis, the effects of learning orientation, as manipulated through verbal instructions, will be explored.

The aim of this thesis is to explore how the content of learning changes as a result of manipulating learning orientation, and to examine the theoretical implications of such effects. This chapter will briefly outline the historical development of learning theory, describing the origins of conditioning and its connection with the contemporary field of associative learning. Theoretical debates in learning theory relevant to this thesis will then be discussed, along with the significance of learning orientation within each debate. It will be shown that the effects of orientation have implications for the question of whether humans possess separable learning processes, as well as the question of whether a common mechanism underlies human and animal learning. Finally, the three different paradigms (prototype-category learning, sequence learning, post-discrimination
generalization) and manipulations used in this thesis will be summarized and research questions posed.

1.1 The Nature of Associative Learning

Learning in this context refers to enduring changes in an organism that result from experience and make a change in behavior possible. By this definition, learning encompasses a wide range of behaviors which can include anticipatory behavior towards stimuli that signal pleasant or aversive consequences, faster responses to impending events that can be predicted based on the recent history of trials, or an ability to generalize knowledge of experienced instances to novel instances that have never been encountered. The goal of learning theory is to formulate a set of general principles under which a wide range of behavior can be understood. This section will briefly outline how, in pursuing this goal, the psychology of learning has been historically linked to ideas of associationism and automaticity.

Learning theory has been dominated by associationism (De Houwer, 2009; Shanks, 2010), the idea that learning occurs as the result of the formation of associative links between mental representations of events or stimuli, such that activation of one idea can ‘bring to mind’ another. This idea of associative thought can be traced back to Aristotle, who presented a number of laws governing the kinds of ideas that could be brought to mind by other ideas (Aristotle, trans. 1906). For example, Aristotle’s ‘Law of Contiguity’ states that recall of an object will elicit the recall of other things that were experienced with that object, with these associations forming the basis for memories. Associative links were also credited as the building blocks for the emergence of complex ideas by British empiricists such as Hume and

---

1 Aristotle’s laws of Contiguity (recall of things experienced with an object), Similarity (recall of things similar to an object), and Frequency (increased probability of recalling something that has been
Locke. For example, Locke believed that since we could not directly observe causality, our causal understanding of the world was an inference arising from empirical observation of associations between events (Locke, 1690). Thus, the idea of associations being the building blocks of mental life emerged quite early in philosophy.

Despite the prominence of associationism in philosophy, associative learning as a field of scientific enquiry did not develop until the discovery of classical conditioning by Ivan Pavlov. Pavlov conducted experiments on animals where he observed changes in their behavior as a result of pairing stimuli together in their environment. Pavlov (1927) termed this process ‘classical conditioning’, and described it as the result of repeated pairings of an initially neutral conditioned stimulus (CS) with a biologically relevant unconditioned stimulus (US), with the CS acquiring the ability to elicit a conditioned response (CR) through its association with the US. Conditioned responding was thought to be reflexive, elicited automatically by external stimuli and operating without careful thought or evaluation. In fact, Pavlov himself termed the CR the “psychic reflex”, likening conditioning to the automaticity of a biological reflex (Pavlov, 1927).

Instrumental conditioning, whereby animals learn to perform specific behaviors as a result of associations between responses and their consequences and antecedents, was similarly thought to operate automatically. Instrumental learning was first investigated by Thorndike, who coined the “law of effect” (Thorndike, 1911) to describe how behavior was controlled by its consequences. According to the law of effect, behaviors that are followed by pleasurable consequences are reinforced, meaning that the animal is more likely to repeat the action when in a similar context. Thorndike explained how stimuli in the environment came to control behavior
through the “stamping in” of connections between a stimulus and a response (S-R) through reinforcement. Thorndike believed that this S-R learning could occur implicitly or without awareness of the learning process (Thorndike, 1911). Thus it is clear that early theorists in both classical and instrumental conditioning conceived of the process of learning as automatic and associative.

The contemporary field of ‘associative learning’ encompasses a wider range of associations than just those of simple conditioning between a CS and a US or a stimulus and a response. Associative learning is concerned with associations that form between mental representations of any events or stimuli. The idea of learning being reflexive was gradually replaced as researchers began to discover that many conditions needed to be met in order for associations to form and learning to occur. For instance, the phenomenon of blocking (Kamin, 1969), shows that the mere co-occurrence of two stimuli is not sufficient for an association to form. Blocking occurs when a stimulus (A) is first paired with an outcome (+) in an initial phase, and is then presented with a novel stimulus along with that same outcome in a second phase (A+/AB+). Compared to control cues that are presented in compound only in the second phase (CD+), learning about the blocked cue B is usually weaker than learning to C or D, despite the blocked and control cues being paired with the outcome an equal number of times. Blocking was first demonstrated by Kamin (1969), and was significant in showing that when multiple stimuli were presented with an outcome, competition amongst those stimuli for association with the outcome (i.e. cue competition) occurred.

---

2 Thorndike is often credited as the pioneer of implicit learning, for this reason.
3 Although, even at these early stages, precursors to the cognitive view were starting to emerge. For example, Tolman believed that learning was goal-directed and intentional and that focus on internal states was necessary to explain behavior (Tolman, 1938, 1948).
Cue competition effects such as blocking demonstrate that associations do not form unconditionally. Models of learning formulated after the discovery of blocking thus attempted to specify the conditions that dictated whether and to what extent learning occurs. For example, according to the Rescorla-Wagner model (Rescorla & Wagner, 1972), in order for learning to occur, there must be a discrepancy between what an animal expects will happen, and what actually happens. In other words, there must be a prediction error that makes a stimulus surprising in order for learning to occur. Using the notion of prediction error, cue competition effects are explainable using the same associative links used to explain simple conditioning, and do not require explanation in terms of higher-order problem-solving or decision processes. That is, learning is controlled by the predictive value inherent in the stimulus’ contingency with other events and is detected by the animal through its experience of the environment. Other associative models describe how learning is dependent on, and modulated by attention (Mackintosh, 1975; Pearce & Hall, 1980), or place emphasis on concepts such as short-term memory (Wagner, 1981). These associative models are considerably more complex than the simple CS-US or S-R links originally proposed to explain learning, and also acknowledge a role for other cognitive functions.

One interpretation of these models is that despite the modulating effects of processes such as attention and memory, associative learning is automatic in that it is essentially governed by mechanistic operations. In other words, the models are assumed to contain all the mental components that are necessary for learning to occur. Therefore, in this context, automaticity does not mean that learning always occurs, but rather once the relevant conditions are satisfied, learning (the formation of associations) proceeds in an incremental, continuous, and lawful fashion. It is this
definition of automatic that will be used throughout this thesis (see Hasher & Zacks, 1979; Schneider & Shiffrin, 1977; and Shiffrin & Schneider, 1977, for more comprehensive discussion of the concept of automaticity). Note that most associative models are silent regarding the effects of effortful cognitive processes, implying that their operation does not depend on the intentions of the individual as long as the relevant information is processed. Such associative mechanisms are remarkably simple but have been hugely successful in explaining a wide range of sophisticated behavior in animals.

1.2 Learning in Humans

The translation of animal learning paradigms and application of associative learning theory to humans during and after the 1980s brought novel challenges in designing human paradigms that were analogous to animal paradigms, and the related difficulty in accounting for the obvious fact that humans, when confronted with a task, will tend to reason, formulate hypotheses and derive rules about the situation. This is especially the case if humans are given instructions that encourage them to learn or solve the problem at hand. In other words (to state the obvious), humans think, and importantly for learning theory, they think in ways that are said to be beyond the capabilities of other animals (Herrnstein, 1990; Penn, Holyoak, & Povenelli, 2008). Much of the early human learning research was aimed at establishing that conditioning effects found in animals do in fact occur in humans (e.g. Alloy & Abramson, 1979; Dickinson, Shanks, & Evenden, 1984). However, it also became clear that humans were capable of forms of learning that lie beyond the capabilities of animals. For example, humans are able to form abstract rules that allow them to generalize beyond the physical features of their experiences to novel instances that
may not share many physical features with previous experiences in the experimental context (Shanks & Darby, 1998; Penn et al., 2008; Wills & Mackintosh, 1998). Further, while humans demonstrate similar learning effects to that of conditioning in animals, the strength of particular learning effects can depend on the broader assumptions about the underlying rules of the causal scenario (Shanks, 2010), making it difficult to attribute learning effects in humans solely to associative learning mechanisms.

This departure from the cognitive abilities of animals has been attributed to the capacity for humans to use language and higher-order cognitive processes to form rules and reason about associations in their environment (Carruthers, 2002; Penn et al., 2008). This capacity also means that human learning can be influenced in ways in which animal learning cannot. The precise instructions given about the task, whether the learner is given additional explicit knowledge about the content to-be-learned, and whether learning is intentional or incidental to the task performed, can all potentially affect the content of learning. If learning in a given task is altered by instructions, this has practical implications for experimental design, and theoretical implications for the interpretation of the results, especially if the aim is to investigate associative processes and the experiment is designed to be analogous to animal experiments. Researchers who wish to compare results between different paradigms or even between studies that use the same task need to know what potential differences could arise through altering the instructions and demands of the learning task.

More generally, the effects of learning orientation are important because a complete theory of learning should be able to account for the ways in which learning is altered as a result of cognitive manipulation. The potential impact of verbal instructions is suggestive of an additional contribution to learning that is verbally
mediated. In other words, if learning is susceptible to instructional manipulations, associative learning mechanisms may well be cognitively penetrable and are not sufficient to explain human learning.

1.3 Dual-Process Theories

Dual-process theories postulate two qualitatively different processes through which humans learn. Within such theories, separate debates exist about whether learning is driven by a combination of associative and ‘propositional’ processes (Mitchell, De Houwer, & Lovibond, 2009), and whether there exist separable ‘implicit’ and ‘explicit’ modes of learning (Reber, 1987). In order to frame the theoretical significance of the manipulations contained in this thesis, the associative/propositional and implicit/explicit distinctions will be briefly reviewed, as well as the distinction between rules and associations. These reviews will be brief and highly selective, since a more specific review of the theoretical issue with respect to particular paradigms will be undertaken within each of the following chapters. Note that while the subject of this thesis is relevant to these dual-process debates (and will be interpreted with respect to each), the aim of the ensuing chapters is not to conclude in favor of one process or another, nor to defend the single- or dual-process approach, but rather to characterize the nature of the interaction of these processes, should they each exist. Note that this approach of comparing learning conditions is equally relevant to single-process theories of learning, since they too should be able to specify how a single process can come to elicit different behavior as a result of changing verbal instructions.

---

4 Note that the debate will be framed in terms of dual-processes and not dual-systems. While these terms are often used interchangeably in the literature (e.g. Stanovich & West, 2000), the idea of ‘systems’ implies modularity and complete independence, while ‘processes’ might interact within a single system.
Neisser (1963) stated, “The psychology of thinking seems to breed dichotomies” (p. 1). Indeed, dual-process theories abound in cognitive psychology, with two processes posited to explain a wide range of cognitive functions such as memory (Jacoby, 1991; Schacter, 1987), reasoning (Evans, 2003; Sloman, 1996), and decision-making (Kahneman & Frederick, 2002). Although there is some variation between theories, attempts have been made to map dual-process theories onto a general System I and System II (Stanovich & West, 2000). System I is usually described as a more primitive but automatic system operating non-analytically on associative information, while System II is described as a more sophisticated but effortful system, operating analytically on symbolic information (Evans, 2008). The distinction between systems can be drawn in different ways, but most theorists agree that what is important in distinguishing between systems is their computation (e.g. Sloman, 1996). In other words, the learning process differs in qualitative ways, even if the content of learning and effects on behavior might appear to be the same. Nevertheless, the content of learning (and in particular, awareness) has often been taken as evidence of the operation of a specific learning process. As will be discussed below, this has proved to be highly controversial.

1.3.1 Associative and Propositional Learning

The distinction between associative thought where ideas simply ‘come to mind’ and effortful, reasoned thought can be traced back to William James (1890). The notion that learning is sometimes driven by a carefully controlled reasoning process, and at other times driven by an automatic link-formation mechanism has intuitive appeal, which may account for the proliferation of dual-process theories in cognitive psychology. It seems natural that at some times, learning is effortful,
carefully reasoned and results in conscious knowledge, but at other times, it occurs without much careful thought or intention. As will be discussed, one of the most controversial issues in learning theory concerns whether the obvious role of higher-order cognition in human learning means that an associative mechanism, as traditionally conceived, is not needed to explain human behavior (Shanks, 2010).

The notion that humans and animals share a common learning mechanism was popularized with the discovery that many of the conditioning phenomena found in animals were also found in human studies where participants were required to actively learn causal relationships (e.g. Dickinson, Shanks, & Evenden, 1984). However, despite similarities between human and animal behavior, it was not clear whether learning was accomplished through the same primitive mechanism, leading some to propose a rejection of associative mechanisms in humans altogether (Mitchell et al., 2009). Multiple lines of evidence motivated this conclusion. Several reviews concluded that there was little evidence to support the idea that conditioning occurs in the absence of awareness (e.g. Brewer, 1974; Lovibond & Shanks, 2002), early studies showed that the same conditioning effects could be found by simply giving participants the equivalent instructions (e.g. Colgan, 1970; Cook & Harris, 1937), and manipulating factors that influenced reasoning also affected cue competition effects in learning (e.g. Mitchell & Lovibond, 2002; Waldmann & Holyoak, 1990). Together, these studies suggest that rather than learning being the result of automatic associative links, learning involves deliberate reasoning and higher-order cognitive processes.

This was exactly the stance taken by Mitchell et al. (2009) in proposing the propositional account of learning, which attributes associative learning phenomena in
Propositional learning involves manipulation of symbolic information to generate inferences. The output of this learning process is conscious and propositional in nature, meaning that it contains symbolic information and can be evaluated to be true or false (e.g. X was followed by Y). Such a learning process is clearly different to an associative mechanism, both in process and in the psychological qualities of the output. However, distinguishing between learning that arises due to associative and propositional mechanisms is not a trivial task. This is because many human learning phenomena are explainable using both systems (see Shanks, 2010, for a comparison of associative and cognitive accounts of blocking). One of the more popular ways of distinguishing between the two systems has been awareness (Lovibond & Shanks, 2002). Propositional learning is assumed to be necessarily conscious, while an associative model has no explicit representation of awareness. This is because associative models were formulated to explain animal behavior, and consciousness has not traditionally been attributed to animals (Heyes & Dickinson, 1990). Therefore, showing that learning is possible in the absence of awareness would be strong evidence for an associative process (Lovibond & Shanks, 2002).

In practice, convincing evidence of unaware learning has been difficult to attain due to various methodological problems in measuring awareness (see Lovibond & Shanks, 2002). For example, researchers do not usually ensure that their measure of awareness is equally sensitive to their measure of learning, often having a large number of training trials (measuring learning) but only a small number of test trials (measuring awareness). Lovibond and Shanks concluded that using their strict criteria

---

5 Within this thesis, propositional learning and associative learning will be treated as qualitatively different processes, ignoring the fact that propositions or abstract representations might be an emergent feature of more complex associative networks (e.g. see McClelland & Rumelhart, 1985, McClelland, Rumelhart, & the PDP research group, 1986).
for assessing awareness, there was no convincing evidence of unconscious learning. While the methodological issues raised by Lovibond and Shanks concerning previous assessments of awareness are valid, there is still considerable controversy over whether unconscious learning occurs, especially from the implicit learning literature, which will be reviewed shortly. Note also that if it is accepted that an associative mechanism can produce declarative, explicit knowledge (e.g. see Shanks, 2007), then the usefulness of awareness as a distinguishing feature becomes questionable.

An alternative approach to distinguishing between associative and propositional processes uses the ‘irrationality’ of behavior as indicative of associative learning (Shanks, 2007). A good example is the phenomenon of second-order conditioning (e.g. Rescorla, 1973), where one stimulus is paired with an unconditioned stimulus (A+), and subsequently a second stimulus is paired with the first stimulus (AB). The second stimulus (B) usually elicits a conditioned response, despite never being paired with the outcome. This effect is non-rational from a propositional perspective, since the addition of B to the compound AB produces no outcome, so participants should learn that B is inhibitory rather than excitatory (Karazinov & Boakes, 2007). Second-order conditioning is uniquely predicted by associative theories and accordingly, has been found to eventuate in causal learning scenarios only when participants undergo paced training trials that restrict their opportunity to think (Karazinov & Boakes, 2007, Lee & Livesey, 2012). Further, second-order conditioning disappears when participants are presented with all the relevant contingencies and asked to make the ‘correct’ inference (Lee & Livesey, 2012). However, Mitchell et al. (2009) have noted that it is possible for the propositional system to reason incorrectly, and therefore the irrationality of learning is not convincing evidence for an associative process. Thus, it is clear that using the
output of learning to determine whether an associative or propositional system is operating is far from definitive.

Mitchell et al. (2009) state that in contrast to the clear evidence for the role of higher-order reasoning in human learning, there is little evidence to support an automatic associative mechanism in humans (using awareness or otherwise), and therefore all human learning should be explained propositionally. While this conclusion is extreme in abandoning associative mechanisms, it is impossible to deny the role of higher-order cognitive processes in human learning. Cognitive manipulations and instructions have large effects on learning, and are even able to replace experience of the actual contingencies (e.g. Colgan, 1970; Cook & Harris, 1937; Lovibond, 2003). Thus, one way to account for the susceptibility of learning to cognitive influences is to adopt a dual-process stance and claim that humans learn propositionally and associatively.

Proposing the existence of both an associative and propositional learning system has multiple advantages. It allows us to retain associative learning mechanisms to explain animal learning and suggests a degree of continuity in the mental faculties of humans and animals. Most importantly for this thesis, a dual-process theory readily accounts for any potential effects of instructional manipulations, as we can postulate that verbal instructions affect the relative contribution of each learning process to behavior. Situations in which humans are given verbal instructions to learn within a task might allow the propositional system to dominate as the learner has additional motivation to learn and would actively test hypotheses and reason about salient events in their environment. In contrast, associative learning mechanisms may only be evident under incidental learning.

---

6 Indeed, even the radical behaviorists did not deny that ‘private’ mental events existed, they just rejected them in the pursuit of objective, scientific investigation.
conditions where the participant does not have the intention to learn (Mackintosh, 1995), or cognitive control is low due to the nature of the task (McLaren et al., 2014). Such a dual-process theory implies that the propositional learning process has primacy in determining behavior, with associative processes only able to influence behavior when propositional learning is hindered in some way (Mackintosh, 1995; McLaren et al., 2014). This is entirely plausible since propositional learning entails conscious knowledge of both the output and the process of learning, providing a stronger justification for behavior. Any effect that is not readily explained by associative theory can be accounted for by assuming that the propositional system had a larger influence on behavior. The dual-process approach thus readily accommodates differences in the content of learning that arise as the result of verbal instruction or changing learning orientation.

It is important to note at this point that this flexibility of the dual-process model to accommodate any result post-hoc by attributing it freely to either mechanism has led to some suggesting that it is unfalsifiable (Mitchell et al., 2009). While in practice it is true that such a dual-process model is difficult to test, some attempt has been made by McLaren and colleagues to specify the manner in which these processes interact (e.g. McLaren et al., 2014). This is important since a dual-process theory should be able to specify a priori the kinds of situations under which associative and propositional mechanisms will dominate, rather than simply attributing results to the more appropriate system. Dual-process theories assume that propositional processes are more controlled and deliberate, have a large effect on learning and therefore the capability to suppress the operation of associative processes, or at least mask its effects (Livesey & McLaren, 2009; McLaren et al., 2014). However, another assumption made by some dual-process theorists is that associative
processes operate automatically while propositional processes can be engaged and disengaged more or less at will (McLaren et al., 2014). This implies that the dual-process view should be able to predict differences in learning that arise when comparing conditions where propositional processes are encouraged versus situations where they are not.

McLaren et al. (2014) have provided the most precise specification to date of a dual-process theory involving associative and propositional processes. They propose that the relative influence of each process is determined by the degree of cognitive control exerted over the behavior. One implication of this is that propositional processes will dominate under situations where cognitive control is high and evidence of associative processes might only manifest in situations where cognitive control is low. Similarly, Mackintosh (1995) argued that in order to study associative learning in people, we either need to distract them from the task by presenting them with a cover story, make them work rapidly, or overload their working memory so they cannot ‘figure out’ what is going on. In other words, associative learning can only be observed if measures are taken to ensure that participants are not actively thinking about the task at hand, but may occur incidentally in the absence of conscious thought. If we accept this potential interaction between learning mechanisms, then it follows that verbal instructions can be used to change the relative influence of propositional and associative processes, by encouraging or hindering the use of propositional learning.

Alternatively, under a single-process view, it may be the case that people always learn propositionally, and that instructions facilitate more effective and accurate reasoning. For example, additional instructions may simply enhance propositional learning processes by increasing motivation to learn, or drawing
attention to relevant aspects of the task or content to-be-learned that otherwise would have been ignored. While this is certainly possible, the propositional account as it stands currently does not specify the way in which learning can differ as a result of changing learning orientation, nor what the minimal conditions are for propositional learning to occur (e.g. whether incidental learning is possible). Thus, while the role of instructions in learning does not provide clear evidence for either a single- or dual-process view of learning (and indeed it is not the aim of this thesis to discriminate between these two views), it is certainly relevant for future theory development from both perspectives. Human learning is flexible, and can be influenced by a wide variety of factors. Specifying how learning orientation can affect learning is needed for a complete understanding of behavior, and will aid in characterizing the potential interaction between propositional and associative processes, should they both exist. The need to move away from proving the existence of associative processes, towards characterizing the interaction between associative learning and other forms of cognition has been voiced by multiple researchers from both sides of the debate as a more fruitful avenue of future research (McLaren et al., 2014; Mitchell et al., 2009; Sloman, 1996). Adopting this research goal would lend predictive utility to the dual-systems approach as well as offer more careful specification of the exact contribution of propositional processes to human learning.

1.3.2 Implicit and Explicit Learning

A separate debate has been raging for the past 50 years concerning whether there are independent implicit and explicit learning processes. Unlike associative learning, the field of implicit learning grew from an interest in language acquisition borne out of the ‘cognitive revolution’ that swept psychology in the 1950s. The term
“implicit learning” was originally coined by Reber (1967), whose research was inspired by the observation that humans seemed adept at learning complex rules inherent in language without the ability to report knowledge of those rules or awareness of the process of learning. Reber’s experiments involved presenting participants with nonsense letter strings that conformed to a specific artificial grammar consisting of rules determining whether different combinations of letter strings were permissible. Despite participants only being told to memorize the strings, they were subsequently able to correctly categorize novel strings as grammatical or nongrammatical (Reber, 1967). Reber interpreted this result as evidence of a fundamental ability to induce rules implicitly in the absence of awareness. Reber’s initial findings sparked a wealth of research investigating whether it was possible to learn implicitly without awareness, and whether implicit learning represented an independent mode of learning.

Controversy over how to define implicit learning and the exact characteristics of implicit learning has dominated learning research since. In particular, Reber’s (1967) original conceptualization of implicit learning as rule abstraction in the absence of awareness was challenged by findings showing that artificial grammar learning tasks could be learnt through memorizing specific pairs or chunks of the grammar (e.g. Perruchet & Pacteau, 1990), and that an ability to recognize grammatical strings was accompanied by explicit knowledge given an appropriate awareness measure (see Shanks & St. John, 1994). The field of implicit learning moved away from the idea of unconscious rule abstraction and today, is most commonly defined as learning without awareness (Reber, 1987). Thus, the role of awareness in learning has become paramount in demonstrating the existence of implicit learning (Shanks & St. John, 1994), and as such, much of the implicit
learning literature is devoted to this issue. The debate about the necessity of
awareness in implicit learning continues to this day (see Lovibond & Shanks, 2002,
and Shanks, 2010, for parallel discussion of the role of awareness in conditioning).

While some researchers might agree that implicit learning is a qualitatively
distinct mode of learning, it is much more mysterious what its defining properties are.
Different authors use the term ‘implicit’ to refer to a wide range of things such as the
nature of the learning process, the nature of the memory retrieval process, and the
type of resultant knowledge (Frensch, 1998). Different authors also ascribe different
meanings to the word ‘implicit’, further complicating the debate. Frensch classified a
wide range of meanings of implicit learning into two broad categories:
unconscious/unaware and nonintentional/automatic. He concluded that the most
scientifically useful way to define implicit learning was nonintentional/automatic
rather than unconscious/unaware, since awareness tests could never be ensured to be
sufficiently sensitive to show that awareness is absent (Shanks & St. John, 1994), and
defining implicit learning as nonintentional/automatic leads to a number of testable
predictions regarding its functional characteristics (see Hasher & Zacks, 1979, for a
discussion of differences between automatic and controlled processes).

Indeed, many of the paradigms used to demonstrate implicit learning are tasks
in which participants are either not informed about the presence of structured material
they could learn, or are performing other tasks that made any learning that occurred
incidental. Therefore it seems reasonable to assume that implicit learning does not
require an intention to learn, and in this way can be said to be different to explicit
learning, which necessarily involves deliberate, controlled processes. The fact that
implicit learning occurs incidentally has led some to propose that implicit learning
represents a by-product of attentive information processing (Jiménez, 2003). However,
a single-system account might describe the phenomenon of implicit learning (as stated) as a situation where learning is easily accomplished with minimal cognitive effort, and in which the resultant knowledge is difficult to report, perhaps due to the nature of the content being complex or difficult to describe verbally (see Perruchet & Amorim, 1992; Shanks & St John, 1994).

Indeed, there are many demonstrations of reportable explicit knowledge in implicit learning paradigms given appropriately sensitive awareness tests (e.g. Perruchet & Amorim, 1992). Under a single-system account, learning occurs given the necessary conditions, and awareness is always present, but whether it is detected or not is determined by the methodology used for a particular experiment. This graded view of implicit learning allows for awareness to emerge at some threshold along the continuum, an idea that is supported by connectionist networks of implicit learning (Cleeremans, 2008). If it is accepted that concepts such as awareness are dynamic then implicit and explicit can be thought of as simply two extremes on a single continuum (Cleeremans, 1994, 1995, 2008), removing the (false) dichotomy between implicit and explicit learning.\(^7\)

Nevertheless, many researchers still view implicit learning as qualitatively distinct from explicit learning. In addition to its functional characteristics, another way in which implicit learning is generally more accepted to differ from explicit learning is in its resultant output. In contrast to Reber’s (1967) original idea of implicit learning as unconscious abstraction of rules, implicit learning as a mechanism might be better thought of as an accumulator of statistical regularities (Cleeremans, 1996). Thus, the output of implicit learning is represented in terms of instances or simple associations, rather than abstract rules of the sort that Reber (1967) originally

---

\(^7\) Interestingly, it was Reber (1993) who cautioned against being seduced by the “polarity fallacy”.

19
surmised. It has also been suggested that representations that result from implicit learning are weaker than that of explicit learning, but are still able to influence behavior (Cleeremans, 2006). This thesis will not attempt to clarify the properties of implicit learning as a process or assess the qualitative properties of implicit knowledge. However, a working definition is needed due to the multiple definitions of implicit learning that punctuate the literature (Frensch, 1998). Thus, for the purposes of this thesis, implicit learning will be conceptualized as a learning mechanism that operates incidentally without the intention to learn, and whose output is approximated by an incremental learning mechanism accumulating knowledge from instances.

In addition to disagreement about the definition of implicit learning, there is also disagreement on how implicit and explicit processes interact. Much of the implicit learning literature has focused on establishing its existence as a qualitatively distinct learning process or showing that awareness is absent, with far less research testing the potential interaction between implicit and explicit learning, should the distinction between the two be meaningful. Note that conceiving of implicit learning as unconscious, or non-intentional, or as an incremental accumulator of information, is silent as to the effects of abstract, conscious knowledge. Some researchers have adopted the position that implicit learning is completely independent of explicit processes and knowledge (e.g. Berry & Broadbent, 1988; Curran & Keele, 1993; Reber, 1990; Stanley, Mathews, Buss, & Kotler-Cope, 1989; Willingham, Nissen, & Bullemer, 1989). This extreme position claims that implicit learning is “cognitively impenetrable”, with implicit and explicit learning processes producing independent forms of knowledge. This view is supported by studies showing that the degree of implicit learning is not altered by providing participants with the intention to learn or
giving them knowledge about the content to-be-learned (e.g. Jimenez, Mendez, & Cleeremans, 1996). If there are situations in which explicit processes do not improve learning, this might suggest that any learning observed under normal incidental conditions can only be explained using implicit learning processes, since the addition of explicit learning has no effect. In other words, if there are some situations in which explicit processes are not efficient, there must be another (implicit) learning process that operates optimally under those conditions (e.g. Reber, 1989).

Alternatively, it could be that while implicit and explicit learning processes are separable, they interact in such a way that the addition of explicit knowledge changes learning qualitatively or quantitatively. In this thesis, quantitative changes refer to changes in the amount of learning, while qualitative changes refer to changes in the content of learning (e.g. learning of rules vs. associations). Quantitative changes in learning are not especially useful in determining whether a task is tapping into implicit learning since such an effect can easily be explained by the additional instructions boosting explicit learning processes or strengthening the resulting knowledge. Qualitative changes however, have been argued to provide stronger evidence that implicit and explicit processes operate differently and are dissociable (Jones & McLaren, 2009).

Thus in implicit learning, the comparison between learning under incidental conditions, and learning under intentional conditions where participants are given additional information about the task have been utilized to show that there are separable implicit and explicit learning processes (e.g. Dominey, Lelekov, Ventre-Dominey, & Jeannerod, 1998; Jones & McLaren, 2009; Sanchez & Reber, 2013). In order to draw these conclusions, many of these studies (and the studies contained in this thesis) assume that incidental conditions are sufficient to engage implicit learning,
and intentional learning conditions are sufficient to engage explicit learning. It will not be assumed that each process is exclusively engaged due to the difficulty in ensuring that tasks are ‘process-pure’ (Merikle & Reingold, 1991; Reingold & Merikle, 1988). Thus, the aim of this thesis is not to prove that learning under incidental conditions is implicit but rather to examine the effects of verbal instructions and explicit learning on learning in tasks where it is believed that learning can occur implicitly.

1.3.3 Rules and Associations

At this point, it is worth defining and introducing the significance of rules in learning, since they are often cited as a type of learning that is uncontroversial in incorporating explicit higher-order abilities such as reasoning and hypothesis-testing (Sloman, 1996). Rules in this context refer to mental representations consisting of variables and logical statements. Since they can apply to a whole class of stimuli that satisfy its constraints, they are necessarily abstract (Sloman, 1996). This means that rules allow for generalization of knowledge in ways that go beyond the surface similarity of a novel instance and a known instance (Shanks & Darby, 1998). Most importantly, rules are assumed to be real psychological entities that have a causal relationship with behavior (i.e. behavior that is rule-following), in contrast to a description that merely happens to be consistent with behavior (i.e. behavior that is rule-conforming, see Smith, Langston, & Nisbett, 1992, for more on this distinction). Sloman (1996) has stated that all representations can be couched in terms of rules, and thus it is important to clearly define what is meant by a rule so that systems of learning that do not use rules can be identified empirically and the term does not become empirically vacuous. Therefore, henceforth the word ‘rule’ will be used to
describe abstract, symbolic representations only and not descriptions of simple
associations or any other behavior traditionally considered explainable in associative
terms (e.g. responding to a novel instance based on the similarity to a known instance).

With this definition of rules in mind, rule learning can be clearly contrasted
with learning that is instance-based or purely associative in content. As discussed
above, rules are necessarily abstract and therefore their content is markedly different
from that of simple associations. In certain tasks, the presence of rules is clearly
distinguishable in the pattern of behavior exhibited. For example, in a simple
discrimination task between two stimuli lying on the same dimension (e.g. a green
circle leading to no reward and a blueish-green circle leading to a reward),
participants can either learn about the instances (i.e. that each individual stimulus
leads to its respective outcome), or they can derive a relational rule describing the
difference between the stimuli (i.e. that bluer circles lead to rewards). Whether
learning is instance-based or rule-based can then be discerned from testing how
participants generalize to novel stimuli along the dimension. If participants have
merely learned about the instances they have seen, they should generalize on the basis
of similarity to the known instances, with their behavior indicating the highest level of
expectancy of reward for stimuli similar to the blueish-green circle. If, however,
participants have learned a relational rule, their behavior should indicate the highest
level of expectancy of reward for the bluest circles. Thus throughout this thesis, some
of the instructions will encourage participants to form rules to ensure use of
propositional and explicit learning, and in order to provide a clear contrast to
situations in which participants may be learning associatively or implicitly.
1.4 Some Commonalities

While the associative/propositional and implicit/explicit debates have been presented here as independent, there are obvious similarities between associative and implicit learning. Conceiving of implicit learning as the result of incremental learning mechanisms sensitive to statistical structure in the environment bears obvious resemblances to an associative learning mechanism. This has led some theorists to propose that implicit learning effects in humans are indicative of an associative system in which associations are learnt automatically (Mackintosh, 1995; McLaren, Green, & Mackintosh, 1994). If this notion of implicit learning is adopted, this would suggest that what controls the relative influence of associative and propositional processes is the nature of the task itself. Many implicit learning tasks contain complex stimuli that are difficult to encode propositionally and are not conducive to reasoning (e.g. contextual cueing, Chun & Jiang, 1998), or are speeded tasks that do not allow the participant to reason about the content of learning (e.g. serial reaction time tasks, Nissen & Bullemer, 1987). Implicit learning tasks do not involve telling the participant that there is content to be learned and are easy to perform without learning this content, and thus the participant does not usually have the motivation to reason or formulate rules about the task. In other words, implicit learning tasks may be instances where propositional influences have little chance to operate, allowing behavior to be determined primarily by associative mechanisms.

1.4.1 Learning without Awareness?

As discussed above, independence from awareness has been cited as a defining feature of both associative and implicit learning, and much of the literature has been devoted to examining this issue (Lovibond & Shanks, 2001; Shanks & St.
However, if it is accepted that awareness can emerge as a result of associative and implicit learning (e.g. see Cleeremans, 2008; Dickinson, 1980), then the usefulness of awareness as a criterion disappears. If this stance is adopted, it is certainly difficult to see how awareness as a causal determinant of behavior and awareness as an epiphenomenal accompaniment to learning can be empirically distinguished. In any case, the focus on awareness detracts from what might be argued as the more important difference between learning processes – their functional characteristics.

1.4.2 Learning without Thought?

Even if implicit learning is not considered to be equivalent to associative learning, both mechanisms suggest that it is possible for learning to occur incidentally, in the apparent absence of reasoned thought or conscious intention. This carries significant theoretical weight, as it would show that deliberate and effortful mechanisms are not sufficient to explain all behavior, providing evidence for a dual-process theory of learning. However, it is possible that learning that occurs under incidental conditions may in fact be explicit and intentional. It may be the case that participants are naturally suspicious within experimental situations and actively look for regularities and rules when confronted with a task. This could occur at the beginning of a task despite the experimenter devising a distracting cover story or alternative task, or could occur during the task when the participant has become competent and mastered control over their responses. Thus what appears to be incidental learning could be the result of explicit processes. This argument may be harder to defend if there are genuine differences between incidental and intentional learning orientations, or if the task was sufficiently engaging so it would be difficult
to learn explicitly even if the intention to do so was present. Then, the difficult question becomes how to reconcile the obvious role of higher-order cognitive faculties in learning with the apparent lack of need for those same processes. This question will be addressed in the General Discussion in Chapter 5.

1.4.3 Theoretical Questions

The associative/propositional and implicit/explicit debates arise from different contexts and have different motivations for their respective methods of explaining behavior, but they ask similar theoretical questions couched in different terms. Researchers in associative learning ask comparative questions, attempting to connect learning in humans with that of other animals, and delineating what the differences in mental capacities between humans and non-humans are. Implicit learning on the other hand, attempts to understand whether complex phenomena such as language acquisition can be accomplished in the absence of effort and/or intention. What researchers in these seemingly disparate areas of research share in common is a desire to understand situations in which participants learn X or Y, where X and Y are acknowledged to be different in content. The main theoretical issues concern whether a single learning process is sufficient to describe all human behavior, or if multiple processes are needed. If participants can learn in different ways, one driven by System I and the other by System II, then it is to our advantage to discover what situations lead to the dominance of one process over another. This will allow us to predict how the two processes work together and whether there is an optimal process for particular tasks. Alternatively, if a single-process approach is adopted, it is nevertheless important to specify the conditions under which a single mechanism can produce learning that has different qualitative properties.
At the heart of both of these debates are three fundamental questions: 1) do both processes exist, and if so, 2) how can we tell which process is operating, and 3) how do the two processes interact? As discussed, the first and second questions have proved to be highly controversial in both literatures and are currently unresolved. The third question has received much less direct attention in the literature. This may be because specifying the nature of interaction is difficult when it has not yet been decided how best to determine the operation of each learning system. Previous efforts to use awareness have not provided a clear answer, and miss the critical point that what differs between learning processes is not just awareness of the output of learning, but the computation. Experimental approaches which attempt to increase or decrease the likelihood of a given learning process operating are not only more informative in investigating potential differences between learning processes, but also allow for investigation into what most dual-process theories fail to specify – their mode of interaction (Mitchell et al., 2009; Sloman, 1996). Note that failure to specify how associative and propositional processes interact is one of the key criticisms leveled at dual-process theories (e.g. Mitchell et al., 2009). Manipulating verbal instructions to change learning orientation seems to be one way in which to start investigating this potential interaction.

1.5 General Aims

The general aim of this thesis is to investigate the effects of verbal instructions on different learning paradigms in humans. As mentioned above, the aim of this thesis is not to prove the existence of any particular mechanism in humans, nor to conclude in favor of a single- or dual-process account. Whatever approach is adopted, it is undeniable that verbal instructions alter how people learn, and what they learn.
However, if humans are capable of learning associatively and implicitly, testing the effect of learning orientation induced by instructions bears directly on the question of how best to characterize the nature of interaction between associative and propositional processes. Using instructions to induce an intention to learn or to give participants additional knowledge about a task will presumably encourage the use of explicit, propositional processes. A comparison to a condition where participants are not given this additional knowledge therefore provides information on exactly how the propositional system uses such knowledge and applies it to the task. This additional information can either increase what is learned, change what is learned, or have no effect at all. General improvements in learning might suggest that the effect of instructions may simply be to increase the effectiveness of propositional learning processes, which are operating in both conditions. Alternatively, qualitative or selective differences in learning may be indicative that dissociable learning processes exist.

While different instructional manipulations will be used for different paradigms, they are all similar in the sense that they are useful instructions or hints designed to aid the participant through providing explicit knowledge about their task. In other words, they are not misleading and should help participants to learn within each task. In Chapter 2, a hint will be provided where participants are told to look for an underlying rule that summarizes a set of probabilistic contingencies. In Chapter 3, participants will be told to memorize a set of stimuli for an impending memory test. In Chapter 4, participants will receive a hint to attend to a category dimension that will help them form a relational rule to aid them in a difficult discrimination task. While these manipulations are varied, two common themes emerge. The instructions either attempt to induce an intention to learn which would otherwise not be present
(comparing incidental and intentional learning conditions, Chapters 2 and 3), or direct participants to uncover a relational rule that might be difficult otherwise (comparing rule and associative learning, Chapters 2 and 4). In all chapters the comparison will be to conditions where this additional information is not presented (Chapters 2 and 4) or where participants are performing a task where the goal is not related to learning (Chapters 2 and 3).

1.6 Research Questions

There are three research questions examined in this thesis:

1. What effect do the additional instructions have on learning?
2. To what extent does learning occur incidentally, in the absence of these instructions?
3. What are the theoretical implications for human learning?

The first question involves a comparison of learning between the two conditions. If a dual-process approach is taken, this comparison should be informative in exposing the role of explicit processes in learning, if certain assumptions are made. Firstly, a kind of ‘subtraction logic’ is needed whereby implicit or associative processes are assumed to be equally operational in both conditions but explicit or propositional processes are assumed to be more operational in Condition 2. If associative or implicit processes are assumed to be automatic, then one criterion of automaticity is that they function similarly in different situations (Hasher & Zacks, 1979). Of course, this assumption may be misguided. It may be that instructions simply enhance motivation to learn so more attention is given to the task at hand, perhaps resulting in more efficient learning. A comparison between conditions that yields general additive effects would support this possibility. However, if qualitative
differences in learning emerge between the conditions, then this conclusion may be less tenable. In any case, results from Chapters 2-4 will primarily be interpreted as resulting from the interaction between dual-processes, but analogous conclusions can be made for single-process theories, given that they too should be able to account for the sensitivity of learning to orientation and instructions.

The second question alludes to the possibility of observing evidence of associative and/or implicit learning processes under incidental learning conditions when participants have low motivation to use explicit or propositional strategies. This requires the assumption that whatever task participants are engaging in, the contribution of explicit or propositional learning will be minimal. This assumption is problematic because it is impossible to ensure that the influence of a particular process in any given task is zero (see Merikle & Reingold, 1991; Reingold & Merikle, 1988, for a discussion of process-pure tasks). As mentioned above, incidental learning has been cited as a defining feature of implicit learning (Frensch, 1998), and may provide a means to observe associative processes in humans (Mackintosh, 1995), and thus what is learnt in the uninstructed or incidental conditions will be informative not just in comparison to the instructed or intentional conditions, but in testing the limits of associative and/or implicit processes.

1.7 Outline of Chapters

What follows, are a series of experiments in three different paradigms comparing learning in a condition where participants perform a task where they might learn incidentally, against a condition where participants are given additional instruction that gives them the intention to learn or derive a rule. The three paradigms used in this thesis are the serial reaction time (SRT) task (Chapter 2), the prototype
distortion task (Chapter 3), and a dimensional category learning task (Chapter 4).

While the content of learning in each of these paradigms and the manipulations across chapters are markedly different, it is hoped that they will enable general conclusions to be drawn about the nature of human learning.

Chapters 2 and 3 examine learning effects that have been labeled as implicit for various reasons. Sequence learning, as investigated in the SRT task, has been singled out as the best incidental learning paradigm currently available (Destrebecqz, 2004; Destrebecqz & Cleeremans, 2003), since participants are engaged in a task involving speeded responses to targets and do not usually suspect that the sequence of the targets is structured and therefore predictable. Participants usually learn about complex deterministic and probabilistic sequences in these types of tasks in the absence of instructions to do so, and sometimes in the absence of reportable knowledge about the sequence (e.g. Sanchez, Gobel, & Reber, 2010). Chapter 2 will involve a manipulation where one group of participants are given a rule about a probabilistic sequence underlying the contingencies, and another group will not be told anything about the underlying contingencies.

Chapter 3 concerns learning in the prototype distortion task. In this task, participants are able to learn about the similarity structure of a category of stimuli centered on a prototype simply through exposure. These prototype effects have typically been deemed to be implicit due to intact categorization of novel stimuli in amnesic patients in the absence of accurate recognition of seen exemplars (e.g. Knowlton & Squire, 1993). Chapter 3 will test an often-assumed property of prototype effects (that they result under incidental conditions) in a more conservative way than has been done in the past. To create an incidental learning condition, a visual search task will be appropriated from another implicit learning paradigm.
(contextual cueing, Chun & Jiang, 1998) to consume participants’ explicit cognitive functions such that any learning can be more confidently deemed as incidental. In Chapter 3, the ability of participants to generalize to novel stimuli within the ‘category’, and ability to discriminate between new and old stimuli, will be compared between a group who are exposed to the stimuli incidentally while searching through the stimuli, and a group who are told to memorize the stimuli for a subsequent recognition test.

Chapter 4, like Chapter 2, will involve a manipulation where participants are given a hint to discover a relational rule on one of two relevant category dimensions. Participants will be required to first discriminate between two categories of stimuli that are very similar perceptually, but differ on two dimensions (one of which will be manipulated to be attended, and the other unattended). Note that unlike Chapters 2 and 3, instructions will be used here to manipulate attention rather than learning orientation. The experiments in this chapter contained an additional manipulation of rule applicability that substantially influenced the pattern of results. As such, much of the focus in Chapter 4 is on this manipulation. The conclusions within each chapter will be self-contained and will not draw upon the results or conclusions from the previous chapter, however the General Discussion (Chapter 5) will draw parallels between the three paradigms and integrate their separate conclusions into a general theoretical conclusion about the nature of learning in humans.
Chapter 2: Sequence Learning

2.1 Introduction

Implicit learning theorists assume that individuals learn about their surroundings incidentally, in a manner that does not easily give rise to verbalizable propositions, but rather a form of tacit knowledge that improves their ability to engage with their environment and anticipate events (e.g. Reber, 1989). It is generally accepted that explicit learning requires higher-order cognitive processes such as reasoning and hypothesis-testing, and results in knowledge that is declarative (Anderson, 1976) and/or propositional (Mitchell, De Houwer, & Lovibond, 2009). In contrast to explicit learning, definitions of implicit learning have been much more varied. Frensch (1998) has noted that different researchers often use the word ‘implicit’ synonymously with either unconscious/unaware (e.g. Reber, 1989, 1993; Lewicki, Czyzewska, & Hoffman, 1987; Shanks & St. John, 1994), or non-intentional/automatic (e.g. Cleeremans & Jiménez, 1998; Jiménez & Méndez, 1999). Much of the early implicit learning studies focused on the former definition, with results from initial studies alluding to the possibility that learning was possible despite some participants being unable to verbalize what they had learned (e.g. Reber, 1967; Willingham, Nissen, & Bullemer, 1987). These observations sparked a wealth of research into implicit learning as “learning without awareness” (Reber, 1987), and fierce debate regarding the validity of different awareness measures (Perruchet & Amorim, 1992; Shanks & St. John, 1994). The role of awareness in implicit learning still remains a point of contention today (e.g. Frensch & Rünger, 2003; Shanks & St. John, 1994).
While awareness does carry significant theoretical weight in differentiating implicit from explicit learning, experiments that simply test for the presence of awareness as proof of implicit learning miss a crucial point— that awareness may directly influence learning itself, but may also emerge epiphenomenally with learning (Sanchez & Reber, 2013). It has been argued that what is more important theoretically in differentiating implicit and explicit learning is not the nature of the resulting knowledge, but the nature of the learning process itself (Frensch, Lin, & Buchner, 1998). Frensch (1998) concluded that conceiving of implicit learning in terms of the automaticity of the learning process (rather than awareness) was the most scientifically useful since it leads to a number of testable predictions which can provide a more meaningful contrast with explicit learning. According to this account of implicit learning, its defining characteristics are that it occurs incidentally without the intention to learn (Lewicki, 1986; Reber, 1989, 1993), and therefore requires little mental effort (Frensch, 1998; Jiménez & Méndez, 1999). This definition accords with the emphasis that many implicit learning researchers place on the automaticity of implicit learning (e.g. Reber, 1989, Berry & Dienes, 1993; Frensch, 1998; Perruchet & Gallego, 1997; Underwood & Bright, 1996), as well as on the independence of implicit learning from the explicit thoughts and conscious motivations of the individual (Berry & Dienes, 1993; Destrebecqz, 2004; Jiménez, Méndez, & Cleeremans, 1996; Perruchet & Vinter, 1998; Song, Howard Jr., & Howard, 2007; Willingham, Nissen, & Bullemer, 1989).

In this regard, one of the best examples of implicit learning is sequence learning as investigated in the serial reaction time (SRT) task. In a typical paradigm, a target appears in different locations on the screen and participants have to respond to the location of the target with a corresponding key press. Unbeknownst to participants
performing the task, the location of the target follows a deterministic or probabilistic sequence, such that some or all of the responses can be predicted. Learning is evident when reaction times are faster on trials that follow the sequence, compared to trials that do not. Sequence learning, as measured in this task, is a robust finding and occurs despite participants having no intention to learn about an underlying sequence, and poor ability to verbalize their knowledge on more direct tests (e.g. Jiménez, Méndez, & Cleeremans, 1996; Willingham, Nissen, & Bullemer, 1989). The SRT paradigm has been cited as the best task to investigate implicit learning in humans since responding is speeded so participants have little opportunity to think before they respond, and learning is truly incidental since the task is easy to perform and participants normally do not suspect an underlying sequence (Cleeremans, 1993; Cleeremans & Jiménez, 1998; Destrebecqz & Cleeremans, 2003).

Thus, sequence learning seems to embody some of the characteristics of implicit learning. However, it has been argued that stronger support for implicit learning as a separate learning process would be provided if it could be shown that explicit knowledge or intention to learn do not affect SRT performance (Jiménez, Méndez, and Cleeremans, 1996; Reber, 1989), or alternatively, produce dissociable patterns of learning (Jones & McLaren, 2009; Stadler & Frensch, 1994). The former result, that learning is not amenable to conscious control, would show that learning is not cognitively penetrable. This can be taken as evidence that implicit learning operates automatically but also independently from explicit learning, meaning that one process cannot affect or share information with the other (Curran & Keele, 1993; Lewicki, 1986; Stadler & Frensch, 1994). The latter result, where explicit and implicit orientations produce dissociable patterns of learning would suggest that while implicit

---

8 This can be manipulated either within- or between-subjects.
learning is cognitively penetrable, manipulating intention to learn engages a separate explicit learning process that operates in a qualitatively different way (Jones & McLaren, 2009; Stadler & Frensch, 1994). In either case, the assumption is that explicit processes can be engaged at will and thus encouraging these processes in one group but not another might reveal the contribution of explicit processes to so-called implicit learning effects.

Recent work has therefore focused on demonstrating that implicit learning is in some sense automatic or operates independently of voluntary control by manipulating explicit knowledge and/or the intention to learn. For example, Curran and Keele (1993, Experiment 1) compared learning of a simple six-item repeating sequence between an incidental group who were not told anything about a sequence, and an intentional group who were explicitly taught the sequence to which they would be responding. When participants in the incidental group were divided post-hoc into ‘more aware’ or ‘less aware’ based on whether they could produce at least four of the six sequence positions in a questionnaire, the ‘less aware’ participants showed less learning in the training phase of the SRT task than the ‘more aware’ and intentional groups. A similar result was reported by Frensch and Miner (Experiment 1, 1994), who tested participants on a 12-item fixed, repeating sequence interspersed periodically with random sequences in a four-choice SRT task. The intentional group were told that the sequence of the target followed a repeating pattern, while the incidental group were simply told that they needed to respond as quickly and accurately as possible. Again, the intentional group showed a larger difference between their RTs for sequential and random sequences (see also Destrebecqz, 2004).

9 Other methods to demonstrate automaticity have involved showing that learning is impervious to secondary tasks (e.g. Jiménez & Méndez, 1999; Jiménez & Vázquez, 2005), or the presence of explicit cues (Jiménez & Méndez, 2001).

10 Interestingly, these differences did not transfer to a phase where a secondary task was added.
These results suggest that explicit knowledge and intention to learn aid performance in tasks that are assumed to measure implicit learning, suggesting that such knowledge enhances learning, at least when deterministic sequences are used (but see Sanchez & Reber, 2013, for an exception with a modified task). Probabilistic sequences, on the other hand, are less amenable to such explicit strategies because any attempts to induce rules are often hindered by elements of the sequence that do not conform to the rule (Cleeremans & Jiménez, 1998). Under these circumstances, participants may either give up searching for a rule, or their attempts to do so may not benefit their RT performance, producing no difference between incidental and intentional learning. Thus, it is possible that this advantage for intentional learners may be confined to deterministic sequences like those used by Curran and Keele (1993) and Frensch and Miner (1994). If this were the case then one might question the relevance of these findings given that sequence learning has been found in many probabilistic tasks, and probabilistic sequences arguably provide a better laboratory model of the imperfect contingencies between events that we experience in most environments (Deroost, Zeeuws, & Soetens, 2006).

Jiménez, Méndez, and Cleeremans (1996) provide support for the hypothesis that the learning of probabilistic sequences might be resistant to the explicit intentions of the individual. They manipulated learning orientation by telling participants allocated to the intentional group that a set of rules determined where the stimulus would appear, and that they should try to work out the rules for a subsequent generation test where they could earn extra payment for accurate performance. Both groups were trained on a six-choice SRT task where on 85% of the trials, the next target location could be predicted, either with 50% accuracy (i.e. from one of two possible locations), or with 100% accuracy. On the other 15% of trials, the location
was made unpredictable by substituting the target location for another random location. Despite the intentional group being more accurate and slower in their responding (suggesting they were trying to work out the sequence), the two groups showed equivalent evidence of learning in responding in the SRT task and a continuous generation task where participants had to predict the location of the next target with feedback on where the ‘correct’ (grammatical or random) target was.

A similar result was obtained by Stefaniak et al. (2008) who looked at the effect of intention to learn on probabilistic and deterministic sequences. In Experiment 1, participants in the intentional group first responded to an 8-item deterministic sequence, and then had to generate the sequence themselves until they could produce the sequence in its entirety twice in a row. Those in the incidental group began the training phase without the chance to learn about the sequence beforehand. Reaction times for the intentional group during the training phase were significantly faster than those for the incidental group during the initial 12 sequential blocks, and the difference in RT for the trained sequence in block 12 (the last sequential block) and another, untrained sequence introduced in block 13 was larger in the intentional group. Experiment 2 investigated the same group manipulation using a sequence with half of its positions deterministic and the other half probabilistic, which amounted to an improbable (i.e. infrequent) position occurring on 10% of the trials. The comparison of RTs for the training sequence in block 12 versus an untrained sequence in block 13 showed that both incidental and intentional groups learned about the sequence equally well, such that the increase in RT for the untrained block was the same between groups. Stefaniak et al. concluded that although participants possessed relevant sequence knowledge (based on better performance

11 This stands in contrast to the effect of intentional instructions in deterministic sequences, where the intentional group are usually faster at responding than the incidental group (e.g. Destrebecqz, 2004).
than the incidental group in a subsequent generation test), it could not be used effectively due to the probabilistic nature of the sequence, and therefore any learning observed must be due to implicit learning mechanisms (see also Cleeremans & Jiménez, 1998, and Song, Howard Jr., & Howard, 2007). However, Stefaniak et al. note that there was no difference between RTs for improbable, compared to probable positions, suggesting that participants did not learn these contingencies, making these results difficult to interpret.

In summary, the finding that intentional learning produces no advantage in sequence learning over incidental learning (Cleeremans & Jiménez, 1998; Jiménez, Méndez, and Cleeremans, 1996; Stefaniak et al., 2008) has been used as evidence in favor of an independent, implicit learning mechanism that operates effectively in situations where explicit mechanisms are hindered (e.g. Cleeremans & Jiménez, 1998). This argument is based on the assumption that the difference between incidental and intentional conditions is the addition of explicit processes\(^{12}\), and therefore a lack of difference between these learning orientations implies that explicit mechanisms are not useful and therefore not responsible for any learning effects observed under incidental conditions. Thus, one way to show that implicit learning exists is to demonstrate that there are some conditions under which explicit processes do not aid sequence learning.

Rather than searching for this particular result as a proof of existence, what may be more informative for characterizing implicit learning at a process level is distinguishing the situations in which explicit knowledge is helpful for sequence learning from the situations in which it is not. Stefaniak et al. (2008) have proposed that whether the sequence is deterministic or probabilistic may determine whether

---

\(^{12}\) Note that this does not assume that explicit processes are absent under incidental conditions.
explicit knowledge is helpful or not, while other authors have suggested that the
difficulty of the sequence is the determining factor (Destrebecqz, 2004; Stadler &
Frensch, 1994). Both these suggestions adequately account for the findings reviewed
above regarding the effects of explicit knowledge (but see Sanchez & Reber, 2013).

However, some authors have suggested that more convincing evidence for
separable implicit and explicit sequence learning processes would be provided if
incidental and intentional learning orientations produced qualitatively different results
provide such a demonstration in a 2-choice (left or right) SRT task by comparing
reaction times between a control group trained on a pseudorandom sequence and an
experimental group trained on a probabilistic sequence. The sequence could be
summarized as: “on two-thirds of the trials, if the previous two trials were different
(e.g. XY or YX), then the next trial will be Y, if the previous two trials were the same
(e.g. XX or YY), then the next trial will be an X”. The test phase was a pseudorandom
sequence that did not follow this rule and was the same for both groups, enabling any
group differences to be attributed to sequence learning and not sequential effects
intrinsic to the task.

Learning on different subsequences was examined by dividing up trials into
the 4 possible subsequences based on trial history (XX, YY, XY, YX) and subtracting
the RT for the consistent trials (e.g. XXX, YYY) from the RT for the inconsistent
trials (e.g. XXY, YYY), producing a difference score. Experiment 1 tested a control
and experimental group under incidental learning conditions. Both groups were told
that the purpose of the task was to respond as quickly and accurately as possible to
maximize their earnings. The experimental group showed significantly larger
difference scores than the control group on subsequences that ended with an
alternation (YYX and YXY), but there were no group differences in learning subsequences that ended in a repetition (XXX and XYY). Interestingly, in Experiment 3 where participants in the experimental group were told that they could use the sequence history to predict the location of the target, this pattern of learning reversed, with the best learning occurring for XXX and XYY, and no group differences for the sequences that ended in an alternation (YYX, YXY). Jones and McLaren argue that their study offers evidence of dissociable learning systems that are based on associative-, and rule-based learning processes.

The results of Jones and McLaren (2009) contradict the claim made by Stefaniak et al. (2008) and Cleeremans and Jiménez (1998) that there are no differences in SRT performance between incidental and intentional learning conditions with a probabilistic sequence. One noteworthy difference between the sequence in Jones and McLaren (2009) and those of other studies is the fact that their sequence can be summarized with an abstract relational rule, that is, one where the next response can be predicted (at least probabilistically) from a comparison of the previous two responses ("same" vs. "different") and this rule applies regardless of the exact physical properties of the last response cue. This rule is distinct from the type of explicit knowledge participants may acquire about the individual subsequences. For instance, knowing that Left-Left-Left is more likely to occur than Left-Left-Right does not require abstract knowledge of the rule that connects the first two responses with the third. However, even though the sequence could be summarized using a probabilistic rule, the rule itself was still relatively difficult and it is not clear whether participants improved because they acquired this rule. Jones and McLaren suggest that those in the intentional condition learned best about the XXX subsequence purely because the repetition of target location was highly salient. Thus it is clear that the
intention to learn in this probabilistic SRT task affected learning, but this may not reflect explicit knowledge of the rule itself. The effect of actually giving participants knowledge in this abstract rule-based form has not been explored in any depth in the sequence learning literature.

A plausible explanation for the results to date is that additional explicit knowledge or intention to learn potentially affects sequence learning in any form, whether it be deterministic or probabilistic, but the impact of this knowledge is a function of how easy it is to discover the relevant rules governing the transitions, as well as how easy it is to apply this knowledge to the task. In this regard, an obvious reason why explicit knowledge did not improve probabilistic sequence learning in the studies by Stefaniak et al. (2008) and Jiménez, Méndez, and Cleeremans (1996) is the complexity of the sequences themselves. One can assume that participants required considerable cognitive resources to discover, retain and apply the multiple rules describing the sequence. The rule used in Jones and McLaren’s study also required maintenance of both the rule and the previous two trials in memory for successful prediction. Since the SRT task is speeded, simplifying the rule might increase the likelihood of its use. To date, there are no studies that examine the application of an explicit abstract rule where the rule itself is just as simple or easier to apply than learning the individual contingencies between the responses in the sequence.

The current study achieved this by devising a set of response contingencies that can be explicitly described in very simple relational terms. The value of acquiring explicit knowledge is especially obvious when a complex set of contingencies between events or properties is encapsulated by a simple relational concept. Examples abound from other learning paradigms where discovering a relational rule or abstract structure is associated with improved performance and more flexible transfer of
learning to new problems (e.g. Chi, Feltovich, & Glaser, 1981; Gick & Holyoak, 1983; Harris & Livesey, 2008; Livesey & McLaren, 2009; Shanks & Darby, 1998). In sequence learning, especially simple versions of the SRT task where the spatial location on screen dictates the appropriate response, one of the most salient relations is the direction of motion of the target, that is, the position of the current target relative to the last. This is a simple property to describe and one that can be easily manipulated such that there is a prevailing direction of motion with which the current transition may be consistent or inconsistent.

The aim of the following experiments was to investigate to what extent an explicit hint about a relational rule describing a set of probabilistic contingencies could impact on performance in a three-choice SRT task. As discussed, one conceptualization of implicit learning is that it is independent of explicit knowledge. While the experiments attempting to demonstrate learning in the absence of explicit knowledge have produced mixed results and suffer from methodological problems (Shanks & St. John, 1994), testing the effect of explicit knowledge on sequence learning may be a more useful method of investigating the relationship between learning and explicit knowledge. This manipulation was similar to intentional learning conditions reviewed above except that the hint was designed to guide participants towards explicit relational knowledge that captured the probabilistic contingencies of the entire sequence in a very simple description. A SRT task (previously reported in Lee & Livesey, 2013, and Lee, Beesley, & Livesey, 2016) with three response locations (left, top and right of the screen) was used, with a prevailing direction of motion randomly chosen for each participant (either clockwise or anticlockwise). The contingencies were arranged such that 75% of the time, the target would appear to be moving in one direction (e.g. clockwise) and on the other 25% of trials the target
would move in the opposite direction (e.g. anticlockwise). For example, if the target appeared at the top position, there was a 75% chance that it would next appear at the right position, and a 25% chance that it would next appear at the left position (see Figure 2.1). Since the target could not appear in the same position twice in a row, trials could be classified as either cued or miscued: where the location of the target can either be consistent with the particular direction of motion (curved, bold lines in Figure 2.1), or inconsistent with the direction of motion (straight, dotted lines in Figure 2.1) respectively.

*Figure 2.1.* Schematic diagram of the three-choice SRT task. The target could appear in either the left, top, or right position on screen (dotted circles, not seen in actual experiment) and participants had to respond by pressing the corresponding arrow key. The target could never appear in the same location twice in a row, which meant that the target would transition in a clockwise or anticlockwise direction on each trial (i.e. if the target appeared at the top, the next target location would either be left or right). In Experiment 1, the target direction was randomly determined so that there was a 50% chance of transitioning clockwise or anticlockwise on each trial. In Experiment 2 onwards, there was a cued direction of motion whereby the target would travel in a predominant direction 75% of the time (in this example, clockwise, as represented by the bold, curved lines), and in the miscued direction 25% of the time (anticlockwise, as represented by the straight, dotted lines).
Experiments 1 and 2 were first performed to determine the pattern of sequential effects present in this task since arranging three target locations around the screen had not been implemented in an SRT task before, and controlling for sequential effects (differences in responding that occur due to recent history of trials and not the presence of contingencies) has been noted as important in accurately measuring sequence learning (Jones, Curran, Mozer, & Wilder, 2013). Experiment 1 investigated what pattern of sequential effects would be shown in the absence of contingencies, while Experiment 2 tested whether the same pattern of sequential effects would result when probabilistic contingencies were added to the task. Experiments 3-5 compared learning between two groups: one given an explicit hint about an underlying rule that described the contingencies (Hint group), and another group not given any additional information (No Hint group). Experiments 4-6 also included an additional transfer phase after the training phase, to determine whether the hint affected the persistence of the cueing effect when the contingencies were removed and aspects of the task were changed. Finally, Experiment 6 clarified a discrepancy in results between Experiments 3-4 and Experiment 5 by comparing two different modes of hint delivery.

2.2 Experiment 1

Before attempting to measure sequence learning in a novel task, it is important to investigate the pattern of sequential effects that may arise, as they may obscure or inflate any evidence of learning (Jones et al., 2013). Sequential effects are transient differences in performance that arise as a function of trial history, and have been studied most extensively in SRT procedures, the same procedure used to investigate sequence learning. When the task is entirely unstructured, such that there is no
consistent sequence to the target’s movement between positions (and the target’s location cannot be predicted), participants are nevertheless faster to respond on certain trials. These sequential effects suggest that in the absence of any predictive information, participants’ responses are still influenced by recent events. In sequence learning experiments, sequential effects are often regarded as variance to be controlled for or minimized on test (Anastasopoulou & Harvey, 1999; Jones et al., 2013). The methods that researchers have employed to this end include devising an appropriate sequence to minimize sequential effects (e.g. avoiding first-order repetitions, Cleeremans & McClelland, 1991), or using a control group who are trained with a pseudorandom sequence containing no contingencies and tested with a trial order that would produce equivalent sequential effects to an experimental group (e.g. Jones & McLaren, 2009).

Sequential effects have been studied most extensively in two-choice SRT tasks, where the possible number of events (e.g. left and right) and transitions (repetitions and alternations of target location) is constrained. In a two-choice RT task (e.g. left and right responses) where the appearance of the target is randomly determined, participants are usually fastest to respond on trials where either repetitions or alternations of target location have occurred consecutively (e.g. Bertelson, 1961; Cho et al., 2002). This means that if a target had just appeared on the left 3 times, participants are usually faster to respond left than they are to respond right (i.e. LLLL would be faster than LLLR). Conversely, if participants have just experienced a series of alternations (left, right, left), they are faster at responding right than left (i.e. LRLR is faster than LRLL), but this facilitation is usually observed to be weaker than the equivalent effect for repetitions (e.g. Bertelson, 1961; Cho et al., 2002; Remington, 1969). These patterns of sequential effects have been attributed to
participants’ subjective expectancies (Soetens, Boer, & Hueting, 1985), which in this context refer to the predictions generated by some internal learning process. However, it is worth noting that these expectancies have been shown to be independent of the individual’s explicit beliefs about impending events: recent work that has directly compared trends in choice RT and trends in explicit expectancy for the relevant events has found them to be widely divergent (Barrett & Livesey, 2010; Livesey & Costa, 2014; Lee Cheong Lem, Harris & Livesey, 2015, see also Hale, 1967, and Hyman, 1953, for earlier informal observations of similar trends).

While the presence of sequential effects is well established in two-choice SRT tasks where the target can repeat its location, it is unknown what pattern of sequential effects will result in the current three-choice SRT task since tasks with three targets are rarely employed (but see Gökaydin, Ma-Wyatt, Navarro, & Perfors, 2011)\textsuperscript{13}. In the current SRT task, three target locations were arranged on the edges of a computer screen (e.g. left-top-right) and repetitions of target location (e.g. top-top) were prohibited. This meant that the sequential effects concerned the repetition and alternation of the direction of target transitions, rather than target location (see Figure 2.1). Using this paradigm, sequential effects were assessed by allowing an equal probability of clockwise or anticlockwise transitions such that there was no dominant direction of motion.

The aim of Experiment 1 was to explore the sequential effects present in a three-choice SRT task that contained no response repetitions, such that all sequential effects would be based on sequences of transitions between target locations. Since neither direction of motion prevailed consistently, for any given target location, the other two

\textsuperscript{13} In Experiment 3 of Gökaydin et al. (2011), participants responded to one of three geometric shapes that appeared in the same location. Participants pressed the corresponding button with one finger, returning their finger to a central position at the start of each trial.
target positions were equally likely to follow and therefore the target always moved in either a clockwise or anticlockwise direction. However, taking any three consecutive responses, the direction of motion itself would repeat if the three different target locations were shown consecutively in any order (in abstract terms, responses X, Y, then Z), whereas the direction of motion would reverse or alternate if the last response was the same as that occurring two presentations prior (that is, Z, Y, then Z again). A difference between these trial types can be thought of as second-order sequential effects (i.e. XYZ vs. ZYZ). A similar logic was applied to examine third-order sequential effects, that is, RT on the last of four consecutive responses that constitute three directional transitions, and fourth-order sequential effects, that is, RT on the last of five consecutive responses that constitute four directional transitions (see Figure 2.2 and Table 2.1). A response-stimulus interval (RSI) of 500ms was chosen because this delay between responses should be long enough to avoid response priming effects that dramatically alter sequential effects with short RSIs (< 200ms) (e.g. Vervaeck & Boer, 1980) but short enough for participants to retain a sense of directional transition from one target location to another.

Table 2.1.

<table>
<thead>
<tr>
<th>Fourth Order</th>
<th>Third Order</th>
<th>Second Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSS YZXYZ (RRR)</td>
<td>SS ZXYZ (RR)</td>
<td>S XYZ (R)</td>
</tr>
<tr>
<td>DSS XZXYZ (ARR)</td>
<td>DS YXYZ (AR)</td>
<td></td>
</tr>
<tr>
<td>DDS ZYXYZ (RAR)</td>
<td>DD XZY (RA)</td>
<td>D ZY (A)</td>
</tr>
<tr>
<td>SDS XYXYZ (AAR)</td>
<td>SD ZY (AA)</td>
<td></td>
</tr>
<tr>
<td>DDD YXZYZ (RRA)</td>
<td>SSD XYZYZ (RAA)</td>
<td></td>
</tr>
<tr>
<td>SDD ZXZYZ (ARA)</td>
<td>DSD ZYZYZ (AAA)</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* The subsequences read from left (past trials) to right (current trial). X, Y and Z represent any one of the 3 target locations left, top, and right. R and A represent whether each subsequence consists of a repetition or alternation of direction, referenced from the previous trial. S and D represent whether the nth-order transition is the same (S) or different (D) direction to the first-order transition (YZ).
Figure 2.2. Example of how a fourth-order subsequence was coded as a series of target locations (XYXYZ), a series of movements with reference to the direction of the first-order transition (SDS), and a series of directional repetitions and alternations (AAR). X, Y, and Z can stand for any of the three target locations (left, top, right), with subsequences reading from left (past trials) to right (current trial, n), and therefore direction 1 and direction 2 can represent either clockwise or anticlockwise. Subsequences were entered into the ANOVA based on whether transitions at the nth level were in the same direction (S) or different direction (D) to the first-order transition (YZ, direction 1 in this example). Subsequences can also be conceived of as a series of repetitions (R) and alternations (A) of target direction, which are referenced from the direction of movement on the previous trial.

2.2.1 Method

2.2.1.1 Participants and Apparatus

Fifteen participants (11 female, $M$ age = 26.87, $SD = 7.90$) who were either first year Psychology students at the University of Sydney or respondents to an online
advertisement took part in the experiment\textsuperscript{14}. Students received course credit and respondents received payment (AUD$15/hour) for their participation. All experiments were programmed using Psychophysics Toolbox for Matlab (Brainard, 1997; Pelli, 1997) and run on Apple Mac Mini desktop computers connected to 17 inch CRT monitors, refreshed at a rate of 85 Hz. A standard Apple keyboard and mouse were used, and testing was conducted in individual cubicles in groups of up to six. Participants in this, and all subsequent experiments, gave informed consent and the study was approved by the University of Sydney Human Research Ethics Committee.

2.2.1.2 Procedure

Participants were told that the purpose of the task was to respond as quickly and as accurately as possible to a target (a magenta circle) that would appear in one of three positions on the screen. Participants had to press the ‘left’ arrow key if the target appeared on the left, the ‘up’ arrow key if the target appeared at the top, and the ‘right’ arrow key if the target appeared on the right of the screen. Participants were not told that the target could not appear in the same location twice in a row, and were not explicitly encouraged to attend to the movement of the target. Participants were asked to use their non-dominant hand to respond during training. If participants used their left hand, they placed their index finger on the right arrow key, their middle finger on the up arrow key, and their ring finger on the left arrow key. The target stayed on screen until a response (correct or incorrect) was made and after a blank RSI of 500ms, the next target appeared. After a short practice phase (48 trials), participants completed 720 trials where the location of the target had an equal chance.

\textsuperscript{14} Target sample sizes for experiments in this chapter were 24 per group for experiments testing the hint manipulation and 15 otherwise. These target sample sizes were based on initial pilot experiments aimed at testing the general paradigm. The final sample size was affected by availability of participants.
of moving clockwise or anticlockwise and thus its location could not be predicted. Trials were randomized in blocks of 12, maintaining the 50/50 ratio of clockwise and anticlockwise transitions within each block. The experiment was completed in one continuous block without a break and lasted for approximately fifteen minutes.

**2.2.2 Results and Discussion**

All subsequent RT analyses refer to mean RTs for correct responses excluding any greater than one second and Greenhouse-Geisser corrections were performed for violations of sphericity. Participants took on average 337ms ($SD = 40.9$) to respond with 96% ($SD = 2.1$) accuracy. Henceforth X, Y, and Z will be used to describe the various subsequences with X, Y and Z representing any one of the 3 positions left, top, right (see Figure 2.1). This coding of subsequences is designed to capture the sequence of target locations, but not the direction of movement. Thus, an XYZ subsequence could equally stand for either a left-top-right or left-right-top sequence. Reaction times and error data always represent performance on the final trial of each subsequence (Z). Trials were divided into subsequence type at the second-, third-, and fourth-order level (see Table 2.1). Within each level (n), trials were classified according to whether the transition at the nth level was the same (S) or different (D) from the direction of motion of the first-order transition (Y to Z, see Figure 2.2 for an example of how a fourth-order subsequence was coded in this way), as well as whether the subsequence contained a series of alternations (A) or repetitions (R) of target direction. This yielded 2 different subsequences at the second-order level (R, A), 4 subsequences at the third-order level (RR, RA, AR, RR), and 8 subsequences at the fourth-order level (RRR, RRA, RAR, RAA, ARR, ARA, AAR, AAA).
Figure 2.3 displays mean RTs and errors for the eight fourth-order subsequences, split according to whether the fourth-order transition was the same (left side of the figures) or different (right side of the figures) to the first-order transition. It is firstly apparent that the overall pattern of sequential effects is very similar for the RTs (top panels) and errors (bottom panels), and that there are very large differences between the two second-order subsequences (XYZ and ZYZ shown as separate lines), with performance on subsequences with a final repetition of motion (XYZ) faster than subsequences with a final alternation of motion (ZYZ). Within XYZ and ZYZ subsequences, the recent history of alternations and repetitions seemed to further impact performance.

To examine the pattern of sequential effects, an (2 x 2 x 2) analysis of variance (ANOVA) with fourth-, third-, and second-order transition as within-subjects factors was performed on mean RTs and errors for the fourth-order subsequences. Note that the subsequences were entered into the ANOVA coded according to whether the nth transition was in the same or different direction to the first-order transition (see Table 2.1). For the RT data, there was a main effect of fourth-order, $F(1,14) = 6.40, p = .024, \eta^2_p = .314$, and third-order, $F(1,14) = 5.31, p = .037, \eta^2_p = .275$, and a very large main effect of second-order, $F(1,14) = 103.3, p < .001, \eta^2_p = .881$. In the error data, there was also a main effect for second-order, $F(1,14) = 17.11, p = .001, \eta^2_p = .550$, but no significant main effect of third- or fourth-order, largest $F(1,14) = 2.68, p = .124, \eta^2_p = 161$. 
Figure 2.3. RTs (a) and proportion of errors (b) for each fourth-order subsequence (coded as a series of repetitions (R) and alternations (A) of target direction) in Experiment 1. Subsequences that end with an alternation are shown on the top and subsequences that end with a repetition are shown on the bottom. Subsequences are divided according to whether transitions at the second-order (XYZ vs. ZYZ, shown as separate lines), were the same or different direction to the first-order transition (YZ). Within each pair of connected data points, the left point has the same third-order transition and the right has a different third-order transition to the direction of the first-order transition (YZ). The pairs of connected data points on the left side of the figure have the same fourth-order transition, and the pairs on the right side of the figure have a different fourth-order transition to the first-order transition (YZ).

It is clear that the strongest main effect in both RT and errors was at the second-order level, specifically comparing the XYZ subsequences (RT: $M = 317$ms, $SD = 39.5$, errors: $M = .023$, $SD = .013$) to the ZYZ subsequences (RT: $M = 363$ms, $SD = 44.1$, errors: $M = .061$, $SD = .035$). Participants were on average 46ms faster and also made on average 3.8% fewer errors when the target travelled in a consistent
direction on 2 consecutive transitions (those ending in XYZ), compared to when the
target appeared to alternate directions (those ending in ZYZ). The main effects of
third- and fourth-order in the RT data show that participants were 10ms faster to
respond when the third-order transition was the same direction as the first
(subsequences of the form S_ were faster than D_ ), but 4ms slower to respond when
the fourth-order transition was the same direction as the first (subsequences of the
form S__ were slower than D__). Note however, that these main effects are qualified
by the significant interactions discussed below.

In RTs, there was a significant third-order x second-order interaction, $F(1,14) = 7.12, p = .018$, $\eta_p^2 = .337$, and significant fourth-order x second-order interaction,
$F(1,14) = 5.31, p = .037$, $\eta_p^2 = .275$. A significant 3-way interaction between fourth-, third-, and second-order factors was found in both RTs, $F(1,14) = 31.34, p < .001$, $\eta_p^2 = .691$, and accuracy, $F(1,14) = 16.13, p = .001$, $\eta_p^2 = .535$ (see Figure 2.3). The
easiest way to interpret the three-way interaction is by conceptualizing the
subsequences as a series of directional repetitions and alternations (see Table 2.1). If
we examine the 4 fourth-order subsequences where the fourth-order transition was
consistent with the first-order transition (left side of Figure 2.3), it is clear that RT and
accuracy were influenced primarily by second-order differences. That is, whether the
last trial in the subsequence contained a repetition (R) or alternation (A). However,
within the XYZ subsequences, responding was facilitated when the subsequence
contained several repetitions in a row (RRR was easier to respond to than AAR), and
within the ZYZ subsequences, responding was facilitated when the subsequence
contained a repetition before the last alternation (ARA), compared to when there were
2 alternations to respond after (RAA).
This pattern seems to reverse for the examination of those subsequences where the fourth-order transition is inconsistent with the first-order transition (right side of Figure 2.3). For the XYZ subsequences, there seemed to be little difference between whether the third-order transition contained a repetition (ARR) or alternation (RAR), as responding seemed to be generally facilitated by the repetition of a direction of motion on the last trial of the subsequence. On the other hand, for the ZYZ subsequences, responding was both faster and more accurate when the subsequence contained a series of alternations in a row (AAA) than when it contained a series of repetitions and then a final alternation (RRA). The pattern of data in Experiment 1 can be summarized in the following way: general facilitation in responding occurred when the target moved in the same direction a few times in a row (i.e. there was a repetition of a direction of motion), and responding was hindered when the direction alternated, except when the direction alternated several times (i.e. ZYZYZ).

It is clear from this experiment that higher-order sequential effects exist in this task, and while some interactions between second-, third- and fourth-order levels of subsequences were significant in this experiment, by far the most substantial difference was at the second-order level between the XYZ and ZYZ subsequences. The biggest determinant for whether responding in this task was facilitated was whether the previous direction of motion was consistent with the current direction of motion. It appears that sequences of trials in which the target changed direction were particularly difficult to respond to, leading to slower responses than when the target moved in a consistent direction on two consecutive trials. Note that the fastest ZYZ subsequence (ZYZYZ) was still numerically slower than the slowest XYZ subsequence (XYXYZ, see Figure 2.3). These sequential effects informed the choice
of analyses to follow in later experiments, but first it is necessary to examine whether they are perturbed by the introduction of probabilistic contingencies required for sequence learning.

### 2.3 Experiment 2

The aim of Experiment 2 was to explore whether the pattern of sequential effects in Experiment 1 could also be found when the probabilistic contingencies were added to the task. In Experiment 2, the direction of motion was biased in a predominant direction such that the target would appear to be moving in one direction 75% of the time during training (see Figure 2.1), and a transfer phase was added where these contingencies were removed and the target direction had equal probability of transitioning clockwise or anticlockwise. This transfer phase tests whether participants have learned the contingencies, with learning evident (i.e. a cueing effect present) if participants are faster to respond on trials where the target moved in the previously cued direction of motion (cued trials) than the previously miscued direction (miscued trials). This transfer phase also enables assessment of sequential effects under conditions that are similar to Experiment 1, where no contingencies were present. Thus for Experiment 2 the focus was on the data from the transfer phase (see Appendix A for the results from the training phase).

#### 2.3.1 Method

##### 2.3.1.1 Participants

All fifteen participants (9 female, $M_{age} = 25.47, SD = 7.22$) in Experiment 2 were respondents to an online advertisement and were paid AUD $15/hour for their participation.
2.3.1.2 Procedure

The procedure was similar to Experiment 1 except for the following changes. After completing a short practice phase (48 trials) with no contingencies, participants responded to 720 trials where the target moved in a prevailing direction of motion on 75% of trials (which was randomly chosen to be clockwise or anticlockwise for each participant) and 360 trials where there were no contingencies (there was no prevailing direction of motion). For the initial 720 trials with prevailing direction of motion, trials were randomized in blocks of 12 trials maintaining the 75% cued and 25% miscued ratio of contingencies within each block. Participants continued to use the same response keys and hand to respond and there was no break between the training and transfer phase, such that there was nothing to mark the transition into a separate phase for participants. The instructions given to participants were exactly the same as Experiment 1, meaning that participants were not informed that there was a bias in the direction of motion.

2.3.2 Results

The data were analyzed in a similar way to Experiment 1, with cueing added as a within-subjects factor. A (2 x 2 x 2 x 2) within-subjects ANOVA with cueing, fourth-, third-, and second-order as factors was run on RTs (Figure 2.4a) and errors (Figure 2.4b) for the transfer phase. For the RTs in the transfer phase, there was a significant main effect of cueing, $F(1,14) = 45.96, p < .001, \eta_p^2 = .767$, with faster responses for cued trials indicating that participants learned about the cued direction of motion. As with Experiment 1, there was also a main effect of second-order, $F(1,14) = 42.02, p < .001, \eta_p^2 = .750$, and a significant interaction between fourth-,
third-, and second-order, $F(1,14) = 49.31, p < .001, \eta_p^2 = .779$.

**Figure 2.4.** RTs (a) and proportion of errors (b) for cued and miscued trials for each fourth-order subsequence (coded as a series of repetitions (R) and alternations (A) of target direction) in the transfer phase of Experiment 2. Subsequences that end with an alternation are shown on the top and subsequences that end with a repetition are shown on the bottom. Subsequences are divided according to whether transitions at the second-order (XYZ vs. ZYZ, shown as separate lines), third-order (left vs. right points connected by lines), and fourth-order (left vs. right side of the figures) level were the same or different direction to the first-order transition (YZ). Cued and miscued trials are shown as separate lines.

In errors, there was a main effect of second-order, $F(1,14) = 9.25, p = .009, \eta_p^2 = .398$, a significant interaction between fourth- and third-order, $F(1,14) = 12.81, p = .003, \eta_p^2 = .478$, and also a significant 3-way interaction between fourth-, third-, and second-order, $F(1,14) = 13.43, p = .003, \eta_p^2 = .490$. This broadly replicates the sequential effects found in Experiment 1, where participants showed large differences
in the speed and accuracy of their responses between XYZ and ZYZ subsequences, and produced both repetition and alternation effects that explain the 3-way interaction. Interestingly, while participants showed very strong cueing effects overall, the 4-way interaction was not significant in either RTs or errors, $F < 1$, nor were any other interactions significant, largest $F(1,14) = 2.06$, $p = .174$, $\eta^2_p = .128$, suggesting that the pattern of sequential effects did not differ on cued and miscued trials (see Figure 2.4).

### 2.3.3 Discussion

Using a novel three-choice RT task where the target locations were arranged on the left, top and right of a computer screen, a robust pattern of sequential effects was found in the absence of a cued direction of motion in Experiment 1, and for cued and miscued trials in a transfer phase following training with a biased direction of motion in Experiment 2. When the contingencies were biased in one direction 75% of the time, participants appeared to learn this probabilistic sequence by showing a cueing effect once the contingencies were removed in the transfer phase of Experiment 2. While RTs in the transfer phase were generally lower for cued than for miscued trials, the pattern of sequential effects did not appear to differ between cued and miscued trials. Participants responded fastest to subsequences containing repetitions of target direction (YZXYZ trials), similar to advantages for repetitions of target location observed in the two-choice RT literature (e.g. Bertelson, 1961). Interestingly, within the ZYZ subsequences, the fourth-order subsequence that was responded to most rapidly was the one where the target direction alternated consistently (ZYZYZ), again similar to the alternation effects found using two-choice RT tasks (e.g. Cho et al., 2002). While higher-order sequential effects were found
suggesting that responding was influenced by the target location occurring 4 trials back, it was clear that the largest sequential effect was at the second-order level, specifically between XYZ subsequences (e.g. left-top-right or right-top, left), where the target direction travelled in a consistent direction on consecutive trials, and ZYZ subsequences (e.g. left-top-left, right-top-right), where the target alternated its direction (see Figure 2.3 and 2.4). Participants found trials where the target rotation travelled in the same direction twice in a row much easier to respond to than trials where the target travelled in different directions on successive trials.

2.4 Experiment 3

The sequential effects found in Experiments 1 and 2 highlight an important consideration for the experiments that follow, concerning the relative proportion of XYZ trials to ZYZ trials when the contingencies are added to the task. Table 2.2 displays the frequencies of the various subsequences for cued and miscued trials. A natural consequence of introducing a prevailing direction of motion is that some subsequences occur more often on cued trials and less often on miscued trials (or vice versa). At the second-order level, if we divide all trials according to whether they are cued or miscued, and XYZ or ZYZ subsequences, there are a greater proportion of cued trials that are XYZ trials (.5625) than ZYZ trials (.1875), whereas the reverse is true for the miscued trials (.0625 are XYZ and .1875 ZYZ, see Table 2.2). Given that participants naturally respond faster to XYZ trials, this means that any cueing effects obtained could be inflated when averaging over all cued and all miscued trials.
Table 2.2. Frequencies and probabilities of subsequences at second-, third-, and fourth-order levels.

| Subsequence | Cued/Miscued | p(subsequence) | p(target t | target t-1) |
|-------------|-------------|----------------|-----------------|
| XYZ – Cued  | C-C         | 0.5625         | 0.75            |
| XYZ – Miscued | M-M       | 0.0625         | 0.25            |
| ZYZ – Cued  | M-C         | 0.1875         | 0.75            |
| ZYZ – Miscued | C-M      | 0.1875         | 0.75            |
| ZXYZ - Cued | C-C-C-C     | 0.421875       | 0.75            |
| ZXYZ – Miscued | M-M-M       | 0.015625       | 0.25            |
| YXYZ – Cued  | M-C-C       | 0.140625       | 0.75            |
| YXYZ – Miscued | C-M-M    | 0.046875       | 0.75            |
| XYZ – Cued  | M-M-M-C     | 0.03515625     | 0.75            |
| XYZ – Miscued | M-C-M-M | 0.01171875     | 0.75            |
| XYZ – Cued  | M-M-M-C     | 0.01171875     | 0.75            |
| XYZ – Miscued | C-C-C-M   | 0.01171875     | 0.75            |
| YXYZ – Cued  | C-C-C-M     | 0.10546875     | 0.75            |
| XYZ – Miscued | M-C-M-M | 0.10546875     | 0.75            |
| YXYZ – Cued  | M-C-M-M     | 0.01171875     | 0.75            |
| XYZ – Miscued | C-C-M-M   | 0.01171875     | 0.75            |
| YXYZ – Cued  | C-C-M-M     | 0.01171875     | 0.75            |
| XYZ – Miscued | C-M-C-M | 0.01171875     | 0.75            |
| YXYZ – Cued  | C-M-C-M     | 0.10546875     | 0.75            |
| XYZ – Miscued | M-C-M-M | 0.03515625     | 0.75            |
| YXYZ – Cued  | M-C-M-M     | 0.03515625     | 0.75            |
| XYZ – Miscued | M-C-M-M | 0.03515625     | 0.75            |
| XYZ – Cued  | M-C-M-M     | 0.03515625     | 0.75            |

This confound would not be the cause of any observed group differences, since the relative proportions of each trial type would be the same in all groups. However, it means that any calculation of overall cueing effects may reflect a combination of learning and sequential effects. To minimize this confound, in the subsequent experiments the cueing effect was calculated separately for XYZ and ZYZ subsequences. This confound, coupled with the strength of the main effect at the second-order level compared to other levels found in Experiment 1 and Experiment 2,
provided the impetus for focusing on analyses at the second-order level for the following experiments. This distinction between the second-order subsequences turned out to be critical in the experiments that follow.

The aim of Experiment 3 was to examine the effect of a hint about a relational rule on SRT performance by comparing learning between a Hint group and a No Hint group. The rule adequately summarized a set of probabilistic contingencies using a very simple spatial relation, namely the prevailing direction of motion (clockwise or anticlockwise). The ratio of cued to miscued trials used in Experiment 2 (75:25) is high enough that the prevailing direction of motion should be relatively easy to discover (at least when given explicit knowledge of its possible existence). This makes the Hint group more similar to the groups in previous studies who were taught the sequence (e.g. Stefaniak et al., 2008), rather than those who were simply told to look for one (e.g. Jones & McLaren, 2009).

As noted previously, the simplicity of the probabilistic rule differs from previous studies, which tend to use more complex probabilistic sequences that may limit the usefulness of explicit knowledge (e.g. Jiménez, Méndez, and Cleeremans, 1996). Another notable difference between the Hint group and those of other studies is that in addition to giving participants the intention to search for a sequence (e.g. Jones & McLaren, 2009), participants were given a hint to discover a relational rule that captured abstract qualities of the contingencies. The effect of abstract relational knowledge on SRT performance has not been examined in depth, as most SRT tasks are not constructed in a way where such knowledge is immediately applicable. The most relevant explicit knowledge usually consists of memorized sections of the sequence itself in the case of deterministic sequences (e.g. Curran & Keele, 1993), or individual rules about permissible or probable transitions in the case of probabilistic
sequences (e.g. Jiménez, Méndez, and Cleeremans, 1996). Knowing that there is a prevailing direction of motion is abstract but is still closely related to the individual contingencies (e.g. knowing that the target moves mostly clockwise necessarily entails knowledge that the target should appear in the right position after the top position). This knowledge should be easy to apply due to both the simplistic nature of the task (only 3 target locations, only 2 possible directions of target movement), and also the simplicity of the rule, which does not require keeping a large amount of information in working memory.

If Stefaniak et al. (2008) are correct, the probabilistic nature of the contingencies may mean that this knowledge will be difficult to use because it does not allow for 100% successful prediction. On the other hand, if the major determinant for whether explicit knowledge will be utilized is the ease of application to the task rather than whether the sequence was probabilistic or deterministic, then the Hint group should produce a larger cueing effect than the group not given the hint. It was also expected that the Hint group would exhibit higher levels of sequence knowledge on subsequent tests of explicit knowledge. These were included primarily to check whether the hint was successful in producing explicit and transferable sequence knowledge, rather than to assess absolute levels of awareness.

2.4.1 Method

2.4.1.1 Participants

Forty-six University of Sydney students (33 female, $M$ age = 19.6, $SD = 3.40$) participated in this experiment in exchange for course credit or payment (AUD $15/hour). Participants were randomly allocated to either the Hint ($n = 23$) or No Hint ($n = 23$) group.
2.4.1.2 Procedure

The instructions given to both groups were the same as the previous experiments, and did not allude to the fact that the target could not appear in the same location twice in a row, nor that they should attend to the movement of the target. After a practice phase (48 trials) the No Hint group were told that they would now start the experiment, while the Hint group were given a written hint on a piece of paper that read “A hint that will help you in this experiment: The location of pink circles follows a pattern. Most of the time, the location will either go in a CLOCKWISE direction or an ANTI-CLOCKWISE direction. Try and work out which direction it goes. You will be tested on this afterwards.” The training phase was identical to Experiment 2, with 720 trials in total completed in one continuous session. There was no transfer phase in Experiment 3.

After the training phase participants completed a recognition test, and then a prediction test. The recognition test consisted of 10 trials where participants responded to 2 sequences consisting of 24 target locations each. One of these sequences followed the same .75 cued/.25 miscued contingencies the participant had been responding to during training, and the other sequence followed the reverse contingencies, such that the target was travelling predominantly clockwise in one of the sequences and anticlockwise for the other (with the order counterbalanced). After the participant had responded to both sequences they were asked to select which sequence (the first or second) they thought most closely matched the sequence to which they responded to during training.

The prediction test consisted of 3 trials where the target was shown in each of the 3 possible locations in randomized order, and participants had to select which of the remaining two locations they thought the target was most likely to appear next.
2.4.2 Results

A 2 x (2 x 2) ANOVA with group as the between-groups factor and cueing (cued vs. miscued) and subsequence (XYZ vs. ZYZ) as the within-groups factors was performed on both RTs (Figure 2.5) and error data. For RTs, there was a main effect of subsequence, $F(1,44) = 186.5, p < .001, \eta_p^2 = .809$, and cueing, $F(1,44) = 136.7, p < .001, \eta_p^2 = .757$. These effects were also present in the error data, smallest $F(1,44) = 14.84, p < .001, \eta_p^2 = .252$. These main effects indicate that cueing effects were present in both RTs and errors and that participants were both faster and made less errors on XYZ subsequences than ZYZ subsequences, replicating the second-order sequential effects found in Experiments 1 and 2.

![Figure 2.5](image)

*Figure 2.5. Cueing effect (RT for miscued – cued trials) for each group for XYZ and ZYZ subsequences across training quarters in Experiment 3. Error bars represent the standard error of the mean.*

While the group x cueing interaction was marginally non-significant in RTs, $F(1,44) = 3.64, p = .063, \eta_p^2 = .076$, the 3-way interaction between group, cueing and subsequence was significant, $F(1,44) = 4.57, p = .038, \eta_p^2 = .094$, suggesting that the difference in the size of the cueing effect between XYZ and ZYZ subsequences differed between groups (Figure 2.5). This was confirmed by a significant group x
cueing interaction in XYZ RT, $F(1,44) = 7.29, p = .010, \eta_p^2 = .142$, but not for ZYZ RT, $F < 1$. The 3-way interaction in errors did not reach significance, $F < 1$, nor did the group x cueing interaction, $F < 1$, however, there was a significant interaction of subsequence with cueing, $F(1,44) = 15.69, p < .001, \eta_p^2 = .263$, due to the ZYZ cueing effect being larger than the XYZ cueing effect overall. There was no overall effect of group in either RTs, $F < 1$, or errors, $F < 1$.

While the Hint group scored numerically higher than the No Hint group on the recognition test (63.9% vs. 55.2%), the group difference failed to reach significance, $F(1,44) = 2.00, p = .164, \eta^2 = .044$ (see Figure 2.6). This pattern of results was mirrored on the prediction test, with the Hint group scoring 69.6% and the No Hint group scoring 59.4%, but again there was no significant group difference, $F(1,44) = 1.07, p = .306, \eta^2 = .024$ (Figure 2.6). However, the Hint group performed significantly higher than chance on both the recognition, $t(22) = 2.91, SEM = .047, p = .008$, and prediction tests, $t(22) = 2.83, SEM = .208, p = .010$, while the No Hint group were not significantly different from chance performance on both tests, largest $t(22) = 1.36, SEM = .208, p = .188$.

![Figure 2.6](image_url). Performance on the recognition and prediction tests in Experiment 3. Error bars represent the standard error of the mean.
2.4.3 Discussion

Participants were both faster and more accurate at responding on XYZ than ZYZ subsequences (as in Experiments 1-2), and also on cued than miscued trials (i.e. there was an overall cueing effect). It is clear that the 3-way interaction can be explained by the presence of a group difference in the size of the XYZ cueing effect, but no group difference in the size of the ZYZ cueing effect. The hint was therefore successful in increasing the amount of cueing for XYZ subsequences, but not ZYZ subsequences. Surprisingly, the hint was not successful in increasing the amount of explicit knowledge expressed on a recognition and prediction test relative to the group that did not receive a hint. The group differences in training can be explained if we consider that the hint may have resulted in increased attention to the general direction of the target as participants tried to use the hint, and since a continuous direction of motion is only present on XYZ trials, perhaps this meant that these were the only trials on which participants were able to use the hint. Alternatively, perhaps participants in the Hint group found ZYZ trials difficult to express their knowledge due to the target alternating direction on successive trials. The sequential effects in Experiments 1-2 demonstrate that the alternation of target direction disrupted performance in general in comparison to instances when the target direction repeated. This may indicate a lower level of control over performance on ZYZ subsequences, which may in turn have made it difficult to express sequence knowledge on these trials.

In either case, it is clear that the Hint group displayed greater cueing only for XYZ subsequences even though the hint was equally valid for both XYZ and ZYZ subsequences. Thus it appears that whether or not knowledge of a relational rule impacts sequence learning depends more on the ease with which that knowledge can
be applied, rather than whether the sequence is deterministic or probabilistic, as suggested by Stefaniak et al. (2008) and Cleeremans and Jiménez (1998). However, since there were no group differences in the explicit knowledge tests, it is hard to know how the Hint group interpreted the hint or what additional knowledge they possessed compared to the No Hint group. While it is possible that both recognition and prediction tests were simply not sensitive, a likely reason for the lack of group differences in the prediction test may be that the recognition test was quite long and exposed participants to sequences moving in the clockwise and anticlockwise directions. This may have made the recognition test itself somewhat confusing and reduced the sensitivity of any ensuing tests of explicit knowledge. Consequently, in Experiment 4 the recognition test was omitted, and the prediction test immediately followed the SRT task, and a simple forced-choice question that asked participants whether the target was mostly travelling clockwise or anticlockwise was presented at the very end of the experiment.

2.5 Experiment 4

In Experiment 4, the recognition test was omitted to increase the chances of obtaining a group difference in the prediction test. Experiment 4 also tested whether the knowledge acquired by the Hint group would transfer to an additional phase where the target locations were changed to the left, right and bottom of the screen. This transfer phase allows separation of the abstract relation (direction of motion) from the other task properties with which it was correlated during training. For instance, if participants learnt that the target was usually moving in a clockwise direction (e.g. left – top – right), they should be able to use this knowledge despite the new target locations, as a clockwise direction of motion would still be apparent (e.g.
right – bottom – left). However, since the ‘top’ target location was removed, most of
the relevant contingencies between perceptual elements of the task (associations
between left and top and associations between top and right) would be removed in the
transfer test.

Furthermore, since participants still used the same finger to respond to the
bottom target location as they had to respond to the top location in training, this meant
that the new target locations required participants to reverse the motor sequence they
had previously used when the target was travelling in the cued direction (i.e. left-top-
right is the same direction as right-bottom-left, but requires the finger motor sequence
2-3-4 and 4-3-2 respectively). This test was intended to assess whether the cueing
effect in each group could be attributed more to simple motor learning, or relational
rule learning, which would predict cueing effects in opposite directions. The
contingencies were removed in the test phase so that there was an equal opportunity
to show a bias in either direction.

Since the explicit hint emphasized the direction of rotation of the target, it was
hypothesized that the Hint group would still show a cueing effect consistent with the
prevailing direction during training despite perceptual and motor changes that
conflicted with that direction of motion. The No Hint group might instead be expected
to show a cueing effect in the opposite direction to the prevailing rotation of the target
instead if what they had learned concerned the contingencies between motor actions
(i.e. the sequence of key presses). However, the expression of implicit learning may
rely on the reinstatement of similar conditions under which learning occurred, while
explicit learning might be more flexible and thus less dependent on superficial task
changes (Jiménez, Vaquero, & Lupiáñez, 2006). Therefore, it was assumed that
cueing would be less evident in the No Hint group irrespective of the direction of the cueing effect.

2.5.1 Method

2.5.1.1 Participants

Fifty University of Sydney students participated in this experiment in exchange for course credit (17 female, \( M \) age = 20.16, \( SD \) = 3.18) or payment (AUD $15/hour). Participants were randomly allocated to the Hint (\( n = 24 \)) or No Hint (\( n = 26 \)) group.

2.5.1.2 Procedure

The procedure was the same as Experiment 3 except for the following changes. After the training phase participants completed a short transfer phase (120 trials) where they were told that the target would now appear in the left, bottom, and right of the screen, and they were either informed to use the corresponding arrow keys (pressing the ‘down’ arrow key for the bottom target location), or the same keys they had been using previously (pressing the ‘up’ arrow key for the bottom target location). It was important to compare transfer using the same versus different response keys as there is evidence that changing the response keys impairs sequence learning (Willingham, Nissen, & Bullemer, 2000) and it was unknown whether this would occur in the current task. In the latter condition the three arrow keys (left, up, and right) were marked with stickers and participants were told to only ever press those keys in both phases. The transfer phase did not contain any contingencies, such that there was no prevailing direction of motion.
After the transfer phase participants in both groups completed the prediction test, which was the same as Experiment 3 except that it was made clear that the predictions should be based on what was learnt in the first part of the experiment before the target locations changed (the training phase). Finally, participants were told that the target moved in one direction most of the time in the first part of the experiment, and were asked to indicate whether they thought the target moved in a predominantly clockwise or anticlockwise direction by pressing one of two keys.

2.5.2 Results

A 2 x (2 x 2) ANOVA was run with group (hint vs. no hint) as the between-groups factor, and subsequence (XYZ vs. ZYZ) and cueing (cued vs. miscued) as the within-groups factors on RTs (Figure 2.7) and errors in the training phase. Replicating Experiment 3, in RTs there was a significant main effect of subsequence, $F(1,48) = 203.96, p < .001, \eta_p^2 = .809$, and cueing, $F(1,48) = 112.68, p < .001, \eta_p^2 = .701$. Thus there was a significant cueing effect and participants showed the same second-order sequential effects as the previous experiments, responding faster to XYZ subsequences. Replicating Experiment 3, there was a significant interaction between subsequence, cueing and group in RTs, $F(1,48) = 6.19, p = .016, \eta_p^2 = .114$, indicating that the group difference in cueing differed between subsequences. Again, this three-way interaction was explained by the fact that there was a significant cueing x group interaction for XYZ subsequences, $F(1,48) = 6.94, p = .011, \eta_p^2 = .126$, but not for ZYZ subsequences, $F < 1$. There was also a marginally non-significant group x cueing interaction, $F(1,48) = 3.38, p = .072, \eta_p^2 = .066$ (see Figure 2.7).
In the error data there was a main effect of subsequence, $F(1,48) = 80.87, p < .001, \eta^2_p = .628$, and cueing, $F(1,48) = 62.37, p < .001, \eta^2_p = .565$, and also a significant interaction between subsequence and cueing, $F(1,48) = 17.23, p < .001, \eta^2_p = .264$, but the 3-way interaction did not reach significance, $F < 1$, nor did the group x cueing interaction, $F(1,48) = 2.90, p = .095, \eta^2_p = .057$. There was no overall effect of group in either RTs or errors, largest $F < 1$. Therefore in errors, the same second-order sequential effects and cueing effects were found as in RTs, but there did not appear to be an effect of the hint on overall errors or on the cueing effect.

For the transfer phase, a 2 (hint vs. no hint) x 2 (same vs. different response keys) x 2 (XYZ vs. ZYZ) x 2 (cued vs. miscued) ANOVA produced a main effect of subsequence in RTs, $F(1,46) = 131.3, p < .001, \eta^2_p = .741$, and errors, $F(1,46) = 45.75, p < .001, \eta^2_p = .499$, but no other main effects or interactions reached significance. Simple t-test comparisons revealed that none of the transfer cueing effects in either group (for either subsequence, in RTs or errors) were significantly greater than 0, all $t$s < 1, suggesting that there was no transfer of learning in any condition.
Unlike Experiment 2, there was a significant group difference in accuracy for the prediction test, $F(1,48) = 5.65, p = .021, \eta^2 = .105$ (see Figure 2.8). The Hint group scored 80.56% ($SD = 23.9$), which was significantly above chance, $t(23) = 6.26, SEM = .049, p < .001$, while the No Hint group scored 58.97% ($SD = 38.1$), which was not found to be significantly above chance, $t(25) = 1.20, SEM = .075, p = .241$. A chi-square analysis revealed that a higher proportion of participants in the Hint group chose the correct direction of motion (24/24) than the No Hint group (17/26), $\chi^2(1, N = 50) = 10.13, p = .001, \phi = .450$ (see Figure 2.8).

![Figure 2.8. Performance on the prediction test and forced-choice question in Experiment 4. Error bars represent the standard error of the mean.](image)

2.5.3 Discussion

The results in the training phase replicate the group differences in Experiment 3, in that participants were both faster and more accurate on XYZ than ZYZ subsequences, and also on cued than miscued trials. Participants in both groups produced a larger cueing effect on ZYZ subsequences in their errors and the difference in RT cueing for XYZ and ZYZ subsequences interacted with group. This 3-way interaction is again due to the fact that the Hint group produced a larger XYZ cueing effect than the No Hint group, but there was no group difference for ZYZ.
cueing. The significant group differences in the explicit knowledge tests confirm that the hint was successful in helping participants discover the direction of motion (all participants in the Hint group could correctly identify the direction of motion), and provided participants with a level of explicit knowledge that they otherwise would not have obtained in the absence of the hint. The results from Experiments 3 and 4 thus far suggest that cueing effects on XYZ subsequences are affected by relational sequence knowledge while cueing effects on ZYZ subsequences may not be susceptible to this same knowledge.

It seemed that changing the target locations in the transfer phase greatly impaired the expression of any cueing effects, regardless of whether participants used the same or different response keys and regardless of whether they were given a hint to concentrate on the direction of the target. It is not surprising that the No Hint group did not show a cueing effect in the transfer phase given that implicit learning has been argued to be tied to the contextual features of the training conditions (Berry & Dienes, 1993; Jiménez, Vaquero, & Lupiáñez, 2006). However, it is surprising that participants in the Hint group also did not show any cueing effects in the transfer phase given that knowledge that the target mostly travels in a certain direction should still be relevant despite the change in target locations. It appears that the abstract knowledge acquired by the Hint group was not sufficient in overcoming the detriment in performance that occurred when the perceptual task features were changed. There is evidence of sequence learning surviving large superficial task changes (e.g. Jiménez, Vaquero, & Lupiáñez, 2006), but only when learning was deemed to be explicit, and only when the target locations and motor responses remained unchanged. The present results suggest that the motor component as well as the perceptual features in the task constituted a large proportion of what participants learned.
Changing these features in the transfer phase eliminated cueing effects, even when participants had relational knowledge that was directly applicable.

In the following experiment, a different transfer phase was used to test whether the hint would facilitate performance. There is evidence showing that sequence learning effects survive motoric changes such as switching from using three fingers to one (Experiment 2, Cohen, Ivry, & Keele, 1990; Experiment 1, Keele, Jennings, Jones, Caulton, & Cohen, 1995), and even when switching hands completely (Deroost, Zeeuws, & Soetens, 2006; Grafton, Hazeltine, & Ivry, 2002; Japikse, Negash, Howard, Jr., & Howard, 2003). Thus, it is possible that changing the motor requirements while retaining the target locations (and hence the perceptual task features) will afford a better chance of observing transfer as well as any possible advantage that the Hint group may have.

2.6 Experiment 5

The aim of Experiment 5 was to replicate Experiment 4 with a transfer phase where the target locations remained the same but participants switched hands to respond. Participants may show cueing effects because they have learned about the contingencies between the target locations on screen (e.g. left – top), but also if they have learned an association between particular motor responses (e.g. pressing the left key – pressing the up key with the right hand). Since there is evidence that perceptual features of the task are important in transfer (e.g. Cohen, Ivry, & Keele, 1990; Keele et al., 1995), it was expected that participants in both groups would show a significant cueing effect in transfer. However, the Hint group should suffer less generalization decrement from switching hands since they possess explicit knowledge about the direction of motion, which is applicable no matter which hand is used for responding.
Therefore it was also hypothesized that the Hint group would show more cueing than the No Hint group in the transfer phase. Following the results thus far, if responding on XYZ subsequences is under a greater amount of control, the Hint group should display a larger XYZ cueing effect, but there will be no group difference in ZYZ cueing at transfer.

2.6.1 Method

2.6.1.1 Participants

Seventy-six participants who were either University of Sydney students or respondents to an online ad took part in this experiment (54 female, M age = 22.24, SD = 7.23). Students received course credit and ad respondents received AUD $15/hour as compensation. Participants were randomly allocated to either the Hint group (n = 39) or the No Hint group (n = 37).

2.6.1.2 Procedure

The procedure was identical to Experiment 4, except that before the transfer phase participants were told to switch to whatever hand they had not been using previously. The target locations and response keys remained the same (left, top, right), and there were no contingencies in the transfer phase. Due to the large proportion of participants in this experiment who were recruited from the general public (i.e. were not university students), the hint was read out to participants in this study, and extra clarification was given that there was a correct ‘answer’ for the target direction, and that it was either clockwise or anticlockwise. To foreshadow the results, this turned out to be critical in the ensuing group differences.
2.6.2 Results

A 2 x (2 x 2) ANOVA with group as the between-groups factor, and subsequence and cueing as the within-groups factors was run on RT and error data in training and transfer. The main effects of subsequence and cueing were again significant in both RTs and errors, smallest $F(1,74) = 69.09$, $p < .001$, $\eta^2_p = .483$, indicating an overall cueing effect and better performance on XYZ subsequences. There was a significant interaction between cueing and group in RTs with the Hint group producing a larger cueing effect overall, $F(1,76) = 21.70$, $p < .001$, $\eta^2_p = .227$, but unlike Experiments 3 and 4, this did not interact with subsequence, $F < 1$ (Figure 2.9).

![Figure 2.9](image)

Figure 2.9. Cueing effect (RT for miscued – cued trials) for each group for XYZ and ZYZ subsequences across training and transfer phases in Experiment 5. Error bars represent the standard error of the mean.

In the error data, there was a significant interaction between subsequence and cueing, $F(1,74) = 26.75$, $p < .001$, $\eta^2_p = .265$, as in Experiments 3 and 4, indicating that a larger cueing effect was obtained in both groups for ZYZ subsequences, and the cueing x group interaction was marginally non-significant, $F(1,74) = 3.15$, $p = .080$, $\eta^2_p = .041$, as well as the three-way interaction between subsequence, cueing and group, $F(1,74) = 3.20$, $p = .078$, $\eta^2_p = .041$. There was no main effect of group for RTs or errors, largest $F(1,74) = 2.50$, $p = .118$, $\eta^2 = .033$. 

78
In the transfer test, there was an overall cueing effect in RTs, $F(1, 74) = 32.28, p < .001, \eta_p^2 = .304$, and errors, $F(1, 74) = 10.72, p = .002, \eta_p^2 = .127$, that did not interact with subsequence or group, largest $F(1, 74) = 1.81, p = .182, \eta_p^2 = .024$. On initial inspection it appeared that the results confirmed the hypothesis – the Hint group produced a larger XYZ cueing effect in the transfer phase, while both groups showed the same level of ZYZ cueing (Figure 2.9), but the 3-way interaction between group, subsequence and cueing did not reach significance in RTs, $F(1, 74) = 1.38, p = .243, \eta_p^2 = .018$, or errors, $F < 1$. A follow-up analysis revealed that there was no group difference in the size of the XYZ cueing effect in RTs in transfer, $F(1, 74) = 2.78, p = .100, \eta_p^2 = .036^{15}$.

There was a significant group difference in accuracy on the prediction test, $F(1, 74) = 5.42, p = .023, \eta^2 = .068$ (see Figure 2.10). The Hint group scored 77.8% ($SD = 29.9\%$), which was found to be significantly above chance, $t(38) = 5.79, SEM = .048, p < .001$, and the No Hint group scored 61.2% ($SD = 31.9\%$), which was also significantly above chance performance, $t(36) = 2.15, SEM = .052, p = .039$. A chi-square analysis showed that a significantly larger proportion of participants chose the correct direction of motion in the Hint group (36/39) than in the No Hint group (23/37), $\chi^2(1, N = 76) = 9.93, p = .002, \phi = .362$ (see Figure 2.10).

---

15 See Appendix B for additional analyses.
Figure 2.10. Performance on the prediction test and forced-choice question in Experiment 5. Error bars represent the standard error of the mean.

2.6.3 Discussion

Replicating Experiment 4, the hint was successful in increasing the amount of explicit sequence knowledge that participants displayed at the end of the experiment. The significant cueing effects seen in the transfer phase showed that the cueing effect persisted when the contingencies were removed and even when using a different hand to respond. This is consistent with previous studies demonstrating transfer of sequence learning between hands (Deroost, Zeeuws, & Soetens, 2006; Grafton, Hazeltine, & Ivry, 2002; Japikse et al., 2003), and along with Experiment 4, is consistent with studies concluding that sequence learning is not purely a series of motor responses (Cohen, Ivry, & Keele, 1990; Keele et al., 1995). Transfer was completely disrupted in Experiment 4 when the perceptual features of the task (target locations) were changed, and significant cueing effects were found to transfer across hands when the target locations remained the same in Experiment 5. Thus it appears that sequence learning in both the presence and absence of the hint is based primarily on perceptual features of the task and not just on simple motoric learning.
Despite the results being in the predicted direction, it seems that cueing persisted equally for XYZ and ZYZ in transfer, and this did not depend on whether participants received the hint. The effect of the hint was seemingly strongest in Experiment 5 compared to previous experiments, elevating cueing for both XYZ and ZYZ subsequences, and yet this advantage did not persist into the transfer phase for either subsequence. The fact that the explicit knowledge afforded by the hint makes little difference when the cueing effects were measured in the absence of contingencies may suggest that the Hint group were sensitive to the fact that the contingencies had been removed and this adversely affected their cueing effect. An alternative explanation is that the sequence knowledge in the No Hint group transferred effectively because participants were primarily learning about the perceptual features of the sequence. This would mean that sequence learning is more flexible than what some theorists have postulated (e.g. Berry & Dienes, 1993; Cleeremans & Jiménez, 1998; Jiménez, Vaquero, & Lupiáñez, 2006). However, since the performance of the No Hint group was significantly above chance in the prediction test in this experiment, it can be argued that this flexibility may have been due to a subset of participants acquiring some explicit knowledge in the absence of the hint.

The surprising finding from Experiment 5 was the absence of the three-way interaction between group, subsequence, and cueing in training that was present in both Experiment 3 and 4. It appears that in this experiment, the hint increased cueing for both XYZ and ZYZ subsequences. One possible reason for this discrepancy between experiments is the method of hint delivery. A large proportion of the sample in Experiment 5 were members of the general public, whereas in Experiments 3 and 4 all participants were university students. Previous experience with non-University
student samples led the experimenter to give these participants extra attention and explanation prior to the experiment, as they were generally not used to the types of computer-based tasks that are routinely completed by most undergraduate psychology students, and found it difficult to follow the on-screen instructions. Thus extra care was taken to explain the hint in Experiment 5 in that the hint was read out verbally to participants, and it was emphasized that there was a correct ‘answer’ that was either clockwise or anticlockwise. It is possible that this extra clarification enabled the Hint group to use their knowledge on ZYZ trials in addition to XYZ trials. Experiment 6 aimed to test this speculation.

2.7 Experiment 6

The aim of Experiment 6 was to provide evidence that a more careful mode of delivering the hint could selectively increase the amount of cueing for ZYZ subsequences. To achieve this, Experiment 6 compared learning between two groups: a group that were simply given the hint to read as in Experiments 3-4 (Weak Hint group), and a group where the experimenter read out the hint and explained that there was a correct answer which would either be clockwise or anticlockwise (Strong Hint group). Based on the results of Experiments 3-5, it was hypothesized that the groups would differ in cueing for ZYZ subsequences, for which the weak version of the hint had little impact in Experiments 3 and 4, but would display equivalent and relatively strong cueing effects for XYZ subsequences, which appeared to benefit from both forms of the hint across Experiments 3-5.

Thus far, no evidence has been found for an advantage of the Hint group in transfer when the contingencies are removed. As discussed above, one possibility is that the changes to the perceptual features (Experiment 4) or the response demands
(Experiment 5) in transfer may be especially disruptive when attempting to learn the sequence explicitly. If participants in the Hint group are constantly looking for the predominant direction of motion, when the contingencies are removed in the transfer phase this may disrupt performance in general, leading to smaller cueing effects that are equivalent in magnitude to the No Hint group. It is possible that a difference between the Weak and Strong Hint groups may emerge not just in training but also in transfer. To provide the best opportunity for observing a difference in transfer between Hint groups, Experiment 6 used the same transfer phase used in Experiment 2: where participants used the same response keys and hand to respond, and the transition to the transfer phase was seamless (the start of the transfer phase was unmarked). Thus the only change during the transfer phase would be the removal of the contingencies. The hypothesis for the transfer phase was the same as in training: a larger ZYZ cueing effect but equivalent XYZ cueing effect for the Weak and Strong Hint groups.

2.7.1. Method

2.7.1.1 Participants

Forty-three University of Sydney Psychology students (30 female, $M$ age = 19.6, $SD = 3.13$) participated in Experiment 6 in exchange for course credit. Participants were randomly allocated to either the Weak Hint group ($n = 22$) or the Strong Hint group ($n = 21$).

2.7.1.2 Procedure

Experiment 6 consisted of a single training session where participants were told to respond as quickly and accurately as possible. After completing a short
practice phase (the same as the previous experiments), participants responded to 720 trials where the target followed the .75 cued/.25 miscued contingencies (there was a prevailing direction of motion) and 360 trials where there were no contingencies (there was no prevailing direction of motion). Participants continued to use the same response keys and hand and there was no break between the training and transfer phase, such that there was nothing to mark the transition into a separate phase for participants. Participants in Experiment 6 received a written hint after the practice phase if they were in the Weak Hint group, which they were told to read before starting the main part of the experiment. Participants in the Strong Hint group were also given the written hint, but the experimenter read out the hint to the participants. The hint given was exactly the same in both groups except for the addition of the following sentence in the Strong Hint group: “Try and work out which direction it goes (the answer will EITHER be clockwise or anticlockwise).” There were no explicit knowledge tests in this experiment.

### 2.7.2 Results and Discussion

One participant from the Weak Hint group was excluded due to their cueing effect on both XYZ and ZYZ subsequences deviating more than 3 SDs from the mean, leaving 21 participants in each group. A 2 x (2 x 2) ANOVA with group as the between-groups factor, and subsequence (XYZ vs. ZYZ) and cueing (cued vs. miscued) as the within-groups factors was run on RTs and errors in the training and transfer phases. In both the training and transfer phases, there was a significant main effect of cueing and subsequence for both RTs and errors, smallest $F(1,40) = 25.64, p < .001, \eta_p^2 = .391$, and a significant interaction between subsequence and cueing in errors, $F(1,40) = 6.31, p = .016, \eta_p^2 = .136$. Cueing effects were therefore present in
both RTs and errors, and again participants found XYZ subsequences easier to respond to. As with previous experiments, participants produced a larger ZYZ cueing effect in errors only. The interaction of interest, between group, subsequence and cueing, was not significant in training in RTs, $F < 1$, however it was significant at transfer, $F(1,40) = 11.67, p = .001, \eta_p^2 = .226$ (see Figure 2.11). Follow-up analyses revealed that this interaction was due to a group difference in ZYZ cueing, $F(1,40) = 6.69, p = .013, \eta_p^2 = .143$, and no group difference in XYZ cueing, $F < 1$, as hypothesized. The three-way interaction did not reach significance in errors in either training or transfer, largest $F(1,40) = 2.90, p = .096, \eta_p^2 = .068$.

![Figure 2.11. Cueing effect (RT for miscued – cued trials) for each group for XYZ and ZYZ subsequences in Experiment 6. Blocks 1-4 are the training phase, blocks 5-6 are the transfer phase.](image)

While it should be acknowledged that this result would be most consistent with the previous experiments had it been found during training, to the extent that cueing effects in transfer can be interpreted as reflecting a genuine learning effect, it can still be concluded that there is a difference between the two modes of hint delivery. The interaction at transfer for RTs confirms the speculation that stressing to participants that the answer would either be clockwise or anticlockwise, as well as
delivering the hint verbally selectively increases ZYZ cueing effects. It may be that the deliberate search for a pattern (motivated by any form of hint) was reduced to deciding on a direction of motion, substantially simplifying the task at hand and thus allowing the expression of explicit knowledge on both second-order level subsequences. Presumably, the hint in its original form directs participants to pay attention to the general direction of the target, but perhaps entertainment of more complex rules due to the probabilistic nature of the task meant that participants were still searching for patterns throughout the training and transfer phases. It should be emphasized that this explanation does not necessitate that the Strong Hint group possessed a greater degree of sequence knowledge, rather it suggests that the difference in hint delivery between the two groups in Experiment 6 made the task easier in the Strong Hint group which then changed the way they approached the task. In summary, the greater ZYZ cueing effect obtained by reading out the hint and emphasizing that there was a correct answer explains the failure to obtain the 3-way interaction in Experiment 5. In Experiment 5, both XYZ and ZYZ cueing benefitted from the strong version of the hint, producing a group difference in ZYZ cueing that was not obtained in Experiments 3 and 4.

2.8 General Discussion

The current series of experiments explored whether a hint about a relational rule describing a set of simple contingencies could affect learning in a probabilistic three-choice SRT task. This was investigated by comparing a group given a hint about the existence of a prevailing direction of motion against a group who were not given this hint. This manipulation was novel in the sense that participants were aided in discovering a simple relational rule that could summarize all contingencies within the
task and was therefore equally applicable on every single trial. Experiments 1 and 2 showed that the three-choice SRT task produced sequential effects in target direction, similar to the repetition and alternation effects found in 2-choice RT tasks with target location. Participants were generally much faster to respond on trials where there was a repetition of target direction (XYZ subsequences), compared to trials where the target direction alternated (ZYZ trials), but participants were also fast to respond when the target direction alternated a few times (ZYXYZZ subsequences).

Due to the magnitude of the second-order sequential effects (XYZ vs. ZYZ) in comparison to the higher-order sequential effects, cueing effects for each subsequence were analyzed separately to avoid artificially inflating the size of the overall cueing effect when the contingencies were added to the task. This separation turned out to be critical for the ensuing experiments. Experiments 3 and 4 showed a consistent pattern of results in training, with the Hint group producing a larger XYZ cueing effect but no group difference in the ZYZ cueing effect. Experiment 3 failed to produce group differences in performance on a recognition and prediction test, but Experiment 4 found significantly better performance in the Hint group on the prediction test, and a significantly larger proportion of participants in the Hint group selected the correct direction of motion than the No Hint group.

Experiment 4 also added a short transfer phase after the training phase to test whether the sequence knowledge in the Hint group was abstract and could be applied when the target locations were changed to the left, bottom, and right of the screen and the contingencies were removed. The test pitted motor and relational learning against each other such that a significant cueing effect in either direction (in the cued or miscued direction) would give some indication of whether the content of learning in each group was primarily motor or abstract. No cueing effects in the transfer phase
were found in either direction in either group, suggesting that perceptual information constituted a large portion of what participants in both groups had learnt, with the disruption of this aspect of the task eliminating cueing effects completely.

Experiment 5 implemented a different transfer phase where participants simply switched hands to respond. This time, positive cueing effects in the cued direction of motion were found over both subsequences and these cueing effects did not differ between groups. However, the Hint group in Experiment 5 produced a larger cueing effect for both XYZ and ZYZ subsequences in training, in contrast to Experiments 3 and 4. Experiment 6 confirmed that the likely reason for this difference was the manner in which the hint was administered. Delivering the hint verbally with additional emphasis that there was a ‘correct’ prevailing direction (as was the case in Experiment 5) resulted in a larger ZYZ cueing effect in transfer but no change in the size of the XYZ cueing effect, in comparison to just giving participants the hint to read (as was the case in Experiments 3-4).

2.8.1 Effects of Explicit Knowledge

This study suggests that the primary determinant for whether explicit knowledge will be used in an SRT task is how easy that knowledge is to apply to the task, not necessarily whether the sequence is probabilistic or deterministic, as suggested by previous studies. In particular, the Hint group found it easier to apply their knowledge on XYZ trials, perhaps due to the consistent direction of motion on two consecutive trials having a facilitating effect. One possibility for this facilitation is that responding on XYZ trials may just have been easier for both groups, making expression of the additional explicit knowledge in the Hint group more
straightforward. Certainly, the sequential effects in Experiments 1-2 suggest that responding is easier on XYZ trials compared to ZYZ trials.

Previous studies demonstrating a lack of difference between a group given additional explicit sequence knowledge and a group who perform the task in the absence of this information can also be explained on the basis of the ease of application of that knowledge to the task. The sequence used in Jiménez, Méndez, and Cleeremans (1996) was probabilistic but also quite complex, generated according to a finite-state grammar (Figure 2.12). In Figure 2.12, the arrows represent possible or ‘grammatical’ transitions and the letters represent different targets. One can imagine that learning this structure explicitly would be difficult, let alone attempting to apply that knowledge to the SRT task that requires fast responses.

Figure 2.12. The finite-state grammar used to generate sequences in the study reported by Jiménez, Méndez, and Cleeremans (1996). Image from Jiménez, Méndez, and Cleeremans (1996).

The results of another study can be explained on the basis of the utility of that knowledge to the task. A study by Sanchez and Reber (2013) showed that pretraining participants on a sequence did not aid performance despite participants being able to verbalize the sequence structure. This result was especially surprising because the
sequence was a 12-item deterministic (repeating) sequence. However, their task was quite different from traditional SRT tasks in that participants had to monitor a series of “falling” cues and press the corresponding key at the precise time when each cue reached the cue outline at the bottom of the screen (see Figure 2.13). Participants’ accuracy was recorded (hit or miss) in contrast to traditional SRT tasks that record reaction time. The task also differed in that participants were presented with multiple cues on screen at the same time and so could “see” the upcoming sequence of cues. Sanchez and Reber explained their results by claiming that explicit knowledge allows participants to predict future responses, but when the upcoming targets are on screen and therefore do not need to be predicted, the benefit of explicit knowledge disappears. They propose that while explicit knowledge may have a role in skill learning (which the SRT paradigm provides a means to test), it may have more of a ‘scaffolding’ effect rather than representing a direct contribution (see Jiménez, 2003, for a similar conclusion). This means that explicit knowledge can help with planning movements or preventing inappropriate movements when initially learning a skill, but it is not responsible for improvements seen after practice (see Petersen, van Mier, Fiez, & Raichle, 1998).

![Figure 2.13. SRT task used in Sanchez and Reber (2013). Image from Sanchez and Reber (2013).](image)
An alternative explanation of the effect of the hint in Experiments 3 and 4 is that the hint engages explicit learning processes, which selectively benefit the most salient subsequences (Jones & McLaren, 2009). Jones and McLaren (2009) found that when participants were told to predict the location of the target, they showed the best evidence of learning for subsequences that ended in a repetition (e.g. XXX), since runs of repetitions constitute the most salient subsequences. This idea is consistent with the current results if we assume that one effect of the hint was to draw participants’ attention to the direction of the target, which is more salient, or more easily encoded, on XYZ trials. Assuming that the hint made discovering the predominant direction of motion relatively easy (supported by the results of the explicit knowledge tests in Experiments 4 and 5), this selective effect of the hint would appear to be a performance effect, rather than a failure of learning that the hint applies to ZYZ trials. Participants in the Hint group may have known that the target was mostly travelling in one direction, but perhaps they found it difficult to express their knowledge when the target location alternated because it violated their expectations and was thus surprising. Alternatively, it may be that the probabilistic nature of the task makes it difficult for participants to be sure of the correct direction of motion throughout the task. Thus, when the target direction alternates, it is not only surprising because of the inconsistency with the previous trial, but it may also violate the current beliefs of the participant that the target is travelling in their chosen direction. This may make the participant entertain more complex hypotheses such as the possibility that the target direction may switch, or that the sequence is much more complex than the hint suggests. Indeed, when the hint was clarified to participants in Experiment 5 so that the task for the Hint group was reduced to deciding on a predominant direction of motion from two possibilities, participants were able to use
this information to produce a larger cueing effect on both XYZ and ZYZ subsequences.

As discussed, it is unlikely that the difference between the weak and strong versions of the hint is in the quality or amount of resulting explicit knowledge, since performance in the explicit sequence knowledge tests in Experiments 4 and 5 were very similar. Rather, the effect of clarification in the Strong Hint group may have reduced uncertainty and the overall cognitive load of the hint during training, as participants no longer had to search for more complex rules, resulting in the Hint group being able to more easily apply their knowledge to ZYZ trials. Note that this explanation implies that explicit processes actually hampered the expression of learning on ZYZ trials in the Hint groups in Experiments 3-4, and would suggest that simply adding the intention to learn does not necessarily make the task easier if the sequence is probabilistic.

The additional care in emphasizing and explaining the hint in order for participants to successfully utilize that knowledge on both XYZ and ZYZ subsequences might suggest that the hint was not clear enough and participants may have misinterpreted it. Although this is a possibility, performance in the prediction test and forced-choice question was good after the weak version of the hint was given in Experiment 4, so any potential misinterpretation of the hint cannot have been particularly detrimental to discovering the prevailing direction of motion. While the performance of the Hint group in Experiment 3 was comparatively poor, this may well have been because the recognition test presented prior adversely affected the prediction test, since the recognition test involved participants responding to sequences biased in both the cued and miscued direction. This additional exposure to contingencies in the reversed direction may have degraded any existing sequence
knowledge that participants possessed, or made them unsure of the predominant
direction of motion that they had worked out during training.

### 2.8.2 Limits of Explicit Knowledge

A surprising finding from the current experiments was that the hint produced
limited enhancement of the cueing effect in the transfer phases. While the training
data seem to suggest that the hint had a large impact on SRT performance, the failure
to find the same differences in the transfer phases in Experiments 4 and 5 suggest that
the knowledge afforded by the hint was less flexible than expected. The Hint groups
in both Experiments 4 and 5 failed to exhibit greater transfer of cueing effects than the
No Hint groups, with both groups displaying no cueing effects in Experiment 4, and
equivalent cueing effects in Experiment 5. The lack of transfer for either group in
Experiment 4 was explained through changes to the perceptual features of the task
being sufficient to disrupt transfer completely. Clearly, a large part of what is learned
in this SRT task is the sequence of target locations on screen, rather than a series of
motor actions, and this was the same regardless of the hint. In Experiment 5, sequence
knowledge in both groups transferred across hands, further suggesting that whatever
is learned is not restricted to specific motor actions and is primarily perceptual. These
results accord with other studies that have found transfer of sequence learning
between hands and fingers (Cohen, Ivry, & Keele, 1990; Deroost, Zeeuws, & Soetens,
2006; Grafton, Hazeltine, & Ivry, 2002; Japikse et al., 2003; Keele et al., 1995), and
studies that show that retaining the perceptual aspects of the task is more important
than retaining the motor aspects in demonstrating transfer (e.g. Stadler, 1989).

The lack of group differences for either subsequence in the transfer phase in
Experiment 5 is surprising given that there should be limited generalization decrement
in the explicit knowledge possessed by the Hint group. In other words, switching hands to respond should have little bearing on both the quality of the explicit knowledge, and the application of that knowledge in the transfer phase. One reason for this, and the lack of transfer in Experiment 3, may be that participants in the Hint group were more aware of the apparent direction of motion. This may have meant that they were more sensitive to the removal of the contingencies in the transfer phase, and this sensitivity reduced the size of their displayed cueing effect. This is consistent with the results of Jiménez, Vaquero, & Lupiáñez (2006), who found more decrement in responding in an intentional group in transfer blocks (where the contingencies were removed) than an incidental group. They explained this counterintuitive result by claiming that participants who learn about a sequence explicitly may notice a change in the sequence and actively try to suppress their previous knowledge to learn about the novel sequence. While better transfer was not obtained in the No Hint group than in the Hint group, a similar case can be made for why a benefit in transfer cueing was not found for the Hint group. Giving participants explicit sequence knowledge seemed to have little effect in transfer, and highlights again that specific direction and obvious applicability need to be present before explicit knowledge can benefit sequence learning. While it can be concluded that probabilistic sequence learning does not seem to be independent of explicit knowledge or the intention to learn, it does appear that encouraging participants to use this knowledge effectively is surprisingly difficult, in contrast to the robust cueing effects obtained (in both training and transfer) under incidental conditions. In this way, sequence learning may represent an example of a task where explicit knowledge is able to affect learning, but only to a limited degree.
2.8.3 Implicit Learning?

It should be noted that while explicit knowledge had an effect on cueing in these experiments, this does not rule out the possibility that under incidental conditions, implicit learning occurs. Reed and Johnston (1994) have argued that the mere presence of explicit learning (defined in terms of participants' awareness) in one sequence structure does not imply that all sequences will be learned explicitly and does not contradict evidence for implicit or automatic learning processes derived from another sequence structure. Indeed, the robust cueing effects obtained in the absence of the hint may indicate the operation of an automatic implicit learning process that contributes to the cueing effect in both groups. Interpreting the group differences as due to the addition of explicit knowledge or learning in fact assumes this already. The results do not imply that all probabilistic sequences would be affected by explicit knowledge, and does not weaken the existing evidence that learning of complex probabilistic sequences is implicit in some sense.

What the current results do suggest is that there is no general divide between deterministic and probabilistic sequences in the sense that the former are sensitive to explicit knowledge whereas the latter are resistant to this knowledge. Instead, they suggest that tasks in which there is a practical advantage in using the explicit knowledge are likely to show sensitivity to such knowledge. Situations in which explicit knowledge conveys little practical advantage will not show that sensitivity. That practical advantage may be a function of several factors including the general reliability of the given rule, which is inherently weaker in probabilistic compared to deterministic sequence structures, but also the participant's ability to implement the rule while concurrently performing the task. This conclusion assumes that rule use is effortful and requires drawing from a limited pool of cognitive resources. Thus,
studies that attempt to make theoretical conclusions on the basis of the cognitive penetrability of sequence learning need to consider both the nature of the knowledge supplied to participants, and its utility to the particular sequence to be learned in order to predict the effect of additional explicit knowledge.

2.8.4 Future Directions

The current set of results and previous literature show that there are situations in which sequence learning is both susceptible, and resistant to explicit knowledge. Therefore, perhaps the best avenue for future research is to investigate the parameters that predict whether explicit knowledge will impact on sequence learning. A comparison of learning under incidental and intentional conditions using a variety of sequences will be more informative in investigating whether there are any differences between implicit and explicit learning. This stance allows one to maintain the assumption that implicit learning is an automatic process that operates differently to explicit processes in important ways. For instance, even if implicit and explicit processes are not fully independent of one another (in the sense that what is learned implicitly can still be affected by explicit processes and/or become available to conscious awareness), a subset of processes might still be isolable because they continue to operate effectively in situations where explicit mechanisms cannot (see Cleeremans & Jiménez, 1998 for a similar argument). This would entail that the specific requirements of the task will dictate whether learning will proceed implicitly or explicitly, with learning in most situations perhaps being a mixture of both (Reed & Johnson, 1994).
2.8.5 Conclusion

In conclusion, consistent evidence of SRT performance advantages was found for participants given relational knowledge about a simple probabilistic sequence, and robust cueing effects in the absence of this knowledge were also found. Probabilistic sequence learning as measured in this task was not found to be resistant to explicit knowledge, but also seems to proceed in the absence of explicit knowledge or intention to learn. Learning of these probabilistic sequences, just like deterministic sequences, is affected by explicit information given at the start of learning, with the findings contradicting the notion of sequence learning being implicit in the sense of being independent from volitional control. On the other hand, substantial effort was needed to ensure that a general performance benefit (i.e. one affecting all subsequences) was gained from this knowledge, and transfer to similar tasks with the same relational structure was seemingly no stronger than that shown by participants not given the hint. This was despite the relational knowledge being abstract from the outset and therefore directly applicable to all subsequences and both transfer tasks. It seems that the benefits of explicit relational knowledge in the SRT task are very sensitive to the precise nature of the knowledge given, and rely on their applicability to the task at hand being made obvious. A fruitful endeavor for future research may be to characterize the conditions under which explicit knowledge will aid sequence learning, and look for qualitative differences in the particular subsequences that benefit from this explicit learning, rather than attempting to show that there are conditions under which explicit knowledge has no effect.
Chapter 3: The Prototype Distortion Task

3.1 Introduction

A key debate within the category learning literature pertains to whether some kinds of category structures can be learned implicitly and whether they recruit independent neural structures to explicit learning (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby & Maddox, 2005; 2011; Poldrack & Foerde, 2008; Smith & Grossman, 2008). One paradigm that is central to the debate is the A/not A prototype distortion task. In this task, participants are exposed to a set of exemplars that are generated by distorting a prototype, creating a category of stimuli that are physically similar to each other and whose average is the prototype. Learning about the category (A) is evident if participants can subsequently distinguish between novel exemplars and foils (make accurate A/not A judgements), or if the prototype is given the highest category endorsement on test (e.g. Posner & Keele, 1968). Alternatively, learning may be evident if, when judging novel exemplars of varying distortion levels, category endorsements or recognition judgements vary as a function of the similarity of the exemplars to the category prototype. This pattern of generalization, referred to as a prototypicality gradient, is present when the highest category endorsement or recognition is given for low distortions of the prototype, and lowest endorsement or recognition is given for high distortions of the prototype. These effects (collectively referred to as prototype effects) can all be taken as evidence that participants have learnt something about the similarity structure of the category. Learning in the prototype distortion task has been labeled as implicit due to seemingly intact categorization performance in amnesic patients, coupled with the incidental nature of the exposure conditions, each of which will now be reviewed.
3.1.1 Intact Category Learning in Memory-Impaired Patients

Prototype effects are often regarded as implicit due to several striking studies that have reported intact categorization in memory-impaired patients (Bozoki, Grossman, & Smith, 2006; Kéri et al., 1999; 2001; Knowlton & Squire, 1993; Reed, Squire, Patalano, Smith, & Jonides, 1999). The earliest study demonstrating this was by Knowlton and Squire (1993), who measured categorization and recognition performance in amnesic patients and healthy controls after being exposed to a series of dot patterns centered around a prototype (for a description of these stimuli, see Posner, Goldsmith, & Welton, 1967; Posner & Keele, 1968). Amnesic and control participants were asked to point to the dot closest to the middle of each stimulus and were then told to use any knowledge they had acquired in a subsequent categorization and recognition test. While control participants were able to discriminate between category exemplars that they had previously seen, and category exemplars that were novel on test, amnesic patients were impaired at discriminating between new and old exemplars and yet still displayed categorization performance equivalent to the normal controls (see also Knowlton, Mangels, & Squire, 1996, for a similar finding in probabilistic category learning; Knowlton, Ramus, & Squire, 1992, for a similar finding in artificial grammar learning; and Squire & Knowlton, 1995, for a related case study). A similar result was found by Kéri et al. (2001, see also Kéri et al., 1999) using the same dot pattern stimuli, with impaired explicit recognition but equivalent A/not A categorization performance in patients with mild to moderate Alzheimer’s Disease relative to a control group. These studies (see also Bozoki et al., 2006; Reed et al., 1999, for similar results with different stimuli) seem to provide evidence for a dissociation between categorization and recognition memory, suggesting the existence
of a separate system that can learn about categories in the absence of declarative knowledge.

However, there are some issues with these studies that limit their interpretation. The original demonstration by Knowlton and Squire (1993), while compelling, has been criticized on the basis of using stimuli that generate a ‘false’ prototype effect, where a prototypicality gradient can result even in the absence of exposure (Nosofsky, Denton, Zaki, Murphy-Knudsen, & Unverzagt, 2012; Palmeri & Flanery, 1999; Zaki & Nosofsky, 2004). These studies suggest that participants are relying on working memory and learning about the category on test in an explicit fashion (Palmeri & Flanery, 1999). This learning-at-test effect may be especially pronounced if the test phase contains a large proportion of low distortion exemplars and foils which make the contrast between categorical and non-categorical stimuli obvious (Nosofsky et al., 2012), or if the test phase contains a large number of low distortion exemplars and prototype presentations which could highlight the prototypical features of the category (Zaki & Nosofsky, 2004).

Another issue concerns the small sample sizes that these studies typically use (e.g. Knowlton & Squire, 1993; Reed et al., 1999). Since the dissociation requires a null effect, that performance on categorization is equivalent between amnesic and control groups, a small sample size means that there will be less power to detect a possible group difference if one exists. However, a large difference in recognition accuracy that naturally results from comparing memory-impaired participants with healthy participants might still be detected. Zaki (2004) confirmed this hypothesis in a meta-analysis on 12 studies that compared memory-impaired participants against controls on performance in various category learning tasks, a subset of which were prototype distortion tasks. Contrary to the suggestions of the studies mentioned above,
it was concluded that there was indeed an overall impairment in categorization performance in amnesic patients, consistent with the idea that prototype effects are dependent on declarative memory and underpinned by a single memory system.

Assuming that the intact prototype effects in amnesic patients are in fact due to exposure, Nosofsky and colleagues (Nosofsky & Zaki, 1998; Zaki, Nosofsky, Jessup, & Unverzagt, 2003) propose that the dissociations discussed above can still be explained using a single system of memory. Dissociation can be predicted even if a common ability (e.g. sensitivity in discriminating between distinct exemplars in memory) underlies both categorization and recognition tasks if we assume that this ability is deficient in amnesic patients, and that this affects recognition judgements more than categorization judgements. In other words, there is a parameter difference between tasks responsible for the observed dissociations whereby changes in memory sensitivity between groups results in a large impairment in recognition and a small impairment in categorization (see Berry, Henson, & Shanks, 2006 for a related single-system view). This idea has been supported by the finding that after a delay of one week between exposure and test, participants’ ability to discriminate between new and old exemplars was impaired while categorization accuracy was unaffected (Nosofsky & Zaki, 1998). This dissociation in healthy participants supports the idea that, even when assuming a single memory system, changes on one measure do not necessarily entail changes on the other. In summary, while the studies with amnesic populations are striking, the methodological issues concerning the stimuli and the ability of single-system theories to explain dissociations undermine the conclusion that learning in the prototype distortion task is implicit.
3.1.2 Emergence under Incidental Learning Conditions

A less-cited reason that prototype effects can be considered implicit is that experiments demonstrating this effect have tended to expose participants to category exemplars under conditions where they are not encouraged to deliberately encode the stimuli. Most studies require participants to view the stimuli passively, such as thinking about their appearance (Bozoki, Grossman, & Smith, 2006), or perform a task that incidentally exposes them to the stimuli such as pointing to the dot closest to the center of each stimulus (Knowlton & Squire, 1993). It is often assumed that because these conditions do not explicitly mention the existence of a category that any learning that occurs must be incidental. However, these passive viewing situations do not preclude the possibility that participants are intentionally encoding the stimuli in some way during this initial phase assuming (correctly) that they will later be useful. Classifying these exposure conditions as incidental would be more convincing if it could be ensured that participants’ explicit cognitive functions were more fully engaged by performing a more difficult task during the exposure phase. A prototypicality gradient under these conditions would suggest that participants can learn about categories of stimuli that are physically similar to one another in an automatic or incidental fashion, satisfying one of the defining characteristics of implicit learning (Cleeremans & Jiménez, 1998; Frensch, 1998; Stadler & Frensch, 1994). One way in which to achieve incidental exposure is suggested by another implicit learning paradigm, namely contextual cueing in visual search.

In a contextual cueing task (Chun & Jiang, 1998), participants search through a configuration of distractors (usually rotated letter L’s) for a target (usually a rotated letter T). Unbeknownst to participants, certain configurations appear multiple times throughout the experiment. Participants show reliable reductions in reaction time in
response to repeatedly presented (old) configurations when compared to novel configurations during training, yet sometimes fail to explicitly recognize old configurations in a subsequent test, or accurately generate the correct location of the target when presented with the old distractor contexts (Chun & Jiang, 1998; 2003). While explicit knowledge is sometimes present (e.g. Smyth & Shanks, 2008), and several studies can be criticized on the basis of lacking sensitivity in the recognition test in comparison to the large number of trials involved in training (Vadillo, Konstantinidis, & Shanks, 2016), several researchers have concluded that learning via visual search in this paradigm is implicit and independent of explicit knowledge (e.g. Chun & Jiang, 2003; Goujon, Didierjean, & Thorpe, 2015). Contextual cueing effects may reflect cueing of the target location by the surrounding visual context (Chun & Jiang, 1998), or facilitation in detection or responding due to learning associations between distractors for repeated stimuli (Beesley, Vadillo, Pearson, & Shanks, 2015; Kunar, Flusberg, Horowitz, & Wolfe, 2007). While there is ongoing debate about the mechanisms responsible for contextual cueing effects as well as the role of awareness, it is clear that the learning that occurs is incidental to the task being performed. Thus, employing a visual search task that requires a correct response should sufficiently engage participants such that any learning that does occur can be confidently deemed to be incidental, in contrast to a task such as pointing to the middle dot where it is usually difficult to assess whether participants are performing the task properly.

Despite the theoretical significance of incidental learning conditions in establishing prototype effects as implicit, there have been few attempts to compare different methods of exposure in the prototype distortion task. One notable exception is a study by Gureckis, James, and Nosofsky (2011, see also Reber, Gitelman, Parrish, & Mesulam, 2003), who compared implicit and explicit learning conditions as well as
an intentional encoding strategy against an incidental encoding strategy in a 2x2 between-subjects design. Participants were either aware (explicit group) or unaware (implicit group) of the existence of a category, and were either asked to study the stimulus as a configural whole (configural group) or imagine pointing to the center dot (dot group). At test, they found differential activation (comparing exemplars and foils) in the posterior occipital cortex as a function of encoding strategy (configural vs. dot). Meanwhile, telling participants about the existence of a category (implicit vs. explicit) had no consistent effect. They concluded that differential brain activation found in studies comparing implicit and explicit learning orientations (e.g. Reber et al., 2003) were better explained through different encoding strategies, rather than the implicit status of the learner or awareness of impending tests.

While this is an appropriate conclusion from Gureckis et al.’s (2011) results, it should be noted that all four groups produced an equivalent level of categorization accuracy (there were no significant main effects nor interactions), suggesting that while the encoding manipulation was successful in encouraging participants to adopt different strategies, there was no evidence that this made a difference to actual test performance. Thus, while their study suggests that participants were obeying instructions to perform the pointing task when asked to, the issues of whether intentional encoding conditions have an advantage over incidental encoding conditions, and whether other incidental encoding conditions are conducive to prototype effects, remain unresolved.

3.1.3 General Aims and Methodology

The aims of the current experiments were twofold. The first was to assess the implicit status of learning in the A/not A version of the prototype distortion task by
testing whether a prototypicality gradient could result when participants performed a visual search task with the category exemplars. Appropriating a visual search paradigm similar to that used in contextual cueing allows for this often-assumed implicit property of prototype effects to be tested in a novel way. The second aim was to directly compare two methods of exposure: intentional memorization of a set of prototype-centered stimuli for a subsequent memory test (Group Memorize), and a visual search task using the same stimuli (Group Search). This comparison is important since different studies use different methods of passive exposure, with the assumption that because participants are not informed about the existence of a category, then any learning that occurs must be incidental and in some sense equivalent across passive exposure conditions. Comparing a visual search group to a group who are given direct instructions to memorize the stimuli allows the best chance of detecting potential differences between incidental and intentional encoding conditions, if they exist.

A novel aspect of the methodology in this chapter was that instead of measuring categorization and recognition in separate tests, a single continuous measure (familiarity ratings) was chosen to assess both. Learning about the category would thus be evident if participants show a generalization gradient (i.e. a prototypicality gradient), and recognition inferred if participants are able to discriminate between new and old test items at matched levels of distortion in their familiarity ratings. While this is a departure from the majority of studies on the prototype distortion task, it was motivated by the desire to test whether dissociations between generalization and recognition performance were still possible when the conditions of the two key comparisons were fully equated for possible transfer deficits caused by the change in context moving from exposure to test, as well as any
forgetting caused by delay or interference from testing itself. Using the same measure should reduce the amount of task-specific variance and might therefore result in similar group differences in generalization and recognition rather than a dissociation (see Nosofsky & Zaki, 1998).

For the following experiments two novel sets of stimuli were constructed to mimic the statistical properties of the dot patterns with the aim of minimizing potential learning-at-test effects. While the dot patterns are well-established in the prototype distortion literature in terms of their known statistical properties and ability to produce prototypicality gradients, testing a novel set of stimuli may provide insight into whether learning-at-test effects are primarily due to the demands of the task, or the dot patterns themselves. Experiment 1 produced similar generalization gradients between participants who received a categorization test and participants who received a familiarity test, justifying the use of familiarity ratings only for the subsequent experiments. Experiment 1 also compared generalization gradients between groups who were not exposed to the stimuli (but led to believe that they were) against groups who did observe the stimuli. The magnitude of any prototypicality gradients exhibited after no exposure would indicate the degree of learning-at-test effects, which would then serve as a point of comparison to determine whether participants in subsequent experiments learned anything during visual search (see Smith, 2008, for a discussion of this subtraction logic). Experiments 2 and 3 compared prototypicality gradients for new and old test stimuli in a subsequent familiarity test. Experiment 3 doubled the length of exposure from Experiment 2 and added an additional visual search group where the stimulus exposure terminated after the response was made to ensure that participants had no residual exposure time to study the stimulus.
3.2 Experiment 1

The primary aim of Experiment 1 was to test for any potential learning-at-test effects with the two novel stimulus sets to use as a basis for comparison for the subsequent experiments. Previous studies have demonstrated that the dot pattern stimuli that are typically used generate ‘false’ prototype effects in normal participants who do not see the stimuli prior to the test phase (Palmeri & Flanery, 1999; Zaki & Nosofsky, 2004). Participants in these studies undergo a mock-subliminal procedure, where they are led to believe that stimuli are being presented to them subliminally but in fact never see any stimuli. Despite the stimuli being novel on test, they subsequently show a prototypicality gradient in their categorization judgements when presented with high and low distortion exemplars. One way that participants can show false prototype effects is by learning about the category during the test phase, since this usually involves informing participants about the existence of a category, and then exposing them to more category exemplars, as well as the prototype. The degree of learning-at-test effects is especially pronounced if the test phase were to present a large number of low-distortion exemplars and foils on test, highlighting the contrast between categorical and non-categorical stimuli (Nosofsky et al., 2012), or when the prototype and low-distortion exemplars are presented multiple times during test which would result in high repetition of prototypical features (Zaki & Nosofsky, 2004). For this reason, to minimize potential learning-at-test effects new and old exemplars were tested at matched levels of distortion with no foils, and the mock-subliminal procedure was used to test whether participants produced prototypicality gradients in the absence of exposure.

A potential explanation for the observed dissociations in amnesic patients discussed above is that categorization and recognition tests have different task-
specific variance (Nosofsky & Zaki, 1998). Usually, categorization accuracy is assessed by asking participants to make an A/not A category endorsement, and recognition is assessed by asking participants to make an old/new judgement about whether they have seen a particular stimulus before. Participants may have different response thresholds for categorization and recognition judgments, making an equivalent comparison between the two difficult. For example, participants may be more willing to classify a test stimulus as part of the seen category than to say that they recognize the stimulus based on the same level of uncertainty, or adopt a more stringent criterion for endorsing recognition since an exact match is required (Nosofsky, Little, & James, 2012). An equivalent comparison between recognition and categorization tasks is made more difficult when we consider that previous studies have typically measured the prototype effect using categorization judgements, and separately measured recognition in a completely different phase using different stimuli (e.g. Bozoki et al., 2006; Knowlton & Squire, 1993; Reber & Squire, 1999). For this reason, it was important to test whether a dissociation could still occur when the stimuli were equated between groups, and also the test measure. Thus another aim of Experiment 1 was to assess category learning (as indexed by prototypicality gradients) and recognition (as indexed by discrimination between new and old exemplars) using both categorization and familiarity tests and to show that they were equivalent between test measures.
3.2.1 Method

3.2.1.1 Participants

Ninety-two ($M = 19.97$, $SD = 3.38$, 70 female) University of Sydney students participated in this experiment in exchange for partial course credit$^{16}$. Participants were randomly allocated to either a mock ‘subliminal’ (NoEx) or passive exposure (Ex) phase, and either a familiarity (Fam) or categorization (Cat) test phase ($n = 23$ in each of the four groups).

3.2.1.2 Apparatus

All experiments were programmed using PsychToolbox for Matlab (Brainard, 1997; Pelli, 1997) and run on Apple Mac Mini desktop computers connected to 17 inch CRT monitors, refreshed at a rate of 85 Hz. A standard Apple keyboard and mouse were used, and testing was conducted in individual cubicles in groups of up to five. The apparatus was the same for all subsequent experiments.

3.2.1.3 Stimuli

The stimuli were constructed to be complex and have multiple dimensions so that focus on a particular feature or an attempt to derive a verbalizable rule to describe the category would be difficult and less useful than memorizing the whole pattern. There were two sets of stimuli constructed for these experiments, each containing 10 features (circles or lines). For the circle stimuli, there were 10 colored circles on a 600 x 600 pixel black square. Each circle had the following variable properties: hue (saturation and brightness were held constant at their maximum respective values), location (x and y coordinates) within the black square, line thickness, and size (see

$^{16}$ The target sample sizes for this, and subsequent experiments was 25 per group. This was based on pilot experiments conducted to test the general paradigm. The final sample size was affected by availability of participants.
Figure 3.1. For the line stimuli, there were 10 white lines (5 oriented horizontally and 5 vertically) on a black square of the same size (see Figure 3.1). Each line was defined by its starting and ending location on the border of the square, as well as line thickness. Maximum and minimum values were chosen for each property to constrain the possible exemplars that could be created and to avoid placing the circles and lines outside the stimulus boundary (see Table 3.1).

![Figure 3.1](image)

*Figure 3.1. Examples of exemplars from each stimulus set (upper panels: circle stimuli; lower panels: line stimuli) used throughout all experiments. From left to right: prototype (0 distortion), a low distortion exemplar (0.1 distortion), a medium distortion exemplar (0.5 distortion), and a high distortion exemplar (1.0 distortion).*

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Min Value</th>
<th>Max Value</th>
<th>Multiplier</th>
<th>Dist to prototype</th>
</tr>
</thead>
<tbody>
<tr>
<td>Circle location (x)</td>
<td>60 pixels</td>
<td>540 pixels</td>
<td>40</td>
<td>20</td>
</tr>
<tr>
<td>Circle location (y)</td>
<td>60 pixels</td>
<td>540 pixels</td>
<td>40</td>
<td>20</td>
</tr>
<tr>
<td>Circle hue</td>
<td>None</td>
<td>None</td>
<td>.05</td>
<td>.1</td>
</tr>
<tr>
<td>Circle radius</td>
<td>5 pixels</td>
<td>50 pixels</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>Line start (x)</td>
<td>50</td>
<td>550</td>
<td>50</td>
<td>20</td>
</tr>
<tr>
<td>Line start (y)</td>
<td>50</td>
<td>550</td>
<td>50</td>
<td>20</td>
</tr>
<tr>
<td>Line end (x)</td>
<td>50</td>
<td>550</td>
<td>50</td>
<td>20</td>
</tr>
<tr>
<td>Line end (y)</td>
<td>50</td>
<td>550</td>
<td>50</td>
<td>20</td>
</tr>
</tbody>
</table>

*Table 3.1. Minimum and maximum values and multiplier for each stimulus feature dimension.

Note. Location values could range from 0 to 600, and the saturation and brightness were always set at 1 and 255 respectively. For vertical lines, only the x values were varied (y values were held constant at 0 and 600) and for horizontal lines, only the y values were varied (x values were held constant at 0 and 600). Hue values wrapped around 0-1 since the scale was continuous. For each dimension of each feature, if the average of the exposed exemplars was not within the specified maximum distance to the prototype, all exemplar values were created again until this condition was met.
All variable feature properties listed above were varied from exemplar to exemplar except for line thickness, which was held constant at 3 pixels in both stimulus sets. Arbitrary distortion levels were chosen ranging from 0.1 to 1.0 in increments of 0.1, with the smaller levels of distortion indicating exemplars that were similar to the prototype and higher levels indicating category exemplars that were dissimilar to the prototype. While the distortion level numbers are arbitrary, they represent a ratio scale with, for example, a distortion level of 0.1 being 10% of the distortion level of 1.0. Appropriate dimension multipliers were chosen for each feature dimension which would create a dissimilar stimulus when multiplied by the higher distortion levels but a highly similar stimulus when multiplied by the lowest distortion level (see Figure 3.1 for examples).

A different circle and line prototype stimulus was randomly generated according to a seed number, which was different for each participant number. Thus, participants with the same participant number were exposed to exactly the same stimuli throughout the experiment. Each category exemplar was created by distorting the relevant prototype stimulus on a feature-by-feature basis. This meant that the feature multiplier (see Table 3.1) was multiplied by the distortion level (e.g. 0.1) for each of the 10 features (circles or lines) separately. The direction of change (positive or negative) was determined randomly and independently for each dimension of each feature. For example, to create a circle exemplar at 0.1 distortion, the size multiplier was multiplied by 0.1, and then either added or subtracted (randomly chosen for each feature) to the prototype values of the first circle, with this process repeating for the remaining circles, and this process repeating again for each of the circle features (location, hue etc.). If any feature values extended beyond the minimum or maximum
values they were simply changed to either the minimum or maximum value they
crossed (see Table 3.1).

Twenty circle exemplars and twenty line exemplars were shown during the
initial exposure phase (2 unique exemplars at each of 10 levels of distortion for each
set). A further twenty novel circle exemplars and twenty novel line exemplars were
also shown during the test phase (again with 2 exemplars at each level of distortion).
To ensure that participants could not rely on occasional salient features (e.g.
clustering of circles in a fashion that leads to occlusion by overlapping circles) within
the stimulus to aid memory, additional checks were implemented to ensure that there
would be minimal overlap between the circles (at least 50 pixels between the center of
all circles). If this check failed then all circle locations were randomized until this
condition had been met. Note that overlap and occlusion was a regular feature of the
line patterns and thus this check was not implemented for these stimuli since it would
be difficult to use a specific conjunction to aid memory for a specific stimulus. In
addition, to ensure that the average of the exposed exemplars was indeed the
prototype, the average value of the 20 exemplars was computed for each feature
dimension (e.g. x coordinate, size, starting position of lines etc.). If the average was
not within a pre-determined maximum distance from the prototype, all stimulus
values were randomized again until they were sufficiently close (see Table 3.1 for the
maximum distance allowed for each feature dimension).

3.2.1.4 Procedure

The experiment was a 2 x 2 between-subjects design, with exposure (exposure
vs. no exposure) and test (categorization vs. familiarity) as the independent variables.
Participants allocated to the exposure phase saw 20 unique exemplars from each of
the circle and line categories 4 times each, with presentation randomized within each block of 40 trials (4 blocks, 160 trials in total). Participants allocated to the subliminal exposure phase were told that they would be presented with some subliminal stimuli, which would be quickly masked by a black square. They were told that because the presentation was so brief, all they might see is the screen flash before the mask covered the stimulus. In the (sham) exposure phase that followed, no stimuli were presented, but on each trial, the screen would flash from white to black and after 10msecs return to white, along with the black stimulus background (i.e. the ‘mask’) presented for the same time (2 seconds) as in the actual exposure phase. In the two exposure groups, the stimuli appeared on screen for 2 seconds before disappearing. The inter-stimulus-interval (ISI) was a blank white screen, and was 2 seconds in both exposure conditions.

After the exposure phase participants either completed a familiarity or categorization test. Participants allocated to the familiarity test were told that they would be presented with more stimuli, where some of the stimuli would be familiar (i.e. they had seen them before in the first phase) and others novel (they had not seen them in the first phase). They were then asked to rate their level of familiarity towards 84 new exemplars (for each category: the prototype presented twice, 20 old and 20 new exemplars) using a visual analogue scale that ranged from “Definitely NOT familiar” to “Definitely familiar”. If participants were allocated to the categorization test, they were told that all of the stimuli they had seen formed part of a category and that they would be presented with more stimuli that may or may not be part of the same category. For each stimulus, participants were asked “Is this stimulus part of the circle category you saw before?” or “Is this stimulus part of the line category you saw before?” It was made clear that participants should only consider the circle category
when presented with a circle stimulus and similarly for the line category. Participants made their rating on a visual analogue scale that ranged from “Definitely NO” to Definitely YES”.

In both tests each stimulus was presented for 2 seconds and then disappeared. After 500msecs a rating scale would appear along with the appropriate test question. Participants made their rating by clicking a point on the scale with the mouse and had unlimited time to make and change their rating. It was made clear that all questions referred to stimuli they had seen in the first part of the experiment. The midpoint of the scale and both endpoints were marked with ticks, and all ratings were transformed to range from 0-100. The familiarity and categorization tests ran the same way no matter what exposure phase participants were allocated to. For this, and all subsequent experiments, the same set of seed numbers for the random number generator were used in each group, such that the stimuli seen and tested were randomized but matched between groups of participants. This also meant that old and new test stimuli were dummy coded for the no-exposure groups. There was no feedback during the test phase.

3.2.2 Results and Discussion

For all subsequent analyses Greenhouse-Geisser corrections to p-values are reported for violations of sphericity, and the results reported combine both circle and line stimulus sets\(^{17}\). To analyze whether the prototype effect, and recognition, were dependent on exposure and varied with test question, a 2 (exposure group: exposure vs. no exposure) x 2 (test group: categorization vs. familiarity) x (2) (novelty: new vs. old) x (10) (distortion level) ANOVA was performed on the exemplar ratings

\(^{17}\) Analyses on the familiarity ratings were also performed on the circle and line stimulus sets separately, and produced similar results for the critical findings. Instances where the results differed between stimulus sets have been noted.
(excluding the prototype) in the test phase. Figure 3.2 shows the mean ratings for the test stimuli in each of the four groups.

![Graphs showing category endorsement and familiarity ratings for new and old test stimuli for groups of participants who were either exposed to the stimuli or not exposed to the stimuli in Experiment 1.]

Figure 3.2. Category endorsement (upper panels) and familiarity ratings (lower panels) for new and old test stimuli for groups of participants who were either exposed to the stimuli (left panels) or were not exposed to the stimuli (right panels) in Experiment 1.

There was a main effect of novelty, $F(1,88) = 18.85, p < .001, \eta_p^2 = .176$, which did not interact with test group, $F < 1$, indicating that overall, participants rated old test stimuli higher on category endorsement and familiarity than new test stimuli. Unsurprisingly, novelty interacted with exposure group, $F(1,88) = 17.97, p < .001, \eta_p^2 = .170$, such that participants who had been exposed to the exemplars could discriminate between new and old test items (Figure 3.2, left panels), but participants who had not seen the stimuli could not (Figure 3.2, right panels). This was confirmed in a set of separate analyses where there were significantly higher familiarity ratings,
\[ F(1,22) = 10.18, p = .004, \eta_p^2 = .316, \] and category endorsements, \[ F(1,22) = 15.24, p = .001, \eta_p^2 = .409, \] for old exemplars in the two exposure groups, and no differences between new and old exemplars in the two no-exposure groups, largest \( F < 1. \)

There was a significant main effect and linear trend for distortion level, smallest \( F(9,792) = 45.79, p < .001, \eta_p^2 = .342, \) since overall ratings declined as the level of distortion increased, indicating the presence of a prototypicality gradient. The main effect and linear trend of distortion interacted with exposure group, smallest \( F(9,792) = 13.28, p < .001, \eta_p^2 = .131, \) but not test group, largest \( F(9,792) = 1.93, p = .090, \eta_p^2 = .021, \) nor novelty, largest \( F(1,88) = 3.58, p = .062, \eta_p^2 = .039. \) Thus it appears that the prototypicality gradient was stronger (i.e. there was a steeper gradient) when participants were exposed to the stimuli, but did not vary according to the type of test conducted, nor between new and old exemplars.

None of the 3-way interactions nor the 4-way interaction were significant, largest \( F(9,792) = 1.79, p = .079, \eta_p^2 = .020. \) There was a significant main effect of exposure group, \( F(1,88) = 20.37, p < .001, \eta_p^2 = .308, \) and test group, \( F(1,88) = 5.13, p = .026, \eta_p^2 = .055, \) since overall ratings were higher for participants who had been exposed to the exemplars, and ratings for the categorization test were generally higher than for the familiarity test. This may have been due to the fact that a few participants in the no-exposure group rated all test stimuli as 0 on familiarity (Definitely not familiar), while an equivalent number of participants in the categorization test rated all stimuli as 50, indicating a noncommittal response. Ratings for the category prototype were significantly higher for the exposure groups than no-exposure groups, \( F(1,88) = 35.50, p < .001, \eta_p^2 = .287, \) but did not differ according to test, \( F(1,88) = 2.22, p = .140, \eta_p^2 = .025. \) There was also no significant interaction between exposure group and test group, \( F < 1. \)
In summary, it appears that overall, participants were able to discriminate between new and old exemplars, and produce prototypicality gradients, and these effects depended on whether participants underwent the mock-subliminal or exposure phase, while there seemed to be no effect of test question. To further explore what participants in the no-exposure groups learnt and whether equivalent prototypicality gradients were obtained for categorization and recognition tests, the two no-exposure groups were analyzed separately to the two exposure groups in two 2 (test group: categorization vs. familiarity) x (2) (novelty) x (10) (distortion level) ANOVAs.

3.2.2.1 Exposure Groups

For the two groups who were exposed to the exemplars (Ex-Fam and Ex-Cat), there was a significant main effect of novelty, $F(1,44) = 25.09, p < .001, \eta_p^2 = .363$, and significant main effect and linear trend for distortion level, smallest $F(9,396) = 46.00 p < .001, \eta_p^2 = .511$. None of these effects interacted with test group, largest $F(9,396) = 1.53, p = .179, \eta_p^2 = .033$, and there was no main effect of test group, $F < 1$, and no 3-way interaction, $F < 1$. There was, however, a significant interaction between novelty and distortion, $F(9,396) = 2.14, p = .042, \eta_p^2 = .046$, but since novelty did not interact with the linear trend for distortion, $F(1,44) = 1.89, p = .176, \eta_p^2 = .041$, this result is difficult to interpret. Thus, it is reasonable to conclude that equivalent prototypicality gradients were found for both familiarity and categorization tests, and participants were equally able to discriminate between new and old exemplars, giving higher familiarity ratings and category endorsements to old exemplars (Figure 3.2, left panels). The high level of similarity in prototypicality gradients for the categorization and familiarity tests suggests that using familiarity
ratings to assess the prototypicality gradient is appropriate for the subsequent experiments.

3.2.2.2 No-Exposure Groups

A similar analysis was performed on the ratings for the two no-exposure groups. There was a significant main effect and linear trend for distortion level, smallest $F(9,396) = 6.77, p < .001, \eta^2_p = .133$, and a significant main effect of test group, $F(1,44) = 5.34, p = .026, \eta^2_p = .108$. Importantly, the linear trend for distortion did not interact with test group, $F(1,44) = 2.90, p = .095, \eta^2_p = .062$. To quantify the magnitude of this important null result, as suggested by Rouder, Speckman, Sun, Morey, and Iverson (2009), a Bayes Factor (BF) was calculated from the results of a t-test using a JZS prior for the alternative hypothesis, which assumes a Cauchy distribution of effect sizes. This distribution assumes a high likelihood of smaller effect sizes but a larger likelihood of medium-to-large effect sizes than a normal distribution. A $BF_{01}$ of 3 is usually considered the threshold for concluding moderate evidence in favor of the null over the alternative hypothesis (that the means are different). Using this technique, a BF of 4.02 in favor of the null was obtained, indicating moderate evidence that similar prototypicality gradients were obtained between categorization and familiarity tests.

All other main effects and interactions were not significant, largest $F(9,396) = 1.99, p = .107, \eta^2_p = .043$. It appears that the mock subliminal procedure produced a false prototypicality gradient in familiarity ratings and category endorsement, similar to previous studies (e.g. Palmeri & Flanery, 1999). Since an effort was made to ensure that no intrinsic features within the stimuli would enable participants to discriminate between low and high distortion exemplars, the most likely explanation was that
participants were able to learn about the category on test\textsuperscript{18}. Participants may have either deliberately based their judgements on stimuli seen during the test phase in order to have some basis for varying their responses, or found it difficult to discount the stimuli they had seen throughout the test phase.

The mock subliminal procedure appears to have been quite convincing. When asked at the end of the experiment, participants allocated to the categorization test claimed to have seen, on average, 12.04\% of the stimuli ($SD = 23.93$, $min = 0$, $max = 99.2$) during the mock subliminal phase, and participants allocated to the familiarity test claimed to have seen 15.00\% of the stimuli on average ($SD = 29.87$, $min = 0$, $max = 99.60$). While it is possible that a small number of participants misinterpreted the question, the manipulation check indicates that on the whole, participants were convinced that stimuli had been presented to them in the first phase and this should have provided sufficient motivation to make a serious attempt at judging the stimuli presented in the test phase.

In summary, the categorization and familiarity tests produced very similar prototypicality gradients and differences between ratings for old and new exemplars. Thus, the following experiments used familiarity ratings only to assess both of these effects. The familiarity test was chosen over the categorization test to more closely approximate other implicit learning paradigms that use recognition to assess explicit awareness of the learned material, and because the test was directly related to the instructions given to the Memorize group (see Appendix C). Using a familiarity test means that any difference in ratings between new and old stimuli is directly interpretable as recognition, allowing a test of whether it is possible to obtain a

\textsuperscript{18} The false prototypicality gradient was present in the first quarter, and increased from the first to the second quarter when collapsed over groups, $F > 10$. Therefore, learning-at-test effects appeared quite early.
prototypicality gradient in the absence of recognition (as demonstrated in amnesic patients, e.g. Knowlton & Squire, 1993).

3.3 Experiment 2

Experiment 2 compared prototypicality gradients in familiarity ratings between intentional exposure conditions where participants were told to memorize a set of stimuli for a subsequent test (Group Memorize), and incidental exposure conditions where participants performed a visual search task using the same stimuli (Group Search). Participants were required to search for a singleton target, which was defined by line width (see Figure 3.3 for examples). To ensure that participants were attending to the stimuli in the Search group, they were required to respond according to the identity of the line width of the singleton (i.e. was the ‘odd one out’ thicker or thinner than the others?).
Figure 3.3. Examples of circle and line stimuli seen during the visual search task. Stimuli on the left contain a circle/line randomly chosen to have a thinner line width, stimuli on the right contain a circle/line randomly chosen to have a thicker line width. Participants responded according to the identity of the singleton (whether it was thicker or thinner).

Experiment 2 tested whether the prototypicality gradient would result under more cognitively engaging incidental exposure conditions. If the reason why prototypicality gradients were obtained in healthy participants in previous studies was because of an opportunity for explicit encoding, then we may not expect to see any evidence of learning using a visual search task. Further, comparing Group Search to Group Memorize tests whether altering the exposure conditions makes any difference to the prototypicality gradient and ability to discriminate between new and old exemplars. As mentioned previously, the choice of an intentional memorization group
was to maximize any potential group differences that may exist since the available evidence (Gureckis, James, & Nosofsky, 2011) suggests that manipulating encoding strategies in the prototype distortion task does not lead to differences in test performance.

### 3.3.1 Method

#### 3.3.1.1 Participants

Fifty participants (40 female, $M$ age = 19.58, $SD = 5.14$) took part in Experiment 2 in exchange for partial course credit. All participants were first year Psychology students at the University of Sydney. Participants were randomly allocated to group Memorize ($n = 25$) or group Search ($n = 25$).

#### 3.3.1.2 Procedure

The stimuli during exposure and test were generated in the same way as in Experiment 1.

*Exposure Phase.* If participants were allocated to the Search group, the task was framed as a visual cognition task (see Appendix C for exact instructions). Their task was to search for an ‘odd one out’ on each trial, which was defined using line thickness for both the circle and line stimuli. The singleton was created by either adding or subtracting from the line width of a randomly chosen feature in each stimulus, and participants had to respond by saying whether the odd one out was thicker or thinner than the other circles/lines by pressing either A (thicker) or L (thinner) on the keyboard as quickly as possible (see Figure 3.3 for examples of the singleton). Participants were also given a sheet with a visual example using squares where one square was thicker and another was thinner to refer to. It was verbally
emphasized that it did not matter where the odd one out was located within the stimulus, and only the identity of the odd one out mattered for their response. There was no feedback given as to the accuracy of each response to equate the exposure conditions between groups as much as possible. However, participants were given feedback (‘Too slow’) if they failed to respond in the given timeframe (3 seconds from stimulus onset) on each trial, and they were told that they could still respond after the stimulus had disappeared. If participants were allocated to the Memorize group, they were told to memorize the stimuli as a whole for a subsequent memory test, and the task was framed as a visual memorization task (see Appendix C for exact instructions). Note that since participants in the Memorize group would also see a singleton on each trial, the instructions explicitly discouraged participants from attending to specific features. Participants in both groups saw each stimulus for 2 seconds, after which there was a blank ISI of 2 seconds. As in Experiment 1, within each stimulus set there were 4 repetitions of each of the 20 unique exemplars that amounted to 160 trials in total. The identity of the singleton was not consistent across the repeat presentations of each exemplar such that its location could not be predicted. The entire exposure phase was presented in one continuous block with no breaks.

_Familiarity Ratings Test._ The test phase was the same as for the Familiarity condition in Experiment 1 except that participants in the Search group were warned that the “odd one out” would no longer be present and that they should not look for it to aid their judgements. Participants in the Memorize group were not given any instructions about the ‘odd one out’ since participants’ responses on a post-
3.3.2 Results and Discussion

Overall, participants were quite good at the visual search task, performing at 88.8% accuracy (SD = 6.82, min = 75.0%, max = 99.4%). The test data were analyzed in the same way as in Experiment 1, with the presence of a prototypicality gradient inferred by the presence of a significant linear trend in the familiarity ratings, and recognition evident if the familiarity ratings for old exemplars were significantly higher than that for new exemplars. There was a significant main effect and linear trend for distortion, smallest $F(9,432) = 45.81, p < .001, \eta_p^2 = .488$, indicating an overall prototypicality gradient (see Figure 3.4). The main effect and linear trend of distortion both interacted with group, smallest $F(9,432) = 8.42, p < .001, \eta_p^2 = .149$, and novelty, smallest $F(9,432) = 2.05, p = .045, \eta_p^2 = .041$. This was due to the steeper overall generalization gradient in the Memorize group than the Search group, and also the steeper overall gradient for new exemplars than old exemplars (see Figure 3.4).

![Figure 3.4](image.png)

*Figure 3.4. Familiarity ratings in the Memorize and Search groups for new and old test stimuli in Experiment 2.*

---

19 In a pilot experiment, 1 participant out of 14 in the Memorize group reported noticing a singleton in response to the question “Did you notice anything peculiar about the stimuli?”
There was an overall main effect of novelty, with significantly higher ratings for old over new exemplars, $F(1, 48) = 25.76, p < .001, \eta^2_p = .349$, but no significant interaction with group, $F(1, 48) = 3.21, p = .079, \eta^2_p = .063$. Analyses within each group revealed that both groups were able to distinguish between new and old exemplars in their familiarity ratings, smallest $F(1, 24) = 5.95, p = .022, \eta^2_p = .199$. The three-way interaction between group, novelty and distortion level was not significant, $F < 1$, and the Memorize group also rated the prototype as significantly more familiar than the Search group, $F(1, 48) = 12.47, p = .001, \eta^2 = .206$. The Memorize group also produced higher overall familiarity ratings than the Search group, $F(1, 48) = 5.47, p = .024, \eta^2_p = .102$.

It appears that memorizing the stimuli had a more reliable effect on improving the prototypicality gradient (i.e. increased the slope of the generalization gradient) than improving recognition. However, because a visual analogue scale was used, the range that participants used to make their familiarity ratings could have affected the strength of their prototypicality gradient. To illustrate this point, consider a participant from the Memorize group and a participant from the Search group who have equivalent category knowledge. It is possible that despite both participants showing a significant prototypicality gradient, a participant who had been intentionally memorizing the stimuli may rate low distortion exemplars higher on familiarity, and high distortion exemplars lower on familiarity, than their counterpart in the Search group due to higher confidence in their knowledge and therefore greater willingness to use the full range of the scale. The two participants would show the same degree of differentiation between low and high distortion exemplars, but the slope of their generalization gradient would differ. Thus, to test whether the two groups differed on
their sensitivity in distinguishing high and low distortion exemplars, a signal detection analysis was conducted.

3.3.2.1 Signal Detection Analysis

The distribution of responses for each participant was split into sextiles, creating five thresholds. For the prototypicality index, for each threshold, higher ratings for low distortion exemplars (distortion level <= 0.5) counted as a hit, whereas higher ratings for high distortion exemplars (distortion level >= 0.6) counted as a false alarm (old and new exemplars were collapsed). Figure 3.5 shows the receiver operating characteristic (ROC) curve for this analysis, and the resulting sensitivity index ($d_A$) calculated from the linear transformation of the curve (see Simpson & Fitter, 1973). The index of sensitivity in discriminating between high and low distortion exemplars was .859 ($SD = .436$) in the Memorize group, and .419 ($SD = .369$) in the Search group, with significantly higher sensitivity in the Memorize group, $F(1,48) = 14.84, p < .001, \eta^2 = .236$.

A similar analysis was performed (counting higher ratings for old stimuli as hits, and higher ratings for new stimuli as false alarms, levels) on discrimination between new and old exemplars, collapsing over all distortion levels (see Figure 3.5). The Memorize group ($d_A = .232, SD = .239$) was not found to show significantly better recognition than the Search group ($d_A = .096, SD = .269$), $F(1,48) = 3.55, p = .065, \eta^2 = .069$. The results of the signal detection analysis confirm the results reported above, and suggest that memorization enhances the prototypicality gradient but had no or at least much weaker impact on ability to discriminate between new and old exemplars.
Figure 3.5. Results from the signal detection analysis in Experiment 2. Upper panels show ROC curves for discrimination between new and old exemplars (recognition, top left) and discrimination between low and high distortion level exemplars (category learning, top right), with the dotted diagonal line indicating zero sensitivity. Lower panels show dA (sensitivity) measures for recognition (bottom left) and category learning (bottom right) calculated from a linear transformation of the ROC curve, with error bars representing the standard error of the mean.

3.3.2.2 Comparison to Experiment 1

To test whether the prototypicality gradient observed in the Search group was significantly larger than would be expected on the basis of learning at test alone, the Search group was compared to the NoEx-Fam group in Experiment 1. Although this is a between-experiment comparison, the two experiments were run at the same time on the same participant pool. There were significantly higher familiarity ratings overall in the Search group, $F(1,46) = 7.51, p = .009$, $\eta^2_p = .140$, significantly higher
familiarity ratings overall for old exemplars\(^{20}\), \(F(1,46) = 4.38, p = .042, \eta^2_p = .087\), and a significant interaction between novelty and group\(^{21}\), \(F(1,46) = 4.23, p = .045, \eta^2_p = .084\), since only participants in the Search group were able to discriminate between new and old exemplars (see Figure 3.6a). Importantly, while the main effect and linear trend for distortion level was significant, smallest \(F(9,414) = 15.19, p < .001, \eta^2_p = .248\), neither interacted with group, largest \(F < 1\). There was a significant interaction between novelty and the linear trend in distortion, \(F(1,46) = 5.24, p = .027, \eta^2_p = .102\), such that the slope of the generalization gradient was steeper for new exemplars. All other interactions were not significant, largest \(F(9,414) = 1.91, p = .075, \eta^2_p = .040\). There was also no difference in sensitivity in detecting low and high distortion exemplars between the Search group \((dA = .419, SD = 369)\) and the no-exposure group from Experiment 1 \((dA = .528, SD = 1.26)\), \(F < 1\) (see b).

\[\text{Figure 3.6.} \text{ Familiarity ratings for new and old test stimuli in group NoEx-Fam in Experiment 1 and group Search in Experiment 2 (a). dA measures for group NoEx-Fam and group Search (b).}\]

\(^{20}\)This analysis was significant for the circle stimuli only but did not differ according to stimulus set, \(F < 1\).

\(^{21}\)This analysis was significant for the circle stimuli only, but there was no interaction with stimulus set, \(F(1,46) = 2.492, p = .121, \eta^2_p = .051\).
To provide stronger evidence for the equivalence of the prototypicality gradients between the Search group and the No-Ex-Fam group in Experiment 1, a Bayes Factor (BF) test was conducted comparing the slopes and sensitivity index between the two groups. Using the technique suggested by Rouder et al. (2009), a Bayes Factor of 4.1 was found when comparing the slopes, and a Bayes Factor of 4.3 when comparing the sensitivity index ($dA$). Thus, the null hypothesis was over 4 times more likely than the alternative, suggesting that the magnitude of the prototypicality gradients was the same for participants who searched through the stimuli, and participants who produced a false prototypicality gradient after no exposure to the stimuli.

While it is possible that the prototypicality gradient in the Search group in Experiment 2 resulted for different reasons than in the mock-subliminal condition in Experiment 1, the magnitude of the prototypicality gradient (ignoring the overall higher ratings in Experiment 2) is indistinguishable from the prototypicality gradient that emerges in the absence of any exposure, and therefore could be entirely attributed to learning-at-test effects. Nevertheless, it is possible that rather than there being a qualitative difference between exposure conditions, the difference might be quantitative, and the Search group simply learned less about the stimuli than the Memorize group. Therefore Experiment 3 doubled the number of trials in the exposure phase to increase the opportunity for incidental category learning.

Interestingly, participants were still able to discriminate between new and old exemplars even after incidental exposure to the stimuli through visual search. One potential explanation is that the visual search task was too easy and participants were able to encode the stimuli in an explicit way during the exposure phase. In addition to the high levels of accuracy, participants took, on average, 1.47 seconds to respond
(SD = 0.188, min = 1.16, max = 1.85) in the visual search task, meaning that they could have been using the residual exposure time to study the stimulus. Accordingly, Experiment 3 aimed to replicate Experiment 2 with double the number of exposure trials and ensure that any effects obtained in the Search group were actually incidental, by adding an additional group where the stimulus disappeared as soon as a response was made in the visual search task.

3.4 Experiment 3

Experiment 3 was a replication of Experiment 2 where the length of the exposure phase was doubled. Since the prototypicality gradients in Experiment 2 in the Search group were found to be no different to those displayed after no exposure in Experiment 1, increasing exposure to the stimuli may also increase the magnitude of the prototypicality gradient, if it can indeed result from incidental learning conditions. Furthermore, although Experiment 2 found no significant difference between Memorize and Search groups in terms of their ability to distinguish new from old exemplars, small numerical differences in the predicted direction were apparent (e.g. in Figure 3.5, bottom left panel). Thus differential recognition between groups may become clearer with greater exposure to the exemplars.

Experiment 3 also sought stronger evidence that learning in the Search group was actually due to incidental exposure conditions. Because it was necessary to equate the exposure time per trial between the Search and Memorize groups, it is possible that participants in the Search group were not searching through the stimulus for the entirety of the 2 sec exposure duration. In Experiment 3, a second Search group was added (Search-Terminate) where the stimulus exposure terminated after a response was made, ensuring that participants would only be exposed to the stimuli
while searching for and responding to the target. While the total exposure time between the two Search groups would not be equated, a comparison between the two groups would determine whether the residual post-response time makes any difference to test performance.

3.4.1. Method

3.4.1.1 Participants

Fifty-seven University of Sydney first-year psychology students ($M_{age} = 19.61$, $SD = 3.06$, 38 female) participated in this experiment in exchange for partial course credit. Participants were randomly allocated to the Memorize group ($n = 20$), Search group (same as the Search group from Experiment 2, $n = 19$) or Search-Terminate group ($n = 18$).

3.4.1.2 Procedure

The procedure was identical to Experiment 2 except for the following changes. The number of trials in the exposure phase was doubled to 320 trials (8 presentations of each individual exemplar, for both circle and line stimulus sets). Participants in the Search-Terminate group were given identical instructions to the Search group except they were told that the stimulus would disappear once their response was made. If a response was not made within 2 seconds, then the stimulus disappeared regardless. Since the total trial time remained the same between the two Search groups (4 seconds), the blank ISI could range from 2-4 seconds and depended on how long participants took on each trial to make a response. As in the previous experiments, the test phase consisted of 84 trials that included the 20 old exemplars, 20 new exemplars, and the prototype presented twice for each stimulus set.
3.4.2 Results and Discussion

After excluding one participant from group Search for poor performance (45.0% accuracy), the average accuracy in group Search was 92.3% (SD = 5.06, min = 79.4, max = 94.7), which was significantly higher than in group Search-Terminate (M = 83.9%, SD = 8.82, min = 61.2, max = 94.7), t(34) = 3.51, p = .001, SED = .024. The average reaction time (RT) in the visual search task in group Search was 1.42 seconds (SD = .227, min = 0.89, max = 1.75), which was significantly slower than in group Search-Terminate, (M = 1.25 seconds, SD = .153, min = 0.91, max = 1.54), t(34) = 2.55, p = .016, SED = .064. Thus it appears that making the stimulus disappear after participants responded resulted in poorer accuracy but faster RTs.

3.4.2.1 Memorize vs. Search

Since there were three groups in Experiment 3, two separate ANOVAs were run in a similar manner to Experiment 2, one comparing the Memorize group to the two Search groups, and another comparing the two Search groups to each other. For the comparison between group Memorize against the two Search groups, there was a significant main effect of novelty (F(1,54) = 33.39, p < .001, ηp² = .382), which did not interact with group, F < 1. This indicates that there was an overall ability to discriminate between new and old exemplars that did not vary according to exposure condition, similar to Experiment 2 (see Figure 3.7). There was an overall main effect and linear trend for distortion, smallest F(9,486) = 42.50, p < .001, ηp² = .440, but neither interacted with novelty, largest F < 1, suggesting an overall prototypicality gradient that was equivalent for new and old exemplars. Critically, the linear trend for distortion interacted with group, F(1,54) = 5.21, p = .026, ηp² = .088, replicating

---

22 This result was significant for the line stimuli only, but did not differ according to stimulus set, F < 1.
Experiment 2, but the main effect of distortion did not, $F(9,486) = 2.07$, $p = .060$, $\eta^2_p = .037$. The 3-way interaction was not significant, $F < 1$, and the main effect of group was also not significant, $F < 1$. Familiarity ratings for the prototype also did not differ according to group, $F < 1$. Therefore, similarly to Experiment 2, the prototypicality gradient was stronger overall in the Memorize group than the two Search groups, but again there seemed to be no evidence of an advantage in recognition for the Memorize group despite the extension of the exposure phase.

Figure 3.7. Familiarity ratings in group Memorize, Search and Search-Terminate for new and old test stimuli in Experiment 3.

3.4.2.2 Search vs. Search-Terminate

Comparing the two Search groups to each other, there was again a significant main effect of novelty, $F(1,34) = 17.41$, $p < .001$, $\eta^2_p = .339$, that did not interact with
group, $F < 1$. There was a significant main effect and linear trend for distortion, smallest $F(9,306) = 20.03$, $p < .001$, $\eta^2_p = .371$, but neither interacted with novelty, largest $F < 1$, nor group, largest $F < 1$. The 3-way interaction was not significant, $F < 1$, and the main effect of group was also not significant, $F < 1$. Thus, there was no evidence to suggest that the residual exposure time in group Search benefitted overall recognition or the prototypicality gradient compared to group Search-Terminate when the stimulus disappeared after a response was made. Ratings for the prototype alone also did not differ between the two Search groups, $F(1,34) = 3.15$, $p = .085$, $\eta^2 = .085$.

### 3.4.2.3 Signal Detection Analysis

A signal detection analysis was conducted in a similar manner to Experiment 2, comparing the Memorize group against the two Search groups and then the two Search groups to each other (Figure 3.8). For recognition, there was no difference in sensitivity ($dA$) comparing the Memorize group with the two Search groups, $F < 1$, nor when comparing the two Search groups to each other, $F < 1$. The Memorize group did have an advantage in discriminating high and low distortion exemplars when compared to the two Search groups, $F(1,54) = 6.05$, $p = .017$, $\eta^2 = .101$, but there was no difference in sensitivity between the two Search groups, $F < 1$ (see Figure 3.8). Therefore it can concluded that the advantage in the prototypicality gradient for the Memorize group was not due to participants using the scale differently to the other groups.
Figure 3.8. Results from the signal detection analysis in Experiment 3. Upper panels show ROC curves for discrimination between new and old exemplars (recognition, top left) and discrimination between low and high distortion level exemplars (category learning, top right), with the dotted diagonal line indicating zero sensitivity. Lower panels show dA (sensitivity) measures for recognition (bottom left) and category learning (bottom right) calculated from a linear transformation of the ROC curve, with error bars representing the standard error of the mean.

3.4.2.4 Comparison to Experiment 1

To test whether the increased exposure resulted in a prototypicality gradient that was greater than a learning-at-test effect, the two Search groups in this experiment were combined and compared to the NoEx-Fam group in Experiment 1.

As in Experiment 2, there was an overall ability to discriminate between new and old exemplars[^23], $F(1,57) = 8.99, p = .005, \eta_p^2 = .133$, and this interacted with group, $F(1,57) = 8.74, p = .004, \eta_p^2 = .141$, since participants in Experiment 1 did not see any stimuli and thus could not recognize them in the subsequent test phase. The main effect and linear trend for distortion was significant, smallest $F(9,513) = 20.95, p < 1$.  

[^23]: This result was significant for the line stimuli only, but did not differ according to stimulus set, $F < 1$. 

136
< .001, $\eta_p^2 = .269$, but again, importantly, the linear trend did not interact with group, $F(1,57) = 1.70, p = .198, \eta_p^2 = .029$. The signal detection analysis comparing $dA$ between experiments was also not significant, $F < 1$, with similar levels of sensitivity in the two Search groups in this Experiment ($dA = .482, SD = .330$) and the no-exposure group in Experiment 1 ($dA = .528, SD = .126$). While it should be acknowledged that this is a between-experiments comparison, the results are consistent with Experiment 2, and between two methods of analysis. To provide stronger evidence for the null hypothesis, a Bayes Factor test was conducted in a similar manner to Experiment 2 following Rouder et al.’s (2009) technique. This resulted in a Bayes Factor of 2.04 for the comparison of the gradient slopes, and a Bayes Factor of 4.9 for the comparison of the sensitivity index ($dA$) between groups, both in favor of the null. While the BF for the slopes cannot be considered good evidence in favor of the null, the BF for the sensitivity index suggests that, consistent with Experiment 2, participants in the Search groups discriminated between high-distortion and low-distortion exemplars to an equivalent extent as those in the no-exposure group.

The results of Experiment 3 suggest that doubling the number of exposure trials did not exacerbate any potential group differences in recognition ability suggested in Experiment 2, but the advantage for the Memorization group for the prototypicality gradient remained. Increasing exposure still did not result in a prototypicality gradient greater in magnitude than that expected on the basis of learning-at-test alone in the Search groups, implying that the prototypicality gradient is not implicit in the sense of resulting automatically from incidental exposure. On the other hand, the lack of any group differences between the two Search groups suggests that the ability of participants to recognize old exemplars at test does arise from
incidental exposure, and not from using the residual exposure time after responding to explicitly encode the stimuli.

3.5 General Discussion

The current study tested the implicit status of learning in the prototype distortion task by introducing a visual search task as a means of incidental exposure, and tested the effect of manipulating encoding conditions by comparing a visual search group to a group who were simply asked to memorize the stimuli. The methodology was novel in the sense that the same measure (familiarity ratings) and the same test stimuli were used to assess both the prototypicality gradient and ability to discriminate between new and old exemplars. Surprisingly, there was no evidence that participants learned about the similarity structure of the stimuli during visual search, and a dissociation of the opposite nature to those commonly reported in amnesia studies (e.g. Knowlton & Squire, 1993) and healthy participants (Nosofsky & Zaki, 1998) was found, with intentional memorization improving prototypicality gradients but not participants’ ability to discriminate between new and old exemplars.

Experiment 1 showed that the magnitude of the prototypicality gradients and discrimination between old and new exemplars were equivalent for categorization and familiarity tests for participants who were exposed to the stimuli, justifying the use of familiarity ratings for the subsequent experiments to assess the prototypicality gradient. It can be argued that in the decision to use a single measure in Experiments 2 and 3, there was no direct test of participants’ ability to categorize the stimuli. While familiarity judgements and category endorsements are potentially different in terms of the decision rules they engage (Nosofsky, 1988), and the responses they can produce (Squire & Knowlton, 1995), it is clear that the category structure was
reflected in the prototypicality gradients for familiarity ratings, and these prototypicality gradients were sensitive to the group manipulations. It remains to be seen whether these results replicate using a categorization test in place of the familiarity test. Presumably, in this particular task where participants are exposed to a single category, visual similarity would be the primary determinant of both category endorsements and familiarity ratings since the categories are not defined using concepts or rules.

Experiment 1 also addressed an issue brought up by previous studies, that prototypicality gradients can result in the absence of exposure. By comparing generalization gradients between a group of participants who were misled into believing that they had been shown the stimuli in a mock-subliminal procedure (Palmeri & Flanery, 1999), against a group of participants who actually were exposed to the stimuli, ‘false’ prototypicality gradients were found, implying that participants were learning about the category on test. This result, along with the false prototypicality gradients displayed in other studies (e.g. Palmeri & Flanery, 1999; Zaki & Nosofsky, 2004), exemplifies a general problem with assessing A/not A category learning in the prototype distortion task. It is difficult to ensure that effects displayed on test are due to exposure alone, and participants who have not learned anything during the initial exposure phase may feel that they need to provide some variation in their responses and thus seek out information to enable them to do so on test. If this is indeed an unavoidable problem with any stimulus set, a similar procedure to that of Experiment 1 should be employed in future studies before claiming that a prototypicality gradient exists.

Experiments 2 and 3 compared ability to discriminate between new and old exemplars and prototypicality gradients using familiarity ratings between a group who
searched through the category exemplars for a singleton, and a group who attempted to memorize the exemplars for a subsequent familiarity test. Experiments 2 and 3 showed that the Memorize groups produced steeper prototypicality gradients than the Search groups, while there was little evidence for a similar advantage in discriminating new and old exemplars. While the Search groups in Experiments 2 and 3 displayed an ability to discriminate between new and old exemplars that was above chance, the prototypicality gradient (i.e. generalization gradient) displayed after performing the visual search task was found to be no different to the false prototypicality gradient displayed in Experiment 1 after no exposure. Doubling the length of exposure from Experiment 2 to Experiment 3 did not have any effect on the general pattern of results and if anything, weakened the (non-significant) advantage for the Memorize group in recognition in Experiment 2. Further, Experiment 3 showed that there were no differences between a visual search group who were exposed to the stimuli for the set duration (2 seconds), and a visual search group for whom the stimulus disappeared after a response was made. This leads to the conclusion that the ability to recognize old exemplars in the Search group was in fact due to learning that occurred during visual search, and not due to deliberate encoding that may have occurred after a response was made.

3.5.1 Learning during Visual Search

Contrary to claims in the literature (e.g. Smith, 2008; Smith & Grossman, 2008), the current study found no support for the idea that learning in the prototype distortion task is implicit in the sense of resulting from an automatic learning process. If this were the case then the prototypicality gradient should have resulted as a consequence of exposure to the stimuli during the visual search task, despite learning
about the stimuli being incidental to the requirements of the task (searching for, and responding to the identity of the singleton). Since the magnitude of the prototypicality gradient displayed in the Search groups across Experiments 2 and 3 was not found to be greater than the false prototypicality gradient displayed in Experiment 1, there is no strong evidence of learning on the basis of the presence of a prototypicality gradient. While this conclusion rests on a null result, the sample sizes used in these experiments were sufficient to detect significant group differences in prototypicality gradients, and the Memorize groups in Experiments 2 and 3 consistently showed prototypicality gradients that were substantially larger in magnitude than those displayed in Experiment 1. Bayes Factor analyses also support the null hypothesis when comparing the slope of the prototypicality gradient, as well as the sensitivity index between groups.

The failure to find prototypicality gradients after visual search may seem surprising given that there is evidence from other paradigms such as contextual cueing that learning reliably occurs as a consequence of repeated exposures in visual search (e.g. Chun & Jiang, 1998, 2003; Colagiuri & Livesey, 2016). There are, however, several notable differences between the two paradigms. For instance, in contextual cueing, the incidental learning that takes place is still relevant to the task being performed during learning and the repeated presentations contain information about where the target is located (which is one of the explanations for contextual cueing effects). The effect is gauged by an improvement in visual search, which is the task that the participant is practicing throughout training, and performance on that task benefits from consistent task-relevant information about the target itself. In the current visual search task, while features of the category were certainly repeated throughout the exposure phase, and each unique exemplar was repeated a number of
times, the stimulus configurations did not predict the location of the singleton, and the assessment of learning took place after the exposure phase, meaning that it may have been harder to detect incidental learning in general.

Another reason why only a weak prototypicality gradient was observed (i.e. equivalent to no exposure) could be that the Search groups focused their attention on finding a singleton during the exposure phase, which was subsequently removed in the test phase. While the exposure and test stimuli were equated between groups, and the Search groups were explicitly instructed not to look for the singleton on test, it is possible that the Search groups were affected more than the Memorize groups by this small change between the exposure and test stimuli. If there was indeed greater generalization decrement in the Search group, this should have lowered familiarity ratings for the old stimuli more than the new stimuli, weakening the level of discrimination between old and new stimuli. The interaction between group and novelty was not significant in Experiment 2 or Experiment 3, suggesting that familiarity ratings were lower in the Search group by an equivalent magnitude for old and new stimuli. Also, since the Search groups were able to discriminate between old and new exemplars on test, it appears that any generalization decrement suffered by the Search group was minimal.

The failure to find a prototypicality gradient in the Search group is more puzzling since participants were able to give higher familiarity ratings to old stimuli, demonstrating that they had learned something about the stimuli. One way to explain these results is that the prototypicality gradient displayed after the mock-subliminal procedure in Experiment 1 and the prototypicality gradient displayed after visual search were not the consequence of the same learning process despite being of comparable magnitude (see Smith, 2008). In other words, perhaps the latter was a
genuine consequence of learning during training and was not further influenced by learning on test. If the visual search task forces participants to encode the features of the stimuli serially, on test when participants are asked to judge familiarity of the stimuli as a whole, the Search group may resort to searching through the stimuli for specific features that they recognize. This technique might be sufficient to allow a prototypicality gradient to emerge, but one that is not larger in magnitude to that in Experiment 1. Participants who underwent the mock-subliminal procedure in contrast, can utilize the whole stimulus at test to make their categorization or familiarity judgements. However, because they have not been exposed to any stimuli in the exposure phase, what they can learn on test is obviously limited, and thus their prototypicality gradient is similarly small in magnitude to the Search groups. Unfortunately, these experiments provide no means to determine whether this account of the failure to obtain prototypicality gradients after visual search is true. Clearly, further studies are needed to clarify whether incidental learning is conducive to producing prototypicality gradients under different encoding conditions (e.g. serial vs. configural feature encoding).

If we assume that the prototypicality gradients obtained after visual search were due to incidental learning, another way in which the prototypicality gradient can be interpreted as implicit is if it results in the absence of explicit recognition, consistent with the suggestion that prototype effects are not dependent on declarative memory (Knowlton & Squire, 1993; Reber & Squire, 1999). Again, this was not supported by the data, with the Search groups in Experiments 2 and 3 showing higher familiarity ratings for old exemplars, consistent with contextual cueing studies that have found above-chance recognition (e.g. Colagiuri & Livesey, 2016; Smyth & Shanks, 2008).
3.5.2 The Effect of Encoding Conditions

Importantly, Experiments 2 and 3 found that varying the nature of the exposure conditions does affect the ability of participants to show a prototypicality gradient, with participants in the Memorize groups consistently showing an advantage over the Search groups. The results in this chapter stand in contrast to those of Gureckis et al. (2011), who found no differences in test performance between groups of participants who were asked to memorize the stimuli as a configural whole, and participants who were asked to imagine pointing to the center dot. As mentioned previously, one explanation for their results is that their incidental task was not cognitively demanding and thus participants may have been able to explicitly encode the stimuli while performing the task. An alternative explanation is that the task itself (pointing to the middle dot) might result in a similar deployment of visual attention as the memorization task. To speculate, pointing to the middle dot might also involve configural processing of the stimuli as participants would have to encode spatial relations between the features (in effect processing the stimulus as a whole) in order to determine where the center of the pattern was located, thus making the two encoding strategies similar in terms of what participants attend to. This would mean that the reason that strong prototypicality gradients have been obtained in the past under incidental exposure conditions is largely due to the nature of the encoding conditions facilitating later performance, rather than incidental learning.

In contrast, in this chapter, a more cognitively engaging visual search task was utilized, demonstrating that test performance can vary between different exposure conditions. However, visual search tasks require participants to search through individual features of the stimuli in order to find the target, and thus differs both in terms of the requirements of the task and what features (configural vs. specific) of the
stimuli participants focus on. As mentioned above, it is possible that the Search group did learn incidentally but only about the specific features of the stimuli. The Memorize groups on the other hand, were told to memorize the stimuli as a whole and thus were encouraged to encode the configural as well as specific features of the stimuli. A prediction concerning the impact of these encoding strategies on test can be derived by considering how these changes in encoding may affect generalization to previously unseen high- and low-distortion exemplars.

Generalization increases as the similarity between the test stimulus and the seen exemplars increases. The low distortion exemplars are very similar to the prototype, and therefore also very similar to each other, whereas the high distortion exemplars are relatively dissimilar to each other. Thus, the prototypicality gradient in the old test stimuli can be seen to result in the following way: generalization is high (and therefore high familiarity ratings are given) for low-distortion exemplars because of a high degree of similarity to other seen low-distortion exemplars. In contrast, generalization is reduced for high distortion exemplars since these exemplars are not as similar to the seen exemplars (whether they be high- or low-distortion). If we assume that the Memorize group are better able to detect similarity (for example, due to encoding of configural features of the stimuli), then generalization will be higher in this group for the low distortion exemplars and thus produce a steeper prototypicality gradient. In contrast, the Search group may have more difficulty detecting similarity between stimuli (for instance, because they have only encoded the specific features of the stimuli in a serial fashion, and the exemplars are created by distorting each feature individually). As mentioned above, if the Search group were to adopt the strategy at test of searching for individual features that they recognize, this might allow them to

---

24 This explanation can also be framed in terms of generalization from an extracted prototype.
discriminate between old and new exemplars to the same degree as the Memorize group, but would not enable them to produce as strong a prototypicality gradient. Thus, a speculative conclusion is that the visual search task may have resulted in incidental learning, but through a manner of encoding that was very different to the Memorize group.

An alternative way to explain the group differences in Experiments 2 and 3 is through awareness of the subsequent familiarity test, since this was necessarily part of the encoding instructions in the Memorize group. While it is certainly possible that any group differences observed could be attributed to this awareness and not the difference in encoding conditions, the results of Gureckis et al. (2011) and Reber et al. (2003) do not support this idea. Both studies failed to find a difference in categorization performance when comparing groups that were aware of the existence of a category against a group who were unaware of the category. This suggests that awareness of a category either does not lead participants to look for similarities between stimuli to aid them in their category judgements, or that the nature of the stimuli makes this difficult. While this study utilizes a familiarity test rather than a categorization test, participants in Group Memorize knew that they would be tested on their memory and yet in both Experiments 2 and 3 there was little evidence that this facilitated their performance in discriminating between new and old exemplars. Thus if anything, we would need to conclude counter-intuitively that awareness of a subsequent memory test facilitates prototypicality gradients, but not recognition ability. It seems that what matters for learning in the prototype distortion task is not whether participants are aware of the existence of a category or an impending memory test, but what demands the particular encoding conditions make on the participant.
3.5.3 The Dissociation between Categorization and Recognition

Finally, the manipulation of encoding conditions in Experiments 2 and 3 increased the magnitude of the prototypicality gradient, but did not improve discrimination between new and old exemplars. This is the opposite of what has been found when comparing amnesic patients to healthy controls where there is usually a discrepancy in recognition but similar categorization performance (e.g. Knowlton & Squire, 1993). If we assume, as single-system theories do, that the two indices of learning should be related, the results in this chapter might be surprising given that the use of a single measure should have eliminated potential differences in response thresholds, and other task-specific variance such as forgetting and interference between tests. Nevertheless, the fact that dissociation is still observed should be interpreted with caution. While the test measures have been equated, there is still a difference in the statistical measures used for the prototypicality gradient (linear trend in ratings) and recognition (difference in ratings between old and new stimuli). Thus, there is still a parameter difference between these chosen indices. For example, it may be that they are not equally easy effects to obtain, or that discriminating between high and low distortion stimuli and new and old stimuli are differentially impacted by encoding conditions.

It is also worth emphasizing that if participants are learning about the category on test, as Experiment 1 suggests, then discriminating new from old stimuli may become more difficult, since participants would be explicitly encoding both old and new exemplars on test. This would presumably strengthen, or make no difference to the prototypicality gradient but may contribute to a weaker level of recognition. This potential problem also applies to most of the prototype effect literature because test stimuli are typically a combination of old and new stimuli and usually many (if not
all) are structured around the prototype to some extent. Thus, for the same reasons, their presentation on test should strengthen the prototypicality gradient but weaken participants’ ability to recognize exemplars they have seen. Ensuring that difficulty level is equated across assessment of the prototypicality gradient and recognition is thus important for interpretation of dissociations but also difficult to implement.

While the results show a clear advantage for intentional memorization in the slope of the prototypicality gradient, the use of a single-category paradigm means that there was no assessment of the ability of participants to accurately sort stimuli into categories, which is what categorization is typically thought to entail (Homa, Hout, Milliken, & Milliken, 2011). However, a descending gradient in familiarity ratings or category endorsement can be seen to represent knowledge of category structure, which in prototype distortion tasks is defined by stimulus similarity. Single-category paradigms, or A/not A paradigms, are also the paradigms that have traditionally been used in amnesia research (e.g. Knowlton & Squire, 1993), have been singled out as implicit (Smith, 2008), and therefore it is the status of learning in these types of procedures in particular to which these results are most relevant. Nevertheless, further studies may provide insight into whether different encoding conditions affect prototypicality gradients and categorization into multiple categories in the same way.

To summarize, the experiments in this chapter call into question the logic of interpreting single dissociations as evidence for implicit category learning due to the inherent difficulty in ensuring a ‘fair’ comparison between tests of categorization and recognition. Since a pattern of results that is the reverse of the majority of studies comparing amnesics to controls was found (e.g. Knowlton & Squire, 1993), it seems that the way in which the prototype effect and recognition are assessed has a large influence on the pattern of results obtained, and even when attempts are made to
minimize differences between tasks by using the same test stimuli and test measure, dissociations are still possible.

3.5.4 Conclusion

In conclusion, the experiments in this chapter do not support the characterization of learning in the prototype distortion task as implicit in the sense of resulting automatically from incidental exposure during visual search, and found a dissociation opposite to that commonly reported in the literature when the same test measure was used to assess prototypicality gradients and recognition. These findings highlight the need to test for potential learning-at-test effects before claiming that learning exists, and to exercise caution in interpreting dissociations between categorization and recognition due to the difficulty in eliminating potential parameter differences that can cause dissociations. Most importantly, Experiments 2 and 3 found a difference between explicit memorization and incidental learning during visual search, emphasizing the importance of studying the encoding strategies and exposure conditions that are required for so-called implicit learning effects.
Chapter 4: Post-Discrimination Generalization

4.1 Introduction

Generalization concerns the transfer of learning from one stimulus to another, and is an important aspect of behavior since each encounter with a stimulus is arguably different in some form (Shepard, 1987). Generalization was explored in Chapter 3 where participants were exposed to exemplars from a single prototype-centered category. Unlike Chapter 3, in this chapter, participants learned to discriminate between slight variants of two category prototypes rather than being exposed to a single category, and generalization will be tested to stimuli further along the relevant dimension. This type of category learning task is more similar to intradimensional discrimination procedures where animals are rewarded for responding to one stimulus and not rewarded for responding to another stimulus lying on the same dimension (see Honig & Urcioli, 1981, for other ways in which generalization has been explored).

A wide range of studies using these simple discrimination tasks suggest that humans are capable of generalizing on the basis of similarity to the physical features of the stimuli (as seen in Chapter 2) but, importantly, also on the basis of the abstract relationships between the to-be-discriminated stimuli (e.g. Mackintosh, 2000; Penn, Holyoak, & Povenelli, 2008). In contrast, there is still debate about the presence of even simple forms of relational generalization in infrahuman animals (Penn, Holyoak, & Povenelli, 2008). Relational and feature-based generalization appear to capture learning that is qualitatively different in content, but whether they represent the operation of separate learning processes is more difficult to determine empirically. If however, there are two learning processes that govern generalization according to
similarity or relational rules, specifying the conditions under which participants generalize will have implications for characterizing the nature of interaction between those processes.

In animal learning experiments, generalization is often examined by rewarding subjects for responding to a single stimulus (S+), for example, a light of a particular wavelength or a tone of a particular frequency, and then assessing responding to stimuli with varying values along that same dimension of interest (e.g. Guttman & Kalish, 1956; Pavlov, 1927). Such studies in animals typically produce a peaked generalization gradient, with the highest rates of responding at the trained S+, and responding decreasing as a function of the degree of similarity between the test stimulus and the S+ (see Ghirlanda & Enquist, 2003, for a review). These peaked generalization gradients can be explained by assuming that behavior towards a stimulus is governed by the degree of similarity or shared features between the novel instance and the past instance that has already been encountered. This assumption is a fundamental tenet of many associative learning theories (Blough, 1975; Estes, 1955; McLaren & Mackintosh, 2000, 2002; Pearce, 1987; Shepard, 1987; Spence, 1937).

If instead of training responses to a single stimulus, animals are trained using differential reinforcement to discriminate between an S+ and S- that lie close together on the relevant dimension, a phenomenon called peak shift often occurs (Hanson, 1959). Peak shift describes the situation in which the peak of the generalization gradient shifts from the location of the S+ in the direction away from the S- (see Figure 4.1 for an example). This effect is well accounted for by associative models that employ elemental representation, and conceive of the S+ and S- as a series of overlapping elements on a continuum (Blough, 1975; Ghirlanda & Enquist, 1998). Peak shift can be predicted if it is assumed that S+ and S- activate elements on the
continuum in a graded (Gaussian) manner during discrimination training. If the S+ and S- contain a large number of common elements, a large proportion of elements on the dimension will be activated by both stimuli and accrue weak associative strength due to their excitatory association with the S+ and inhibitory association with the S-. Elements that are maximally activated by S+ and minimally activated by S- will accrue the strongest associative strength. Thus elemental theories predict that a stimulus slightly removed from the S+ in the direction away from the S- should accrue the most associative strength.

![Graph showing generalization results based on rules and similarity.](image)

**Figure 4.1.** Simulation of generalization results based on rules and similarity. Generalization on the basis of rules produces a monotonically increasing gradient while generalization on the basis of similarity produces a peaked gradient. The ‘Near’ stimuli represent stimuli along the dimension that are very similar to the training stimuli (‘Train’), while the ‘Far’ stimuli represent stimuli that are very dissimilar to the training stimuli.

A similar analysis can be applied for a discrimination task with two stimuli (i.e. S1 and S2) with competing responses (i.e. R1 and R2). These responses would presumably have an inhibitory association with each other, meaning that S1 serves as a S- to S2, and vice versa (see Blough, 1973). Assuming that the probability of
choosing the associated response is a function of the activation of that response relative to activation of the alternative response, the highest rates of R1 will be to a stimulus slightly removed from S1 in the direction away from S2, and vice versa for S2.

In animal learning, even when a shift in the peak of responding occurs, a decline in responding is usually observed for extreme stimuli that are sufficiently different from those used in training (e.g. Hanson, 1959, see Figure 4.1). This suggests that generalization still occurs on the basis of physical similarity, albeit with a learned bias towards exaggerated versions of the S+. Thus, a generalization gradient with a localized peak of responding is generally interpreted as evidence of generalization based on the number of shared features between the test and training stimulus, consistent with an associative analysis.

While peak shift is readily found in animal studies given appropriate parameters (see Purtle, 1973), its elusiveness in human studies suggests there may be other important factors that dictate the manner of generalization. Analogous procedures in humans, particularly those involving simple forms of category learning, often result in a different pattern of generalization – a monotonically increasing gradient with the highest levels of responding or accuracy at the extreme ends of the dimension, where the stimuli are the most dissimilar from the training stimuli (e.g. LaBerge, 1961; Livesey & McLaren, 2009; Thomas, Lusky, & Morrison, 1992; Wills & Mackintosh, 1998, see Figure 4.1). A monotonic gradient implies that participants have learnt about the relationships between the features of the stimuli rather than just their physical features. If participants have formed a simple relational rule (e.g. concerning brightness: “category 1 is darker than category 2”), accuracy should be greatest at the extreme ends of the dimension (e.g. the lightest and darkest stimuli) since a relational
rule is presumably easiest to apply at these test stimuli. Forming a relational rule requires abstract representations of the difference between the stimuli and thus the manner in which they are related to one another. Higher-order cognitive processes such as hypothesis-testing and symbolic representation are thought to be critical in the successful use of such rules (Mitchell, De Houwer, & Lovibond, 2009; Penn, Holyoak, & Povenelli, 2008). Therefore, both the content of learning and the underlying mechanisms in rule-based generalization may be qualitatively different to the processes that support feature-based generalization in other animals.

There are, however, a few instances of peak shift reported in the human generalization literature that are suggestive of a common mechanism of generalization between humans and animals (Mackintosh, 1997). Most of these demonstrations of peak shift arise from situations in which formation of a relational rule is difficult. For example, peak shift can be found when relational rule learning of the form just described is difficult due to the complexity of the stimuli (Wills & Mackintosh, 1998). Participants in Wills and Mackintosh (1998; Experiment 2), categorized complex stimuli lying on an artificial dimension that was created by varying the number of icons in each stimulus (see Table 4.1). On test, participants performed most accurately on ‘Near’ stimuli, which contained a larger number of prototypical icons than the training stimuli, but performance declined for the ‘Far’ stimuli, which contained fewer prototypical icons than the training stimuli.
Table 4.1.
Average number of icons (letters) present in each test stimulus along the artificial dimension (A-L, adapted from Wills & Mackintosh, 1998). S+ is the reinforced stimulus and S- is the non-reinforced stimulus presented during training. Near stimuli are similar to the training stimuli, distant stimuli are dissimilar to the training stimuli and represent the extreme ends of the artificial dimension.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distant+</td>
<td>0.7</td>
<td>2.9</td>
<td>4.8</td>
<td>2.9</td>
<td>0.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Near+</td>
<td>0.7</td>
<td>2.9</td>
<td>4.8</td>
<td>2.9</td>
<td>0.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S+</td>
<td>0.7</td>
<td>2.9</td>
<td>4.8</td>
<td>2.9</td>
<td>0.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-</td>
<td>0.7</td>
<td>2.9</td>
<td>4.8</td>
<td>2.9</td>
<td>0.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Near-</td>
<td>0.7</td>
<td>2.9</td>
<td>4.7</td>
<td>2.9</td>
<td>0.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distant-</td>
<td>0.7</td>
<td>2.9</td>
<td>4.8</td>
<td>2.9</td>
<td>0.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The letters A-L represent different visual icons.

Another way to obtain peak shift is by having participants make speeded responses to a target with the stimuli presented as incidental cues (Aitken, 1996). Aitken found that participants were able to learn the predictive relationships between the cues and the target location by showing faster reaction times to the targets in the presence of the predictive cues compared with unpredictable filler cues. Interestingly, they appeared to learn these relationships incidentally and were not able to explicitly verbalise the relationship between the cues and the targets. Most importantly, when tested with stimuli along the dimension, participants showed a peak shift, making fewer correct responses to stimuli at the extreme ends of the dimension. While this task is quite different from that used in Wills and Mackintosh (1998), the two paradigms are similar in that they both restrict the formation of relational rules, which perhaps allows responding to novel stimuli to be governed by the degree of similarity to the training stimuli. Other demonstrations of peak shift in category learning rely on reducing the amount of training (Jones & McLaren, 1999), degrading the contingency between the training stimuli and the correct response (Jones & McLaren, 1999), or interleaving two qualitatively distinct sets of stimuli to decrease the opportunity for stimulus comparison between trials (Livesey & McLaren, 2009). Taken together,
these demonstrations suggest that when the training stimuli and procedures minimize the opportunity to form a relational rule, humans generalize on the basis of physical stimulus features, and in a manner that produces the peak shift phenomenon.

Since the literature demonstrates that humans can generalize in different ways, the question remains as to what determines the basis (relational rules or similarity) of generalization after simple discrimination learning. An obvious answer, and one that is consistent with the studies discussed above, is that rule-based generalization depends on the conditions of learning being conducive to participants forming a relational rule. Perhaps the clearest demonstration of this comes from a study by Livesey and McLaren (2009). In their study, they trained participants to discriminate between two shades of green differing only in their hue. In two experiments, a peak-shifted gradient was found in the initial phase of testing which gradually became monotonic throughout testing. In Experiment 2, this change in generalization was only found for participants who had failed to notice the difference between the stimuli during training as assessed by a post-experiment interview (those who reported noticing the relation in training produced monotonic gradients from the outset). Presumably, these participants were able to use what they had learned during training to derive the appropriate rule on test. Their within-subjects demonstration of both a monotonic and peak-shifted generalization gradient shows that in the absence of a relational rule, participants generalize according to the physical features of the stimuli. Assuming that the peak shift was the result of associative learning, Livesey and McLaren went on to speculate that while their results demonstrate that associative learning can be overridden at test by higher-order rule-learning processes, they are also consistent with the idea that in humans, associative learning provides input into higher-order cognitive processes such that the two processes are effectively integrated.
into category judgements (see Natal, McLaren, & Livesey, 2013, for a similar argument).

However, the experimental evidence suggests that the integration of associative and relational information is probably not equal. The fact that peak shift is so elusive suggests that humans predominantly use explicit rules when confronted with discrimination tasks, and if present, this relational rule dominates responding at test. This dominance of rule-based responding may be due to the fact that the majority of human discrimination studies reporting rule-based generalization have used stimuli that are relatively simple with only one relevant dimension (e.g. colored squares, Livesey & McLaren, 2009). Although the discriminations are usually difficult enough to ensure that deriving a rule is not easy, once it has been derived the application of that rule on test is certainly straightforward, given that participants can identify the rule-relevant difference between the stimuli or categories of stimuli. While there have been studies on peak shift and rule-governed responding that have attempted to disrupt rule formation (for instance, by reducing the contingency between the stimuli and correct response during training, Jones & McLaren, 1999), most have focused on preventing the formation of the rule in the first place and not on disrupting its subsequent use on test. To date, none of the post-discrimination generalization studies of this nature have observed disruption of the application of rules on test once that rule has been learned. It may be that associative and rule-based processes interact in a more complex way once both types of learning are acquired, such that other unexplored factors determine the degree to which rules govern generalization.

To summarize, previous findings suggest that knowledge of a relational rule dictates the manner in which participants generalize (e.g. Livesey & McLaren, 2009). However, more evidence is needed to determine whether the basis of generalization
can change qualitatively between physical similarity and relational rules and, if so, how similarity- and rule-based generalization interact to determine responding when such a shift occurs. The current set of experiments approached these questions in a different way to previous studies. Rather than restricting the acquisition of relational rules during training, as has been done in the past, conditions were provided to encourage the use of a relational rule and then its applicability on test (that is, the ease with which the rule can be applied) was manipulated. The hypothesis was that evidence of feature- or similarity-based generalization (i.e. peak shift) would emerge when the applicability of a rule was reduced at test. The results of Livesey and McLaren (2009) show that participants’ generalization gradients progress from being peak-shifted towards being monotonic through the course of testing only if participants derive a rule at test and not training, suggesting that in the absence of a relational rule, participants learn about the physical features of the stimuli. Thus, if knowledge of relational rules and stimulus features are stored concurrently and expression of rule-based generalization dominates under situations where the rule is easy to apply, then reducing the applicability of a rule on test may result in participants reverting to generalizing on the basis of similarity to stimulus features.

The aim of the experiments in this chapter was to see whether reducing the applicability of a relational rule on test could disrupt the application of that rule on trials where the rule should be easy to apply. To reduce the applicability of the rule on test, a situation was provided where a rule was easily applicable, or useful in determining category membership on 50% of test trials but not easily applicable (and potentially not useful at all) in determining category membership on the other 50% of test trials. In order to prevent participants from simply discounting the trials where the rule was difficult to use as completely irrelevant, the test stimuli were created to be
perceptually similar to the training stimuli so that participants would assume that their rule would continue to be valid throughout test. A set of stimuli was constructed that had two relevant category dimensions and therefore two possible relational rules that participants could form. This allowed for disruption of rule use by making one dimension (and rule) relevant on half of the test trials, and the other dimension (and rule) relevant on the other half of test trials. The stimuli were complex with multiple features (9 colored circles on a black background, see Figure 4.2), making rule application on test more difficult than previous studies, and the hue (blue vs. green) and size of the circles (small vs. large) chosen as the diagnostic category dimensions.

![Figure 4.2](image)

*Figure 4.2. Examples of stimuli used in the training phase in Experiments 1 and 2. In this example, the ‘Left’ category has smaller and greener circles while the ‘Right’ category has larger and bluer circles.*

These dimensions were correlated during training, meaning that for example, one category had bluer and larger circles and the other had greener and smaller circles (see ‘Train’ stimuli in the Consistent Group in Figure 4.3). This meant that participants could successfully discriminate between the categories using either category dimension. In order to provide the best opportunity to obtain peak shift, the
stimuli were made to be perceptually similar and included a large amount of noise such that relational differences for any given circle were less reliable than the average of all the circles. This added noise makes the discrimination more difficult but also discourages participants from attending to a single circle, which might have made rule application straightforward and undermined the effect of the test manipulation. Note that making the stimuli noisy and complex might more closely approximate how animals experience stimuli that are perceived as simple to humans (e.g. keylights), and thus be more amenable to producing a peak shift effect.
Figure 4.3. Schematic diagram of the test stimuli in a two-dimensional space for the Consistent and Inconsistent groups for Experiment 1 and 2A. Variations on the attended dimension are on the horizontal axis and variations on the unattended dimension are on the vertical axis. “T” represents the ‘Train’ test stimulus and ‘F’ represents the Far test stimulus. Note that for group Consistent, the ‘Train’ test stimuli for each dimension were exactly the same, hence they are only displayed once. ‘Near’ stimuli not shown for clarity. The white dotted line represents the category boundary. Diagnostic information was always present on both dimensions in the Consistent group, but only present on one dimension for any given test stimulus in the Inconsistent group. Experiment 2A differed from Experiment 1 in that one of the dimensions was attended and the other unattended whereas in Experiment 1 the more attended and less attended dimensions were determined post-hoc. In the Consistent group, the Train test stimuli (T) were also the category prototypes.
To create the test stimuli, the dimensions were varied one at a time such that when one dimension was being varied, the other stayed constant. Whether the information on this non-varied dimension was diagnostic of category membership (i.e. varied reliably between categories) was manipulated between groups. When a dimension (e.g. color) was varied in Group Consistent, the other dimension (e.g. size) was set at the same values used for the training stimuli such that size was still diagnostic of category membership (Figure 4.3), but in group Inconsistent, size was set at a value roughly in the middle of the two categories, making it non-diagnostic of category membership (Figure 4.3). This effectively meant that although the test trials for the two groups looked very similar, on any given test trial, either rule could be used in Group Consistent, but one of the two rules was much more difficult to use in Group Inconsistent. If participants derived a rule using one of the two dimensions, this rule would either be easy to apply on 100% (Group Consistent) or 50% of test trials (Group Inconsistent). Thus, Group Inconsistent would experience a sequence of test trials where their rule was easy to apply on some trials, and very difficult to apply on the other trials, making application of a rule inconsistent. The question of interest was whether manipulating the consistency of rule application on test in this way would affect application of the rule on the test trials where the rule was clearly valid and should thus be easy to apply.

Another aim of the experimenters in this chapter was to provide stronger evidence for rule use by including an additional measure of stimulus similarity. Livesey and McLaren (2009) found a strong concordance between monotonicity in generalization gradients and verbal description of the relevant stimulus relation. Evidence of rule use at test corresponded with participants reporting that they had noticed the relevant dimension during training. However, this type of analysis
inevitably requires post-hoc comparisons from which the relationship between patterns of generalization and articulated knowledge is never completely clear. Furthermore, associative models (e.g. Ghirlanda & Enquist, 1998) can also explain monotonic gradients of generalization through similarity-based processes, by assuming very broad underlying generalization functions. Under these parameters, the hypothetical peak of the gradient lies close to or beyond the bounds of the test range and thus declining response accuracy may not be observed (Livesey & McLaren, 2009). This explanation implies that a monotonic gradient results because participants are over-generalizing to stimuli that they perceive to be physically similar. This stands in contrast to rule use, in which a major advantage is that it allows for the extrapolation of learning to novel situations that have similar structural or relational features but are potentially dissimilar in their surface features (Penn, Holyoak, & Povinelli, 2008; Smith, Langston, & Nisbett, 1992).

A simple way to demonstrate that monotonic generalization is based on stimulus properties other than mere physical similarity is to have participants simultaneously acknowledge that the stimuli at the extreme ends of the dimension (which should show the highest levels of accuracy) are the least similar to the stimuli they have seen during training. Therefore, in addition to category judgements on test, participants were also asked to give typicality ratings as a measure of perceptual similarity. If indeed the presence of a monotonic gradient were due to over-generalization and under-assessment of the full range of the dimension, then one would expect categorization accuracy and typicality ratings to closely mirror each other. On the other hand, if participants are using a relational rule, then categorization accuracy and ratings of typicality should diverge as the test stimuli become less similar to the trained stimuli.
While an effect of the test manipulation relies on participants learning at least one of the two relational rules, Experiment 1 presented training in the absence of any instructions about the relevant dimensions to explore whether participants would self-generate a relational rule concerning one of the dimensions during the course of discrimination learning. To foreshadow the results, Experiment 1 yielded evidence of rule use but little effect of the test manipulation. In order to increase reliance on a rule at training, and therefore increase the probability of observing an effect of reducing rule applicability at test, in Experiments 2A and 2B, participants were explicitly directed to attend to one of the two dimensions, creating an attended and an unattended dimension (either the color or size of the circles served as the attended dimension, counterbalanced across participants). The focus was on how participants generalized on the attended dimension since this was the dimension where rule use was most likely to occur, and therefore the dimension on which rule disruption was likely to occur. However, generalization along the unattended dimension was also assessed to see whether any learning occurred in the absence of directed attention.

Experiments 1 and 2A contained test trials where diagnostic information was always present on both dimensions in the Consistent group, while information from only one dimension was present for any given test stimulus in the Inconsistent group (see Figure 4.3). Experiment 2B equated the critical test trials where the attended dimension was varied across groups (see Figure 4.4 and Table 4.2) while maintaining rule applicability at 100% and 50% across the two groups (the same as Experiments 1 and 2A) in order to rule out the possibility of rule-based generalization in the Consistent group being due to additional information present on the unattended dimension (for an illustration of this, compare variations on dimension 1 in the Consistent group in Figure 4.3 and on the attended dimension in the Consistent group
in Figure 4.4). Typicality ratings were also included as a measure of stimulus similarity in addition to category judgements in all experiments to provide further evidence for relational rule use in addition to the post-experimental questionnaire.

Table 4.2. Distortion values from the base prototype (represented as 0,0) for any given stimulus circle in the test stimuli.

<table>
<thead>
<tr>
<th>Group</th>
<th>LEFT CATEGORY</th>
<th>RIGHT CATEGORY</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Far</td>
<td>Near3</td>
</tr>
<tr>
<td>Exp 1&amp;2A: Group</td>
<td>Att</td>
<td>-5</td>
</tr>
<tr>
<td>Consistent</td>
<td>Unatt</td>
<td>-1</td>
</tr>
<tr>
<td></td>
<td>Att*</td>
<td>-5</td>
</tr>
<tr>
<td></td>
<td>Unatt</td>
<td>-1</td>
</tr>
<tr>
<td>Exp 1&amp;2A: Group</td>
<td>Att</td>
<td>-5</td>
</tr>
<tr>
<td>Inconsistent</td>
<td>Unatt</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Att*</td>
<td>0</td>
</tr>
<tr>
<td>Exp 2B: Group</td>
<td>Att</td>
<td>-5</td>
</tr>
<tr>
<td>Consistent</td>
<td>Unatt</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Att*</td>
<td>0</td>
</tr>
<tr>
<td>Exp 2B: Group</td>
<td>Att</td>
<td>-5</td>
</tr>
<tr>
<td>Inconsistent</td>
<td>Unatt</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Att*</td>
<td>0</td>
</tr>
</tbody>
</table>

Note. Whole numbers shown for clarity. The actual distance between the two category prototypes was larger than the distance between test exemplars. The non-varied dimension is marked with an asterisk. The attended dimension was chosen post-hoc in Experiment 1 but explicitly manipulated through instructions in Experiments 2A and 2B and could be either color or size. Negative numbers indicate that the stimulus belongs to the left category and positive numbers indicate that the stimulus belongs to the right category. The key difference between Experiment 2A and 2B was that the critical test stimuli varying the attended dimension were equated between Consistent and Inconsistent groups in Experiment 2B (same values on attended and unattended dimensions).
Figure 4.4. Schematic diagram of the test stimuli in a two-dimensional space for the Consistent and Inconsistent groups for Experiment 2B. Variations on the attended dimension are on the horizontal axis and variations on the unattended dimension are on the vertical axis. “T” represents the ‘Train’ test stimulus and ‘F’ represents the Far test stimulus. Note that the training stimuli for group Consistent always contained diagnostic information on both dimension, but the training stimuli for group Inconsistent only contained diagnostic information on a single dimension. ‘Near’ stimuli not shown for clarity. The white dotted line represents the category boundary. Diagnostic information was always present on the attended dimension (horizontal axis) in the Consistent group, but was only present on half of the test trials in the Inconsistent group. Compare variations on the attended dimension in the Consistent group to variations on Dimension 1 in the Consistent group in Figure 4.3.
4.2 Experiment 1

The aim of Experiment 1 was to test whether reducing the validity of a rule on test would disrupt its application on trials where it should be easy to apply. The experiment used a between-subjects design in which the key group difference was a manipulation of consistency in rule application at test, with participants randomly allocated to a Consistent or Inconsistent group. For the Consistent group, participants were always provided with diagnostic information from both the color and size dimensions on test, which meant that provided the participant formed at least one rule based on either circle color or size, they could continue to use that rule consistently on test. Figure 4.3 shows how the groups differed in terms of information on the non-varied dimension for the average of the 9 circles within a test stimulus.

For Group Consistent in Experiment 1, for any given test stimulus, when color was varied to create the Near (intermediate) and Far stimuli, the size values were set at their respective category values seen during training, allowing participants to accurately categorize the stimuli using either dimension (see Figure 4.3). In contrast, in group Inconsistent, when color was varied, the size values were set at a value approximately in the middle of the two categories, meaning that any rule formed on circle size would be very difficult to use, since the value for each of the 9 circles would be at, or very close to the category boundary (see Figure 4.3). Therefore, for stimuli where color was varied, only color was informative for categorization, and vice versa for size. Figure 4.5 shows an example of test stimuli seen by the Inconsistent group. Notice that when size was varied (top row), the color values do not change between categories and when color was varied (bottom row), the size values of the circles do not change (but the location of the circles change randomly, providing some noise in the stimuli). This means that if participants were to derive a
rule using color, for example, this rule would be difficult to use on test trials when size is varied (Figure 4.5, top row), and is thus only useful on the half of test trials where color is being varied (Figure 4.5, bottom row).

Figure 4.5. Example of test stimuli seen by the Inconsistent group varying size (top row) and color (bottom row). Note how the values of the non-varied dimension (color in the top row and size in the bottom row) do not change between categories. The values of the non-varied dimension were set at a value roughly in the middle of the two categories, making it non-diagnostic of category membership in the Inconsistent group. The Consistent group had diagnostic information from the non-varied dimension, such that there would be a small difference between categories in terms of color in the top row (the same difference present between the Train stimuli in the bottom row), and a small difference between categories in terms of size in the bottom row (the same difference present between the Train stimuli in the top row).

Experiment 1 was exploratory since it was unknown whether participants would easily form relational rules on both dimensions, on one of the dimensions, or neither dimension. Verbal reports and questionnaires have been extremely useful in previous studies in distinguishing between rule users and non-rule users and explaining differences in generalization that might otherwise be obscured by looking at data averaged over test phases or participants (e.g. Livesey & McLaren, 2009; Natal, McLaren, & Livesey, 2013). Therefore, a detailed questionnaire was included at the
end of the experiment to assess which dimensions participants found useful and
whether they could readily identify differences between the categories in terms of
color and size. It was hypothesized that, amongst those participants who were able to
form at least one relational rule, participants who could apply their rule consistently
throughout test (group Consistent) would show a monotonic gradient of
generalization when their rule-relevant dimension was varied, consistent with
previous studies demonstrating rule-based generalization in humans (e.g. Livesey &
McLaren, 2009; Wills & Mackintosh, 1998). For participants in group Inconsistent,
for whom the rule is no longer fully applicable on test, it was hypothesized that rule
use would be disrupted such that overall test accuracy would be lower and
participants would revert to generalizing on the basis of similarity to physical features.

4.2.1 Method

4.2.1.1 Participants

One hundred and thirty-three University of Sydney students ($M$ age = 20.6, $SD$
= 4.52, 105 females) participated in this experiment in exchange for partial course
credit or payment (AUD$15/hour)$^{25}$. Participants were randomly allocated to the
Consistent ($n = 68$) or Inconsistent group ($n = 65$). Participants who indicated that
they were colorblind were excluded from the analyses (6 participants).

4.2.1.2 Apparatus

The experiment was programmed using Matlab software and PsychToolbox
(Brainard, 1997; Pelli, 1997). Each setup consisted of an Apple Mac Mini desktop
computer connected to a 27 inch Dell monitor, with a standard Apple keyboard and

---

$^{25}$ Target sample sizes for this, and subsequent experiments were 70 per group based on exclusion rates
from a pilot experiment. The final sample size was affected by availability of participants.
mouse. Testing was conducted in groups of up to 4 and participants were asked to wear headphones to block out sound. The apparatus was the same for all subsequent experiments.

4.2.1.3 Stimuli

Each stimulus presented in training and test consisted of 9 circles on a black 600x600 pixel square background. The stimulus background was divided into a 3x3 grid, with each circle confined to a 200x200 pixel cell so that there was no overlap between circles. The location of the circles within each cell varied randomly such that circle location was not predictive of category membership. The critical dimensions that were varied between categories were the color and size of the circles. The minimum and maximum size values (circle radii) were 15 and 50 pixels respectively, and the minimum and maximum color values (hue) were .403 and .555, with saturation and brightness set to 100% and 75% respectively (see Table 4.3). The test stimuli varied between this full range, but the training stimuli were restricted to varying in a mid-range band that included the middle 52% of values. This was done to ensure that the test stimuli were more extreme than the training stimuli to allow an adequate assessment of generalization along each dimension. An example of how stimuli varied along each dimension can be seen in Figure 4.5.

Table 4.3. Min and max values for the whole dimension (Min, Max) and restricted (middle 52%) dimension (rMin, rMax).

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>Max</th>
<th>rMin</th>
<th>rMax</th>
<th>Multiplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hue</td>
<td>.403</td>
<td>.555</td>
<td>.443</td>
<td>.515</td>
<td>.025</td>
</tr>
<tr>
<td>Size</td>
<td>15</td>
<td>50</td>
<td>24.2</td>
<td>40.8</td>
<td>6</td>
</tr>
</tbody>
</table>

Note. All training stimuli values were set between the restricted minimum and maximum values while all test stimuli were allowed to vary between the extended minimum and maximum values.
For each participant, a ‘base’ prototype was first created which contained randomly chosen hue and size values for each of the 9 circles (in Figure 4.3, the base prototype would be at the intersection of the two dimensions). The color and size values for the base prototype could be any value within the respective ranges. The category prototypes used to create the training stimuli (T in Figure 4.3) were then constructed from the base prototype. For each participant, one category (left/right) was randomly chosen to have larger and the other smaller circles, and one category was chosen to have bluer and the other greener circles. From the base prototype, the color and size values of each circle were distorted in opposing directions to the same degree, such that the size value for any given circle in one category were larger than those of the corresponding circle in the other category, and the hue value for a given circle in one category were larger than the hue values for the corresponding circle in the other category. The exact degree of distortion was determined by multiplying the distortion level (arbitrary level of 0.8) by the feature multiplier (see Table 4.3) and then adding or subtracting these values from the base prototype values to form the category prototypes.

These category prototypes then formed the basis for creating the 120 training stimuli, which all contained the same color and size values but different (randomized) location coordinates within its cell in the stimulus grid. Thus all the category exemplars were unique but similar to each other in terms of the relevant dimensions. To make the training phase more difficult so that participants would not immediately work out a rule on both dimensions, each training stimulus had 2/9 of its color values and 2/9 of its size values (randomly and independently selected for each stimulus) swapped with its respective value in the other category prototype. This effectively meant that the category exemplars seen during training were more similar to each
other than the ‘Train’ stimuli seen on test. Randomizing the locations of the circles and swapping the color and size values served to discourage participants from focusing on a single circle in discrimination, and also added some noise to make the initial discrimination harder.

The test stimuli (Train, Near1, Near2, Near3 and Far, see Figure 4.5 for examples) were spaced at regular intervals (arbitrary distortion level of 0.6) and were created by distorting the category prototype. The test stimuli were created in the same way as the training stimuli (varying hue and size of each of the 9 circles), except there was no swapping of values for the other prototype and only one dimension was varied at a time. This meant that there were 20 different test stimuli (5 per dimension varied, for each of the two categories). Four sets of the 20 test stimuli were created, each with randomized location values amounting to 80 test stimuli in total. For the Consistent group, when one dimension was varied, the other dimension was set at the values of the category prototype. For the Inconsistent group, the non-varied dimension was set at the values of the base prototype (see Figure 4.3 and Table 4.2). Since the base prototype was the starting point to create the two category prototypes, the base prototype values were effectively at the midpoint of both categories, and therefore its values were non-diagnostic of category membership.

4.2.1.4 Procedure

The experiment consisted of three phases, an initial training phase, a categorization test, and a short questionnaire. All instructions were presented within the computer program.

Training Phase. For each participant, the size and color of the categories were randomized independently such that one category would have greener and the other
bluer circles, and one category would have larger and the other smaller circles. Participants first read a cover story explaining that they would see various artworks that belonged to different artists (Evan and Justin), and their task was to work out which artworks belonged to which artist through trial and error. They were warned that the artworks were very similar and thus they might find the task difficult at first, but they would be able to learn about the artworks with feedback.

Before training started, a message appeared for 3 seconds that told participants to get ready for the task by placing their fingers on the left and right shift keys. On each trial, participants were presented with an artwork in the middle of the screen and the names of the artists appeared on the sides of the screen (“Evan” on the left and “Justin” on the right). Participants made a categorization response by pressing the left or right shift key, and had 4 seconds to respond before they were timed out. Once a response was made or the timeout was reached, the stimulus disappeared, and feedback (either “correct” in black, “wrong” in red, or “too slow” in red) was shown in the center of the screen, along with the choice of artist they had made on that trial. Feedback was shown for 500ms and 1000ms later the next trial would begin. There were 120 trials in total (with 120 unique stimuli), and the training phase progressed in exactly the same way for both groups.

*Categorization Test. The instructions for the test phase emphasized that participants would now be shown more artworks that belonged to the same artists, and that they should choose which artist the artwork belonged to by again pressing the left and right shift keys, and then make a typicality rating on a scale using the mouse. They were also told that some of the artworks might look like the artworks seen previously, while others would not. It was emphasized in the instructions that the category judgements and typicality ratings were separate judgements, by telling
participants that it was possible to know the correct category due to high similarity between a test stimulus and the trained stimuli (i.e. high typicality), or despite a lack of similarity between the test stimulus and training stimuli (i.e. low typicality).

The test stimuli consisted of the training stimulus (Train), and transfer stimuli that were more extreme along the dimension (Near1, Near2, Near3 and Far stimuli) for each of the two categories, and for each of the two dimensions. The 20 test stimuli were presented 4 times each, with the location of the circles randomized each time, such that each instantiation of each stimulus was unique (but the critical color and size values were the same). The increment in similarity between all transfer stimuli was the same such that the test stimuli for each category were regularly spaced along the dimension (see Figure 4.5 for examples of Near and Far test stimuli).

On each test trial a stimulus was presented along with the question “Who does this artwork belong to?” appearing underneath. Participants had unlimited time to respond and pressed the left or right shift key to make their decision. Once they had made their choice a new question replaced the previous one asking “How typical of [chosen artist]'s art collection is this artwork?” and a rating scale appeared underneath the question, ranging from “NOT typical” to “VERY typical”. Participants were free to alter their category choice and typicality rating until they were satisfied with both (they could change their categorization judgement even after making their typicality rating), after which they could press spacebar to progress to the next trial. There was a blank inter-trial interval (ITI) of 1000ms.

Questionnaire. Participants answered a series of questions following the categorization test. The first question asked participants how useful they found various dimensions (brightness, size, color, location) of the circles. Four visual analogue scales appeared on the same page (one for each dimension) and participants
made a rating on each scale ranging from “Not useful” to “Very useful”. They could make their four ratings in any order and could only progress to the next question once all ratings had been made. They were then asked to answer a three-alternative forced-choice question (3AFC) and then a two-alternative forced-choice (2AFC) question about one of the two relevant category dimensions, and then the other dimension (with the order of dimensions randomized).

The first 3AFC question read:

“You may have noticed a difference between the left (Evan) and right (Justin) categories in terms of the [COLOUR/SIZE] of the circles. If you did notice a difference, when did you notice it?”

1) During the FIRST phase of the experiment (where there was feedback)
2) During the SECOND phase of the experiment (where there was no feedback, and you had to give typicality ratings)
3) I did not notice a difference

The second 2AFC question read:

“One of the categories had mostly [BLUER/LARGER] circles, while the other category had mostly [GREENER/SMALLER] circles. Which of the following do you think is correct?”

1) LEFT category (Evan) had mostly [GREENER/LARGER] circles, RIGHT category (Justin) had mostly [BLUER/SMALLER] circles.
2) LEFT category (Evan) had mostly [BLUER/SMALLER] circles, RIGHT category (Justin) had mostly [GREENER/LARGER] circles

Both the 3AFC self-report question and the 2AFC rule-identification question were answered by pressing the corresponding number key on the keyboard.

Participants also made a confidence rating for the 2AFC rule-identification question
on a scale ranging from “I’m guessing” to “100% confident”. All scale ratings were transformed to a scale from 0-100. Lastly, participants were asked to indicate whether they suffered from any form of color-blindness.

4.2.2 Results and Discussion

4.2.2.1 Data Analysis

For each set of analyses for the categorization and typicality data, a repeated measures analysis of variance (ANOVA) was conducted using the 5 test stimuli (Train, Near1, Near2, Near3, Far). Since the pattern of generalization across test stimuli was the primary focus, the following analyses will focus on the repeated-measures trends and related interactions. In particular, in the categorization data, the presence of linear and quadratic trends was tested across the dimension since these trends capture linearity and curvature suggestive of monotonic and peak-shifted gradients respectively. However, to show convincingly that a peak shift is present in the Inconsistent group, a stronger test is needed to show that the gradient is non-monotonic. Therefore, the point with the highest accuracy out of the Near1, Near2, and Near3 test stimuli was chosen and t-tests conducted to test whether there was a significant rise in accuracy from the Train stimulus to this peak, and a significant fall in accuracy from the peak to the Far stimulus. These t-tests were one-tailed since they necessarily entail a comparison between the highest point and the extreme end points on the dimension. To maintain the Type I error rate at .05, the critical alpha value was Bonferroni-corrected. Thus, if both t-tests are significant at the .017 level, then we can infer that the generalization gradient demonstrates a peak-shift effect (similar to Livesey & McLaren, 2009). To determine whether a monotonic gradient of
generalization was present in the Consistent group, a linear trend analysis was conducted using the five test stimuli.

4.2.2.2 Exclusion Criteria

To ensure that the data analysis included only participants who learned something about the categories, similar to Livesey and McLaren (2009), participants who scored less than 55% in the last half of training were excluded (32 participants, 25.2% of the sample). After applying this criterion, a total of 95 participants remained (52 in the Consistent and 43 in the Inconsistent group).

4.2.2.3 Questionnaire

Figure 4.6 shows the results of the 3AFC self-report question asking participants if and when they noticed a difference between the categories. Participants seemed to have found differences in color more noticeable than differences in the size of the circles, with a larger number noticing color during training and few failing to notice the difference at all, compared to size. Table 4.4 shows the number of participants who noticed vs. did not notice differences in each dimension during training. Participants who claimed to have noticed differences in test were grouped with those who did not notice differences at all since the effect of the test manipulation specifically required participants to form a relational rule during training (as has been found to be critical in Livesey & McLaren, 2009, and Natal, McLaren, & Livesey, 2013). It can be seen from Table 4.4 that the majority of participants noticed differences in only one dimension during training, and that the probability of noticing each dimension was lower if participants had also noticed a difference in the other dimension. This suggests that in the absence of any instructions,
participants were more likely to notice differences in a single dimension, rather than both dimensions. Looking at the results from the 2AFC rule identification question (Figure 4.7), participants remaining after the exclusion criteria were applied tended to perform equally well at identifying the category difference in terms of color and size.

![Figure 4.6](image)

*Figure 4.6. Proportion of participants who selected each option for the 3AFC self-report question (after exclusions) for each dimension (left panels: color, right panels: size) in Experiment 1.*

<table>
<thead>
<tr>
<th></th>
<th>Size</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Noticed in Training</td>
<td>Did not notice in training</td>
</tr>
<tr>
<td>Color</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Noticed in Training</td>
<td>20</td>
<td>42</td>
</tr>
<tr>
<td>Did not notice in training</td>
<td>23</td>
<td>10</td>
</tr>
<tr>
<td>TOTAL</td>
<td>43</td>
<td>52</td>
</tr>
</tbody>
</table>

Table 4.4. *Number of participants who reported noticing a difference in each dimension during training vs. during test/not at all after exclusions.*

probability (noticing size) = .453 (43/95)
probability (noticing size | noticing color) = .323 (20/62)
probability (noticing color) = .653 (62/95)
probability (noticing color | noticing size) = .465 (20/43)
4.2.2.4 Training

The training data were analyzed in a 2x(4) analysis of variance (ANOVA) with test group (Consistent vs. Inconsistent) as the between-subjects factor and training quarter (1-4) as the within-subjects factor. Training accuracy increased over the four training quarters as shown by a significant linear trend, $F(1,93) = 80.4, p < .001, \eta_p^2 = .464$. This linear trend did not interact with test group, $F < 1$, and there was no overall difference in accuracy between groups, $F < 1$. This analysis was also repeated with the full sample and produced the same results.

4.2.2.5 Category Judgements

The hypothesis concerning the effect of disrupting rule application on test requires participants to form at least one relational rule using one of the dimensions. Thus, for the following analyses participants were divided into 2 subgroups: those who reported noticing differences in at least one dimension ($n = 48$ in the Consistent group, $n = 37$ in the Inconsistent group), and those who reported noticing neither dimension ($n = 4$ in the Consistent group, $n = 6$ in the Inconsistent group) during
training. These subgroups were not directly compared due to unequal and low n for those who failed to notice any differences during training. Since the hypothesis for this experiment applies specifically to the dimension on which rule-based generalization was most likely to occur in the Consistent group (and therefore the dimension on which rule disruption was most likely to occur in the Inconsistent group), generalization along the dimension that participants were attending to more was of particular interest. The ‘more attended’ dimension for each participant was classified as the dimension that participants claimed to have noticed a difference in during training. If participants had noticed both or neither dimension during training, whichever dimension participants rated as being more useful was then chosen as the more attended dimension.

Figure 4.8 illustrates the generalization gradients for the more attended dimension, split according to whether participants noticed at least one relevant difference during training. Note that the dimensions have been folded such that the point labeled “Far”, for example, incorporates the Far stimulus from both categories (see Appendix D for the full unfolded dimension). In other words, if the relevant dimension is color, the “Far” point represents the stimuli that have the greenest and the bluest circles. Generalization along each dimension was assessed by a 2x(5) ANOVA with test group (Consistent vs. Inconsistent) as the between-subjects factor and test stimulus (Train, Near1, Near2, Near3, Far) as the within-subjects factor.
Figure 4.8. Categorization accuracy and typicality ratings for variations on the ‘more attended’ dimension in Experiment 1. Participants were split into 2 subgroups for the analyses: those who noticed a difference in at least one dimension during training (\(n = 48\) in the Consistent group, \(n = 37\) in the Inconsistent group), and those who did not notice differences in either dimension during training (\(n = 4\) in the Consistent group, \(n = 6\) in the Inconsistent group). Error bars represent the standard error of the mean. Note the different scales used in the left panels.

‘More attended’ dimension. For participants who noticed at least one dimension (Figure 4.8a), there was an overall linear trend, \(F(1,83) = 4.16, p = .044, \eta^2_p = .048\), and a significant quadratic trend, \(F(1,83) = 13.47, p < .001, \eta^2_p = .140\). Neither of these trends interacted with group, largest \(F < 1\), and there was also no main effect of group, \(F(1,83) = 2.24, p = .138, \eta^2_p = .026\). For participants who claimed to have noticed neither dimension during training (Figure 4.8c), no significant trends nor interactions were found, \(Fs < 1\), and there was also no main effect of group, \(F < 1\). There was a significant linear trend in the Consistent group for the subgroup that noticed at least one dimension, \(F(1,47) = 24.97, p < .001, \eta^2_p = .347\).
but not for the subgroup that did not notice either dimension, $F(1,3) = 3.61, p = .153$, $\eta^2_p = .546$. There was no significant quadratic trend in either subgroup, largest $F < 1$.

For the Inconsistent group, there was a significant rise in accuracy using the Near1 point in the subgroup that noticed at least one dimension during training, $t(36) = 3.89, p < .001$, $d = .697$, but no significant fall in accuracy to the Far point, $t(36) = .947, p = .175$, $d = .159$. There was no evidence of peak shift in the Inconsistent group in the subgroup that failed to notice either dimension, largest $t(5) = .674, p = .530$, $d = .303$. The subgroup who noticed at least one dimension during training also showed a significant linear trend in their generalization gradient, $F(1,36) = 6.59$, $p = .015$, $\eta^2_p = .155$, but no other linear or quadratic trends were significant in either subgroup, largest $F(1,5) = 4.37, p = .091$, $\eta^2_p = .466$. It seems that both the Consistent and the Inconsistent groups generalized on the basis of a relational rule, but only if they noticed differences between the categories in at least one dimension. The test manipulation did not seem to have an effect on the pattern of generalization nor on overall accuracy, although participants in the Consistent group did perform at a numerically higher level of accuracy overall than the Inconsistent group.

‘Less attended’ dimension. For the dimension that participants did not notice in training or rated as being less useful (Figure 4.9a and 4.9c), there was no overall linear or quadratic trend, nor interactions with group, in either subgroup, largest $F(1,83) = 3.78, p = .055$, $\eta^2_p = .044$. There was a significant overall group difference in accuracy for participants who noticed at least one dimension (Figure 4.9a), $F(1,83) = 94.09, p < .001$, $\eta^2_p = .531$, but not for those who noticed neither dimension (Figure 4.9c), $F(1,8) = 1.27, p = .292$, $\eta^2_p = .137$. The high level of accuracy attained by the Consistent group can be explained by the fact that on these trials, information on the other (more attended) dimension was present, and participants must have relied on
this dimension for their category judgements. The Inconsistent group on the other hand, did not have information on the other (attended) dimension and accuracy is therefore around chance\textsuperscript{26}. The flat generalization gradients in each group and subgroup indicate that despite a number of participants self-reporting noticing both dimensions ($n = 20$) or neither dimension ($n = 10$) during training, generalization on test was dominated by the use of a single dimension such that there was no stimulus control by the less attended dimension. There were no significant linear or quadratic trends in any subgroup, largest $F(1,5) = 5.62, p = .064, \eta_p^2 = .529$.

\textbf{Figure 4.9.} Categorization accuracy and typicality ratings for variations on the ‘less attended’ dimension in Experiment 1. Participants were split into 2 subgroups for the analyses: those who noticed a difference in at least one dimension during training ($n = 48$ in the Consistent group, $n = 37$ in the Inconsistent group), and those who did not notice differences in either dimension during training ($n = 4$ in the Consistent group, $n = 6$ in the Inconsistent group). Error bars represent the standard error of the mean.

\textsuperscript{26} Note that overall, categorization accuracy for test stimuli that varied on the unattended dimension was actually significantly greater than chance in the Inconsistent group, $t(42) = 1.94, p = .030$ (one-tailed).
4.2.2.6 Typicality Ratings

‘More attended’ dimension. The typicality ratings are shown in Figure 4.8 and 4.9 (right panels), and were analyzed in a similar manner as the categorization data but only testing for linear trends. For the more attended dimension, for the subgroup that noticed at least one dimension (Figure 4.8b), there was a significant overall linear trend, smallest $F(1,83) = 27.60, p < .001, \eta_p^2 = .250$, that did not interact with group, largest $F(1,83) = 2.62, p = .109, \eta_p^2 = .031$, and there was no overall group difference, $F < 1$. For the subgroup that did not notice either dimension during training (Figure 4.8d), there was no overall linear trend, $F(1,8) = 4.23, p = .074, \eta_p^2 = .346$, no interaction with group, $F < 1$, and no overall group difference, $F < 1$.

‘Less attended’ dimension. There was an overall linear trend in the subgroup of participants that noticed at least one dimension (Figure 4.9b), $F(1,83) = 31.71, p < .001, \eta_p^2 = .276$, but not in those who did not notice either dimension (Figure 4.9d), $F(1,8) = 2.34, p = .165, \eta_p^2 = .226$. Neither subgroup’s linear trend interacted with group, largest $F(1,8) = 1.53, p = .251, \eta_p^2 = .161$. There was no overall group difference in either subgroup, largest $F(1,83) = 3.06, p = .084, \eta_p^2 = .036$. Participants in both Consistent and Inconsistent groups rated more extreme stimuli along both dimensions as less typical of the categories seen during training, and the Train stimulus as the most typical, but only when they noticed differences in at least one dimension during training.

4.2.2.7 Summary

In summary, participants who were able to notice differences in the categories in at least one dimension showed a linearly increasing gradient in their categorization accuracy on the dimension that they were deemed to be attending to most. They also
produced linearly descending typicality ratings along this dimension. In contrast, the small number of participants who did not notice differences in either dimension did not show either of these trends in categorization or typicality. Although in the predicted direction, the Inconsistent group did not show a lower level of categorization accuracy for variations on the attended dimension, and there was no effect of the test manipulation on the pattern of generalization or typicality ratings. One potential explanation is that although the lack of stimulus control by the ‘less attended’ dimension implies that participants were primarily using one dimension on test, a substantial number of participants ($n = 20$) also claimed to have noticed differences in both dimensions during training. Thus, their attention may have been divided over these two dimensions during training and therefore learning (and rule formation) for either dimension may have been poorer than those who focused on one. Since the 3AFC self-report question does not indicate whether participants noticed differences at the beginning or end of training, it may be that a large proportion of participants only formed a rule at the end of training. While this may be sufficient to show a linear gradient on test, it may be that the test manipulation requires reliance on a rule throughout the entirety of training in order for use of that rule to be disrupted by subsequently reducing its applicability. To speculate, perhaps a strongly-formed rule and directed attention to the relevant dimension are needed in order for the Inconsistent group to notice that there are test trials in which their rule is not applicable, or alternatively to make the contrast in rule applicability between training and test more pronounced and lead to the desired degree of uncertainty in their rule. Thus, in Experiment 2 explicit instructions were provided to attend to a particular dimension in order to encourage rule formation and reliance on that rule from the beginning of training.
4.3 Experiment 2

Participants in Experiment 2 were either told to attend to the color (Attend Color group) or the size (Attend Size group) of the circles to increase the reliance on a relational rule, and therefore increase the effect of the test manipulation. Previous studies have shown that verbal instructions can determine the dimension along which participants subsequently generalize for simple stimuli with both shape and color dimensions (Vervliet, Kindt, Vansteenwegen, & Hermans, 2010), and therefore it seemed reasonable that directing participants to attend to a particular dimension would allow participants to derive a rule on one dimension but not the other.

Participants were again allocated to a Consistent or Inconsistent test group, making the experiment a 2x2 between-subjects design. It was expected that providing a hint to attend to one of the dimensions would increase the reliance on a rule during training and therefore a reduction of rule applicability on test would have a more substantial impact on the Inconsistent group relative to Experiment 1. Thus, regarding generalization along the dimension that participants were instructed to attend to, the predictions were that a monotonic gradient of generalization would result in the Consistent Group consistent with rule-based generalization, and there would be lower accuracy and similarity-based generalization in the Inconsistent Group.

Experiment 2 also sought to rule out an alternative explanation for increased rule-based generalization observed in the Consistent group relative to the Inconsistent group. The Consistent and Inconsistent groups differ not just in the ease of application of rules on test, but also in the physical characteristics of the test stimuli. As depicted in Figure 4.3, the Consistent group always has diagnostic information present on both dimensions while the Inconsistent group only has diagnostic information from a single dimension. Thus, considering the critical ‘attended dimension’ trials, the
Consistent group has ‘extra’ diagnostic information on test, which may have aided rule use\textsuperscript{27}. Having diagnostic information available during test also makes the testing context more similar to the training context than in the Inconsistent group. This is particularly important because despite the proposal that rule use should not be affected by the similarity between training and target domains (Smith, Langston, & Nisbett, 1992), research in different paradigms has shown that in practice, rule use \textit{is} affected by the similarity between the training and test stimuli even when the rule is perfectly valid (e.g. Allen & Brooks, 1991; Nosofsky, Clark, & Shin, 1989; Hahn, Prat-Sala, Pothos, & Brumby, 2010) and even when using similarity is detrimental to performance (Hahn et al., 2010). Consequently, to rule out this alternative explanation for any potential group differences, two experiments (Experiment 2A and 2B) were run that were identical except for a minor procedural manipulation to test whether the presence of information on the unattended dimension affected generalization on the attended dimension.

Experiment 2A was similar to Experiment 1 where the critical test trials varying the attended dimension were not equated between groups (see Figure 4.3). In Experiment 2B, the critical ‘attended dimension’ test trials were equated such that neither group had information on the unattended dimension when the attended dimension was varied (see Figure 4.4). Note that this does not change the fact that in both Experiments 2A and 2B, the Consistent group would be able to use a relational rule on their attended dimension on all test trials while the Inconsistent group would only be able to use a rule on their attended dimension on half of the test trials. In addition to showing that any monotonic gradients in the Consistent group are not due to having extra diagnostic information, equating the test trials also provides stronger

\textsuperscript{27} Note that this is highly unlikely due to Experiment 1 showing little evidence of learning about the unattended dimension. Still, it was important to account for this confound in the design.
evidence that any observed group differences are attributable to the reduction of rule applicability.

4.3.1 Method

4.3.1.1 Participants

One hundred and thirty-nine (M age = 20.88, SD = 6.22, 96 females) and one hundred and forty (M age = 19.48, SD = 3.17, 92 females) University of Sydney first-year Psychology students participated in Experiment 2A and 2B respectively in exchange for partial course credit or payment (AUD$15/hour). Participants were again randomly allocated to either the Consistent (n = 134) or Inconsistent (n = 145) group. Similarly to Experiment 1, participants who indicated that they were colorblind were excluded (6 participants).

4.3.1.2 Procedure

The procedure was identical to Experiment 1 except that participants were randomly allocated to attend to either the color or size of the circles (Attend Color vs. Attend Size). The instructional manipulation was presented on a separate screen after the training instructions and read:

“You should try to use as much information about the circles to categorize the stimuli. However, the [COLOUR/SIZE] of the circles may be most helpful in distinguishing Evan’s and Justin’s artworks. Try and attend to the [COLOUR/SIZE] of the circles when learning the categories.”

The experiment was thus a 2x2x2 between-subjects design with attention group (Attend Color or Attend Size), test group (Consistent vs. Inconsistent), and experiment (2A vs. 2B) as the independent variables. A manipulation check was
added to the end of the questionnaire phase asking participants “Which dimension were you asked to attend to? Press 1 for colour or 2 for size”. Participants were still asked the same 3AFC self-report and 2AFC rule-identification questions concerning whether/when they noticed a difference in each dimension, and identifying the difference between the categories, but they were always asked about their attended dimension before their unattended dimension.

4.3.2 Results and Discussion

For the ensuing analyses only the results of relevance to the main hypothesis will be reported (the full set of statistical results can be found in Appendix E).

4.3.2.1 Exclusion Criteria

To ensure that the data analyzed were from participants who used their hint to derive the appropriate rule, participants were excluded if they reported not noticing a difference in their attended dimension during training (58 participants, 21.2% of the sample). It was assumed that despite the difficulty of the discrimination, giving participants a hint should have made it easy to eventually discover a difference between the categories in their attended dimension during training. Thus any participants who failed this criterion were probably not following the instructions. Participants who failed the manipulation check were also excluded (a further 17 participants, 7.9% of the remaining sample) for the same reason. Note that neither of these exclusion criteria require participants getting the 2AFC rule-identification question correct. As with Experiment 1, participants were also excluded if they scored < 55% accuracy in the last half of training (a further 11 participants, 5.6% of the
remaining sample). After applying these criteria there were a total of 187 participants remaining (97 in the Consistent group and 90 in the Inconsistent group).

4.3.2.2 Questionnaire

Figure 4.10 shows the proportion of participants in each of the four groups that selected each option in the 3AFC self-report question for color and size in Experiments 2A (top panels) and 2B (bottom panels) after exclusions. It can be seen that across both experiments in both Consistent and Inconsistent groups, the majority of participants remaining after exclusions failed to notice differences in their unattended dimension during training.
Figure 4.10. Proportion of participants who selected each option for the 3AFC self-report question (after exclusions) for each dimension (left panels: color, right panels: size) in Experiment 2A (top panels) and 2B (bottom panels). Note that participants who reported not noticing differences in their attended dimension during training were excluded.

Figure 4.11 shows the proportion of participants in each group who were able to identify the correct difference between the categories after applying the exclusion criteria in Experiment 2A (a and b) and Experiment 2B (c and d). Accuracy is generally quite high, and there was significantly higher accuracy for color when participants were told to attend to color, $\chi^2(1, N = 187) = 16.94, p < .001$ (left panels), and similarly for size, $\chi^2(1, N = 187) = 7.24, p = .007$ (right panels). The proportion of participants identifying the rule for color and size did not differ between experiments, largest $\chi^2(1, N = 187) = 3.32, p = .068$, but a larger proportion of
participants in the Consistent group were able to identify the correct category difference for color, $\chi^2(1, N = 187) = 8.55, p = .003$, and also for size, $\chi^2(1, N = 187) = 11.72, p = .001$. A possible explanation for these group differences in rule identification is that the reduction of rule validity at test may have resulted in more errors when attempting to identify the rule in the subsequent questionnaire.

![Figure 4.11](image) Accuracy for the 2AFC rule-identification question (after exclusions) for Experiment 2A (upper panels) and Experiment 2B (lower panels).

4.3.2.3 Training

There was a significant increase in accuracy over the four training blocks, $F(1,179) = 130.4, p < .001$, $\eta_p^2 = .421$, that did not interact with test condition, $F(1,179) = 1.57, p = .211$, $\eta_p^2 = .009$, attention group, $F < 1$, nor experiment, $F < 1$.

As with Experiment 1, the analysis was repeated for the whole sample. The results were the same except that the interaction between training block and experiment was significant, $F(1,271) = 4.23, p = .041$, $\eta_p^2 = .015$. 

193
4.3.2.4 Category Judgements

**Attended Dimension.** The categorization results for variations along the attended dimension are shown in Figure 4.12a. Inspecting Figure 4.12a, the overall pattern is similar to Experiment 1 (Figure 4.8a), and the Consistent Group seems to be performing better than the Inconsistent Group. While a monotonic gradient appears to be present in the Consistent Group, the Inconsistent Group appear to show a different pattern of generalization that resembles a peak shift. The categorization data were analyzed in a 2x2x2x(5) ANOVA with experiment, attention group, and test group as between-subjects factors and test stimulus as the within-subjects factor. As predicted, the Consistent group’s accuracy was significantly higher overall than the Inconsistent group, $F(1,179) = 10.59, p = .001, \eta_p^2 = .056$. There was an overall linear and quadratic trend, smallest $F(1,179) = 24.29, p < .001, \eta_p^2 = .120$, and a significant interaction between the quadratic trend and test group, $F(1,179) = 5.00, p = .027, \eta_p^2 = .027$, but no significant interaction between the linear trend and group, $F(1,179) = 1.03, p = .312, \eta_p^2 = .006$.

![Diagram](image)

*Figure 4.12. Categorization accuracy and typicality ratings for variations on the attended dimension in Experiment 2 (combining Experiments 2A and 2B). Error bars represent the standard error of the mean.*
Since there was a significant interaction between the quadratic trend and group (indicating group difference in the pattern of generalization), planned analyses were conducted in a similar manner to Experiment 1, testing for a linear trend in the Consistent group and a peak shift in the Inconsistent group. In the Consistent group, there was a significant linear trend, $F(1,93) = 24.34, p < .001, \eta^2_p = .207$, indicating that a monotonic gradient was present and suggesting that generalization occurred on the basis of a relational rule. To test for the presence of peak shift in the Inconsistent group, the highest of the three middle points (Near1, Near2, Near3) was again compared to the two endpoints (Train and Far) and the critical alpha value was Bonferroni-corrected to .017. Using the Near1 point, there was a significant rise in accuracy from the Train stimulus, $t(89) = 4.90, p < .001, d = .526$, and a significant fall in accuracy to the Far stimulus, $t(89) = 2.42, p = .009, d = .261$, indicating a peak-shifted gradient. The presence of peak shift in the Inconsistent group suggests that when rule validity is reduced on test, rule-based generalization was disrupted and instead, participants generalized on the basis of similarity to physical features of the stimuli.

**Unattended Dimension.** Categorization accuracy for generalization along the unattended dimension is shown in Figure 4.13a. The data were analyzed in a similar way to Experiment 1, but looking only at linear trends and associated interactions. Similarly to Experiment 1, there was no significant linear trend overall, $F < 1$, and this did not interact with test group, $F(1,179) = 1.14, p = .287, \eta^2_p = .006$, attention group, $F < 1$, nor experiment, $F < 1$. Again, the Consistent group had significantly higher categorization accuracy than the Inconsistent group, $F(1,179) = 395.3, p < .001,$
\( \eta_p^2 = .688 \), presumably because the Consistent group were able to rely on their attended dimension while the Inconsistent group could not\(^{28}\).

4.3.2.5 Typicality Ratings

**Attended Dimension.** The typicality ratings for the attended dimension are shown in Figure 4.12b, and were analyzed in the same way as in Experiment 1. There was a significant linear trend, \( F(1,179) = 98.57, p < .001, \eta_p^2 = .355 \), which interacted with test group, \( F(1,179) = 22.08, p < .001, \eta_p^2 = .110 \), but not with attention group, \( F(1,179) = 3.14, p = .078, \eta_p^2 = .017 \), nor experiment, \( F < 1 \). The linear trend was also significant within the Consistent and Inconsistent groups, smallest \( F(1,86) = 13.22, p < .001, \eta_p^2 = .133 \). Similarly to Experiment 1, the linear trends show that for the Consistent group, the highest categorization accuracy was achieved for the test stimuli that participants considered to be least typical of the category, providing further

\(^{28}\) In contrast to Experiment 1, the Inconsistent group did not perform significantly better than chance at categorizing test stimuli that varied on the unattended dimension, \( t(89) = 1.10, p = .137 \) (one-tailed).
evidence that participants in the Consistent group were generalizing on the basis of a
relational rule. A monotonic gradient in categorization accuracy can always be argued
to be the rise of a peak shift that spans a large width of the category dimension
(Livesey & McLaren, 2009), but the presence of these descending typicality gradients
confirms that the monotonic gradient in categorization is based on more than just the
perceived similarity of the test stimuli to those seen in training.

Unattended Dimension. There was a significant overall linear trend for
typicality ratings in the unattended dimension (Figure 4.13b), $F(1,179) = 76.6, p$
< .001, $\eta_p^2 = .300$, with steeper gradients in the Consistent group than the Inconsistent
group, $F(1,179) = 9.56, p = .002, \eta_p^2 = .051$. The linear trend was also significant in
both Consistent and Inconsistent groups, smallest $F(1,86) = 18.52, p < .001, \eta_p^2 = .177$.

4.3.2.6 Summary

In Experiment 2A and 2B, participants were explicitly instructed to attend to a
particular dimension. For category judgements of stimuli varying along the attended
dimension, evidence was found of reduced accuracy in the Inconsistent group, as well
as a different pattern of generalization to the Consistent group. Importantly, neither of
these effects interacted with experiment, suggesting that reducing the applicability of
a rule on test disrupted rule-based generalization, and this was not due to differences
in the physical stimuli presented. Planned analyses revealed that in the Consistent
group, a monotonic gradient of generalization was present, consistent with rule use,
and in the Inconsistent group, a significant peak shift effect was found, consistent
with generalization on the basis of similarity to physical features. Further evidence for
generalization on the basis of a rule was provided by the typicality ratings, where the
Consistent group performed at the highest level of accuracy for stimuli they rated to
be the least typical of the category. Interestingly, the test manipulation affected not just category judgements but perceptions of typicality, with the Inconsistent group showing a flatter gradient in their typicality ratings. While there were some higher-order interactions with attention group, and experiment (see Appendix E for the full results), the key results discussed above were not found to interact with experiment or attention group.

The peak shift in the Inconsistent group is surprising given previous literature demonstrating that peak shift only occurs in the absence of a relational rule (e.g. Aitken, 1996; Livesey & McLaren, 2009), and since the majority (77/90) of participants in the Inconsistent group were able to correctly identify the rule-relevant difference between categories. The fact that participants were explicitly directed to use a particular dimension during training, and the fact that participants were excluded if they claimed that they did not notice any difference between the categories during training given this hint, leaves little chance that the remaining participants were mostly guessing correctly.

It should also be noted that while a peak shift was found in the Inconsistent group, the linear trend was also significant, $F(1,86) = 5.92, p = .017, \eta_p^2 = .064$, suggesting that perhaps participants were generalizing on the basis of a rule and similarity. This might be expected if participants have derived a rule and are simply unsure whether to use it and therefore show a mixture of generalization gradients. The effect of reducing rule validity might therefore be to undermine certainty or confidence in applying the rule, and it may be the case that this certainty changes on a trial-by-trial basis, with the pattern of generalization dictated by rule applicability on the previous trial.
4.4 Sequential Reanalysis

To test this hypothesis, a post-hoc sequential analysis was conducted of the categorization data in both experiments whereby the critical test trials where the attended dimension was varied were divided into ‘repetition’ and ‘alternation’ trials. Repetition refers to trials where the previous trial varied the same dimension (and therefore a rule on the attended dimension was previously applicable in both groups), and alternation refers to trials where the previous trial varied the other (unattended) dimension (and therefore a rule on the attended dimension was not previously applicable in the Inconsistent group but was applicable in the Consistent group). The data from the subgroup in Experiment 1 who were able to notice differences in at least one dimension during training were combined with the data from Experiment 2, with experiment (1 vs. 2) added to the analysis as a between-subjects factor. The results for the sequential analysis are shown in Figure 4.14a, with the corresponding typicality ratings shown in Figure 4.14b. It is firstly apparent that the Consistent groups show roughly monotonic gradients for both repetition and alternation trials, but the Inconsistent groups are showing different patterns of generalization based on the previous trial. Since this is a post-hoc analysis, only the results of interest will be reported.
**Figure 4.14.** Sequential analysis: categorization accuracy and typicality ratings for variations on the attended dimension on trials where a rule was applicable on the previous trial (repetition trials) or was not applicable on the previous trial (alternation trials). The data combine the subgroup from Experiment 1 who reported noticing differences in at least one dimension during training \((n = 85)\) and participants in Experiment 2 \((n = 187)\).

### 4.4.1 Inconsistent Group

The data for the Inconsistent group were analyzed in a 2x(2x5) ANOVA with experiment (1 vs. 2) as the between-subjects factor and previous trial (repetition vs. alternation) and test stimulus as within-subjects factors. There was an effect of previous trial on overall accuracy, \(F(1,125) = 5.34, p = .022, \eta_p^2 = .041\), with lower accuracy for trials where the rule could not be used on the previous trial. There was a significant overall linear trend, \(F(1,125) = 5.42, p = .021, \eta_p^2 = .042\), that did not interact with experiment, \(F < 1\), but importantly, did interact with previous trial, \(F(1,125) = 4.77, p = .031, \eta_p^2 = .037\). The interaction between linear trend and previous trial did not further interact with experiment, \(F < 1\). Planned analyses showed that a significant linear trend was present in the repetition trials, \(F(1,125) = 10.65, p = .001, \eta_p^2 = .079\), but not in the alternation trials, \(F < 1\). Instead for the alternation trials, using the Near1 point, there was a significant rise in accuracy from
the Train stimulus, \( t(126) = 4.99, p < .001, d = .416 \), and a significant fall in accuracy
to the Far stimulus, \( t(126) = 3.68, p < .001, d = .330 \) (see Figure 4.14a). Thus while an
overall linear trend was found, this appears to have been driven by the repetition trials,
suggesting that participants were consistently using a relational rule only on trials
where the previous trial varied the attended dimension, which might have made them
more certain that their rule was applicable. On trials where the previous trial varied
the other dimension and their rule could not be used, participants did not consistently
generalize using a relational rule but instead generalized on the basis of similarity,
showing a peak shift.

4.4.2 Consistent Group

The Consistent group also showed a significant overall linear trend, \( F(1,143) = 11.38, p = .001, \eta_p^2 = .074 \), but unlike the Inconsistent group, this did not interact
with previous trial, \( F < 1 \) (see Figure 4.14a). There was also no main effect of
previous trial, \( F < 1 \), and none of these results interacted with experiment, largest
\( F(1,143) = 1.64, p = .203, \eta_p^2 = .011 \). Whether the same or different dimension was
varied on the previous trial had no effect on generalization in the Consistent group,
who showed consistent rule-based generalization throughout test, since they could use
their rule on every trial.

4.4.3 Summary

Within the same participants in the Inconsistent group, a peak-shifted and
monotonic gradient of generalization were both observed, depending on the
applicability of a rule on the previous trial. In contrast, the typicality ratings in the
Inconsistent group did not appear to differ in the same way (see Figure 4.14b),
suggesting that what was changing on a trial-by-trial basis was participants’ tendency to use a relational rule based on its recent applicability, rather than changing perceptions of similarity between the test and training stimuli. This analysis reinforces the conclusion that disrupting rule use serves to undermine participants’ certainty or proficiency in applying their rule and leads them to switch between generalizing on the basis of rules and similarity on a trial-by-trial basis. The results of this sequential analysis indicate that participants apply rules in a flexible way as a function of recent difficulty in the applicability of the rule, suggesting a possible condition that can limit rule use.

4.5 General Discussion

In two experiments, categorization accuracy and typicality ratings were compared between a group who were able to derive a relational category rule during training and apply that rule consistently on test, and a group who could only apply that rule easily on half of the trials on test and therefore only use that rule inconsistently. During training, the color and size of stimulus features were predictive of category membership and it appeared that in Experiment 1, and especially in Experiment 2, participants attended to one of the dimensions and used that dimension to derive a relational category rule. On test, participants either experienced test trials where information was present on both color and size dimensions, making a rule on the attended dimension relatively easy to apply on all trials (Consistent group), or test trials where information was only present on the attended dimension on half of the test trials, making a rule concerning the attended dimension very difficult to apply on 50% of trials and thereby rendering its application on test inconsistent (Inconsistent group). The test trials of interest were those in which the rule was clearly valid and
easy to apply in both groups (i.e. where the attended dimension was varied). Although participants showed good evidence of learning and reported noticing simple relational rules, when rule application was made difficult on test, a peak-shifted generalization gradient emerged, with declining accuracy at the extrema of the test range. This suggests that participants engaged in both rule-based and feature-based generalization during test, with the expression of each varying according to the difficulty of applying the rule on the previous trial.

Although the general pattern of results were similar between Experiment 1 and Experiment 2 (compare Figure 4.8a and 4.12a), predicted differences in the pattern of generalization between Consistent and Inconsistent groups were not apparent statistically in Experiment 1. Experiment 1 found numerically lower categorization accuracy in the Inconsistent group but the resulting generalization gradient was roughly monotonic in both Consistent and Inconsistent groups (based on a significant positive linear trend and the absence of statistical evidence of peak-shift). In Experiment 2, when participants were explicitly told to attend to a particular dimension in order to encourage reliance on a rule during training, divergent patterns of generalization in the Consistent and Inconsistent groups became much clearer. The Inconsistent group performed at a lower level of accuracy than the Consistent group and also produced a different pattern of generalization. The Consistent group in Experiment 2 showed a monotonic gradient of generalization consistent with rule use. Although the Inconsistent group also showed an increasing gradient as revealed by a significant linear trend, the gradient was also peak-shifted, with a significant decline in accuracy at the extrema despite a large proportion of participants (77/90) being able to identify the relational category rule. This is, to the best of the author’s knowledge, the first demonstration of peak shift in categorization despite self-reported knowledge...
of a simple relational rule derived during training, and suggests that when rule use is disrupted, participants revert to generalizing on the basis of similarity.

Further, the sequential analysis showed a within-participant dissociation in generalization based on whether the rule was applicable on the previous trial. Rule applicability on the previous trial had no effect on the monotonic gradients in the Consistent group, while in the Inconsistent group the overall linear trend interacted with previous trial. Follow-up analyses in the Inconsistent group showed a peak shift only on trials where they could not use their primary rule easily on the previous trial, and a monotonic gradient only on trials where they could use their rule on the previous trial. This not only confirms that participants in the Inconsistent group did acquire a rule but also reinforces the conclusion that reducing rule applicability on test serves to undermine the likelihood of applying that rule. In contrast, participants in the Consistent group showed a monotonic gradient that was unaffected by the nature of the previous trial, presumably since their rule was applicable on all test trials.

4.5.1 Mechanisms Responsible for Monotonic Generalization Gradients

Monotonic gradients of generalization over a wide range of test stimuli are most easily interpreted as indicative of the use of a relational rule. However, one problem with assuming that monotonic categorization is synonymous with relational rule use is that the same pattern can be derived from other psychological processes. An associative model (e.g. Ghirlanda & Enquist, 1998) can, in principle, simulate a monotonic gradient if a broad generalization gradient is assumed which spans a wide range of the dimension (Livesey & McLaren, 2009). The monotonic gradient can thus be interpreted as the rise in accuracy in a peak shift that has not been given the opportunity to come down due to testing a limited range. The inclusion of typicality
ratings addresses this problem to some extent. It is clear that the Consistent group, who show greater evidence of a monotonic gradient, do not show broader generalization on the typicality ratings. In fact, generalization gradients for typicality fell more sharply in this group compared to the Inconsistent group. The typicality ratings complement the evidence from self-report measures that monotonic gradients are correlated with identification of the underlying relationships between stimuli. This evidence is important because relying on self-report measures to divide and compare participants is effectively correlational and, furthermore, it has been argued that participants have poor introspective skills and can report knowledge of rules that were not responsible for their behavior at test (Nisbett & Wilson, 1977). The declining typicality ratings for the Consistent group lend some validity to using self-report measures to assess rule use and category knowledge.

4.5.2 Mechanisms Responsible for Peak-shifted Generalization Gradients

It is tempting to conclude that the peak shift shown in the Inconsistent group in Experiment 2 is due to the same basic associative processes used to explain peak shift in animals, especially since it was found under conditions where rule application was hindered. Error-correction learning models account for the peak shift phenomenon with impressive quantitative precision (Blough, 1975; Ghirlanda & Enquist, 1998; McLaren & Mackintosh, 2002), while it is not clear how a rule-based account could explain why accuracy was best at a stimulus slightly removed from the training values. When considered in light of the human literature on peak shift, the results in this chapter seem wholly consistent with the idea that associative processes always operate in human categorization or discrimination studies, but are usually masked by higher-order rule-learning which dominates performance at test (Livesey
Relational rules, especially simple ones that are easy to articulate and apply, should reduce the difficulty of discriminating between categories and thus it is to the participants’ advantage to search for a rule to improve their performance. If rules have primacy over associative information in governing responding, then only when the task requirements and stimuli make it difficult to derive a rule (e.g. Aitken, 1996; Wills & Mackintosh, 1998) can evidence of peak shift emerge.

It is also possible that participants in the Inconsistent group were generalizing on the basis of a conservative similarity rule that describes the physical characteristics of the experienced category. Note that the content of learning in a rule of this form is still clearly different to the content of learning of a relational rule that describes the relationships between the stimuli. Peak shift can be explained if we assume that participants are adopting a conservative decision-rule (e.g. “respond left when the circles are mostly small, but not too small”) whereby the training stimuli are still too similar to each other to enable confident categorization but the stimuli that are slightly more extreme are less likely to suffer from confusability. An alternative explanation is that participants encode an average value of all the stimuli currently seen, and use this stored stimulus as a referent to then compare all subsequent stimuli against (Capehart, Tempone, & Hébert, 1969; Thomas, 1993). Participants may then form a rule such as “respond left when the circles are slightly smaller than the average”. Again, since the training stimuli only deviate slightly from the average, participants are most accurate in categorization on test stimuli that are very similar to the training stimuli but more dissimilar to the opposite category (i.e. the Near1 test stimulus). Whatever the exact specification of the similarity rule, the results of the sequential analysis can be explained by participants switching between two rules – one based on similarity and the other based on relations.
Alternatively a peak shift may represent relational rule use that is affected by stimulus similarity. Participants may know, for example, that artworks in the left category had smaller circles, but at the extreme ends of the dimension where the circles become very small, participants may be uncertain of the applicability of their rule due to the stimuli being dissimilar to the training stimuli. Although possible, this explanation seems unlikely due to the reasonable limits on the range of each dimension tested (see Figure 4.5). There is no obvious reason why participants would regard their rule as applicable to the Near1 stimuli but not to the Near3 and Far stimuli. Still, the evidence showing that rule use is affected by similarity between test and training stimuli (e.g. Allen & Brooks, 1991; Hahn et al., 2010) suggests that further exploration of the nature of rules that participants derive and their willingness to extrapolate that rule on extreme test stimuli is needed. Whether similarity-based generalization in humans is accomplished via learning an explicit similarity rule, or through non-rule-like associative learning processes is difficult to determine empirically since both accounts make essentially the same predictions. In any case, similarity-based generalization certainly does not require an explanation in terms of rules, and it is clear that the content of learning differs to that of generalization based on a relational rule.

4.5.3 Interaction between Rules and Similarity

Livesey and McLaren (2009) showed that the emergence of a relational rule during test resulted in a peak-shifted gradient being replaced by a monotonic one, leading them to suggest that the operation of rule learning and other executive processes meant that evidence of associative processes would be hard to find. The current study however, shows that the presence of a relational rule does not entail that
evidence of associative learning will be overridden. Rather, the results from the sequential analysis shows that under certain conditions, relational rules can only be fully expressed on trials where participants are confident that it is applicable. This is also consistent with Natal, McLaren, and Livesey (2013), who argued that the content of associative learning enters into category judgements along with rules and is subject to cognitive control. In their study, they found different levels of accuracy and patterns of generalization based on whether participants could verbalize a relational category rule. However, they found that both rule-learners and feature-learners were able to reverse the category assignments when asked to, suggesting that associative learning in this context does not automatically bias actions and instead is subject to cognitive control. Similarly, the fact that peak shift was found in Experiment 2 once rule use was disrupted suggests that knowledge about the physical features of the stimuli and a relational rule are stored concurrently, with associative learning expressed only under conditions where participants are not willing or able to use a rule. Similar to Livesey and McLaren (2009) and Natal, McLaren, and Livesey (2013), the current results indicate that higher-order rule learning and learning about physical features of stimuli are integrated into category judgements and that there is a certain degree of cognitive control over how participants generalize. The results are consistent with the idea that associative learning processes operate alongside rule learning, but may not always be expressed due to rule-based generalization having priority on test (McLaren et al., 2014). The primacy of rules is supported by literature indicating that participants prefer to use a single dimension when confronted with categorization tasks involving stimuli with multiple dimensions (Ahn & Medin, 1992; Regehr & Brooks, 1995), and the success of models which assume that participants
initially search for simple rules and then memorize exceptions (Nosofsky, Palmeri, & McKinley, 1994).

These findings are also relevant to theories of categorization that propose two separate, qualitatively different learning processes (e.g. Ashby et al., 1998; Smith, Patalano, & Jonides, 1998), as well as hybrid models of categorization. Hybrid models assume that rule- and similarity-based processes are qualitatively similar (e.g. Erickson & Kruschke, 1998; Pothos, 2005, Verguts & Fias, 2009), but vary based on the number of dimensions they consider (Pothos, 2005), or the abstractness of the feature that is considered (Verguts & Fias, 2009) or attention to rules or exemplars (Erickson & Kruschke, 1998). In particular, models that assume competitive processes with only a single process ‘winning’ and determining behavior would need to incorporate an assessment of rule applicability to explain these results. With both dual-process and hybrid models of categorization, the utility of a process is assumed to determine its use (Hahn et al., 2010). Thus these models could conclude that the test manipulation reduced the utility of using a rule, and therefore its ability to influence responding decreased. This explanation assumes awareness of the reduction in utility of a rule on test, in the absence of feedback. Further studies investigating factors that cause participants to switch between using relational rules and similarity may further inform the development of models that assume competitive processes, whether qualitatively distinct or similar.

Although the sequential analysis demonstrates that participants switch between monotonic and peak-shifted patterns of generalization, the experiments in this chapter cannot reveal whether the switch was due to a conscious decision to use or refrain from using a rule, or alternatively whether the conditions at test ‘activated’ a representation of the rule which determined its use. Participants’ certainty or
confidence in their rule may have been directly affected as a result of its applicability on the previous trial, and therefore participants may have made a conscious decision whether or not to use the rule based on this changing degree of uncertainty. Alternatively, the changes in rule applicability on a trial-to-trial basis may have strengthened or weakened ‘activation’ of the rule in a similar manner to activation of outcomes via associative strength, and lead to participants being more or less willing to use their rule. Note that this latter explanation implies that there is much less cognitive control over rule use than the former explanation.

4.5.4 Typicality Ratings

Interestingly, typicality gradients were steeper for the Consistent Group than the Inconsistent group in Experiment 2 (Figure 4.12b). The flatter typicality gradient in the attended dimension for the Inconsistent group can be explained if we assume that participants are attending less to their attended dimension than group Consistent. Disrupting rule validity on test may have led participants in group Inconsistent to search for additional information other than their attended dimension to aid their judgements. Since at test it is now obvious that two dimensions are varying, participants may have started focusing on the unattended dimension as well. As these values did not change between stimuli, typicality ratings incorporating the unattended dimension would thus be flatter. However, since the same pattern of results was also found for the stimuli that actually did vary along the unattended dimension, a more likely explanation is that participants may have started to attend to the irrelevant features of the stimuli (e.g. the location of circles), which would also have had the same effect. In fact, any attention directed away from the attended dimension when judging typicality would have made the typicality gradient flatter.
Further, since the group difference did not interact with experiment, these results cannot be explained using differences in the ‘attended dimension’ stimuli presented to the Consistent and Inconsistent groups on test, since these were equated in Experiment 2B. It would seem that when a rule is rendered less valid on test in the Inconsistent group, participants overgeneralize perceptions of typicality to non-typical category members in comparison to the Consistent group, effectively considering dissimilar stimuli as being typical of the category. This may reflect uncertainty in the boundaries of the category as a direct result of the lack of applicability of a rule. This is an intriguing result as it suggests that undermining the applicability of a relational rule has effects not just on categorization accuracy but also when judging perceptions of similarity of novel category members. Whether this was a direct result of disrupting rule use or the result of changes in attention is not determinable in these experiments, and is a potential avenue for future research.

Another interesting result regarding the typicality ratings for test stimuli varying the unattended dimension was that in Experiment 2, descending typicality ratings were found in both groups despite flat generalization gradients for categorization. In other words, despite a lack of stimulus control by the unattended dimension in categorization judgements, it seems that participants in both groups were sensitive to variations in the unattended dimension in their typicality judgements. This might suggest that participants did actually learn about the physical features of the unattended dimension despite the hint directing them to attend to the other dimension. The dissociation in categorization and typicality may reflect the fact that participants were selectively focusing on their attended dimension in order to use a rule on this dimension when making categorization judgements (which would have
resulted in a flat generalization gradient), but considering both dimensions when rating the typicality of the stimuli.

4.5.5 Conclusion

In conclusion, the results in this chapter show that in simple category learning, the presence of a relational rule does not necessarily entail that a monotonic gradient of generalization will result. Rather, whether the test conditions make that rule applicable determines whether participants show a monotonic or a peak-shifted gradient of generalization. When rule application on test was consistent, participants generalized using a relational rule, but when rule application on test was rendered inconsistent, participants may have been uncertain of their rule and reverted to generalizing on the basis of similarity. A within-subjects demonstration of participants switching between rule- and similarity-based generalization as a function of rule applicability on the previous trial was provided, suggesting that both kinds of learning are stored concurrently, and their respective influences on behavior can change dynamically throughout test.
Chapter 5: General Discussion

The aim of this thesis was to investigate the differences in learning that occur as a result of changing the verbal instructions to a range of putatively implicit and/or associative learning tasks. This involved primarily comparing learning under incidental (uninstructed) conditions with conditions where participants had the intention to learn about specific stimulus attributes. This chapter will summarize the results in each paradigm, address the research questions posed in the General Introduction, and discuss the implications of these results for theories of learning.

5.1 Summary of Results

5.1.1 Sequence Learning

Sequence learning, as investigated in the serial reaction time (SRT) task, is considered to be one of the best examples of implicit learning since participants make speeded responses to targets and do not usually suspect the presence of contingencies in the underlying sequence of responses (Cleeremans & Jimenez, 1998; Destrebecqz & Cleeremans, 2003). Chapter 2 looked at the effects of providing a written hint to encourage discovery of a relational rule describing a set of probabilistic contingencies in a SRT task. Participants in both groups were asked to respond as quickly and accurately as possible to a target that could appear in one of three locations (left, top, or right) on screen, by pressing the corresponding arrow key. The target was constrained such that it could not appear in the same location on two successive trials, meaning that the target always moved in a clockwise or anti-clockwise direction on each trial. Unbeknownst to the No Hint group, the target locations followed a probabilistic sequence such that 75% of the time, the target would move in a particular (cued, randomly chosen to be clockwise or anti-clockwise for each
participant) direction, and the remaining 25% of the time, the target would move in the opposite (miscued) direction. The Hint group were given written instructions telling them that most of the time, the target would be moving in one direction, and that they should try and figure out which direction this was, while the No Hint group performed the task in the absence of this information. The dependent measure of interest indexing sequence learning was the cueing effect (difference in RT on cued and miscued trials), but explicit knowledge tests were also added after the SRT task to ensure that the hint was having the desired effect of promoting rule learning and increasing explicit sequence knowledge.

Experiments 2.1 (Chapter 2, Experiment 1) and 2.2 sought to establish the pattern of sequential effects that could result in the novel three-choice SRT task devised in Chapter 2. Sequential effects are differences in performance that eventuate due to the recent history of trials, and occur regardless of whether contingencies (i.e. an actual sequence) are present. It was important to investigate sequential effects because they are well established in other two-choice RT tasks (e.g. Bertelson, 1961; Remington, 1969; Soetans, Boer, & Hueting, 1985), and sequential effects have been noted to contaminate measures of sequence learning (Jones, Curran, Mozer, & Wilder, 2013). Across Experiments 2.1 and 2.2, it was shown that a reliable pattern of sequential effects was found in both the reaction times (RT) and error data. The same overall pattern was displayed in the absence of contingencies in both Experiments 2.1 and 2.2, and on cued and miscued trials following experience with the probabilistic contingencies in Experiment 2.2. While reliable interactions were found when dividing the subsequences according to the directions of the third- and fourth-order transitions, the largest sequential effects were found at the second-order level, specifically comparing subsequences consisting of a repetition of target direction.
(XYZ subsequences) to subsequences consisting of an alternation of target direction (ZZY subsequences). Participants were both slower and less accurate to respond on ZYY trials in both Experiments 2.1 and 2.2, and this difference was present on both cued and miscued trials in Experiment 2.2 after the contingencies were removed, suggesting that sequential effects and sequence learning are additive in this context. Due to the potential for this sequential effect to inflate the magnitude of the overall cueing effect due to the relative imbalance of XYZ and ZYY subsequences in cued and miscued trials (see Table 2.2), Experiments 2.3-2.6 examined sequence learning on XYZ and ZYY trials separately to control for this potential confound. This separation of the analyses by subsequence turned out to be a critical for the group differences found in the subsequent experiments.

Experiments 2.3-2.5 compared cueing effects between a Hint group, who were given a written hint encouraging them to work out the direction that the target would be traveling in “most of the time”, and a No Hint group, who performed the task as usual under incidental learning conditions. Experiments 2.3 and 2.4 found selective effects of the hint in increasing the size of the cueing effect, with the hint improving performance for XYZ subsequences, but not for ZYY subsequences, while Experiment 2.5 found a general advantage for the Hint group for both XYZ and ZYY subsequences. The results of Experiment 2.6 suggested that the difference in the effect of the hint between Experiments 2.3-2.4 and Experiment 2.5 was due to a difference in the way the hint was delivered to participants in Experiment 2.5. Experiment 2.6 compared a ‘weak’ and ‘strong’ version of the hint: one where participants were given the written hint to read (as was the case in Experiments 2.3 and 2.4), and one where the experimenter read out the hint to the participant and emphasized that there was a correct answer that would either be clockwise or
anticlockwise. Experiment 2.6 found that this stronger version of the hint selectively increased the cueing effect for ZYZ subsequences, whereas cuing for XYZ subsequences was equivalent for the weak and strong forms of the hint. This finding from Experiment 2.6 suggested that these subtle differences in the delivery of the hint in Experiment 2.5 compared to Experiments 2.3-2.4 allowed participants in the Hint group to express their knowledge on both XYZ and ZYZ subsequences.

Experiments 2.4 and 2.5 revealed significantly better performance on the explicit knowledge tests by the Hint group suggesting that the hint was effective in increasing the amount of explicit sequence knowledge. Above-chance performance was also found for the No Hint group in Experiment 2.5, but not in Experiments 2.3-2.4, suggesting that the No Hint groups had some degree of explicit sequence knowledge that was difficult to detect. Despite the inconsistent performance of the No Hint groups in the explicit knowledge tests within these experiments, significant cueing effects for both subsequences in RT performance were found across Experiments 2.3-2.5. Experiments 2.4-2.5 also implemented a transfer phase at the end of the training phase to assess whether the hint afforded better transfer in situations where the contingencies were removed but explicit sequence knowledge was still applicable. The transfer phase in Experiment 2.4 where the target locations were changed to the left, bottom, and right of the screen, eliminated cueing effects in both groups. A different transfer phase in Experiment 2.5 showed that cueing effects persisted when participants switched hands to respond but this did not differ between groups or subsequences. These results suggest that the hint did not benefit transfer in SRT performance once features of the task were changed, and that retention of the perceptual features of the task, but not the motor component (i.e. the hand used to respond) was important for exhibiting transfer of sequence knowledge.
The results in Chapter 2 suggest that sequence learning is affected by abstract, explicit knowledge, contrary to previous studies showing a lack of benefit of explicit sequence knowledge with probabilistic contingencies (Jiménez, Méndez, & Cleeremans, 1996; Stefaniak et al., 2008). At the same time, the benefit of explicit learning in SRT tasks was found to be highly dependent on the exact instructions given, as well as the properties of the subsequences to be learned. It is clear that simply drawing conclusions about the cognitive penetrability of implicit learning from the benefit or lack of benefit of explicit knowledge is too simplistic, since within the same participants, Experiments 2.3 and 2.4 found that it is possible for explicit knowledge to benefit learning on some subsequences and not others. Further, the hint had a limited benefit in regards to transfer performance in Experiments 2.4 and 2.5, and large cueing effects were found in the absence of the hint, suggesting that the role of explicit knowledge in sequence learning (and perhaps motor learning in general) may be limited (see Sanchez & Reber, 2013, for a similar conclusion).

5.1.2 The Prototype Distortion Task

Chapter 3 examined another paradigm where learning has been claimed to be implicit – the prototype distortion task. In this task, participants are usually exposed to a set of visually similar stimuli constructed around a category (A) ‘prototype’ (the average or ideal exemplar), and then tested on their ability to categorize novel stimuli into that same category (make A/not-A judgements), as well as their ability to recognize the exemplars that they were shown (make old/new judgements). Learning in this task has traditionally been considered to be implicit because of the intact categorization performance of amnesic patients who exhibit impaired memory of seen exemplars (e.g. Knowlton & Squire, 1993).
The experiments in Chapter 3 tested an often-assumed property of learning in the prototype distortion task – its emergence under incidental conditions. Much of the prototype distortion literature uses an exposure phase where participants are not told about the existence of a category or the nature of subsequent tasks, but require participants to perform an alternative task where they may learn incidentally (e.g. pointing to the center dot, Knowlton & Squire; thinking about the appearance of the stimuli, Bozoki, Grossman, & Smith, 2006). Chapter 3 employed a more cognitively-engaging visual search task as a means of incidental exposure to test whether learning about prototype-centered stimuli could occur automatically, and whether any differences would arise between this Search group and a group who were told to memorize a set of prototype-centered stimuli for a subsequent memory test (Memorize group).

Two novel sets of stimuli were constructed for these experiments – one involving 10 lines and the other involving 10 circles, and their respective features were distorted in systematic ways to create the category exemplars. The dependent measure of interest was familiarity ratings for old (previously seen) and new stimuli at matched levels of distortion. A descending prototypicality gradient (highest ratings for low distortions of the prototype and lowest ratings for high distortions of the prototype) was taken to be indicative of learning the similarity structure of the category, and higher familiarity ratings for old than new stimuli were taken to be indicative of recognition ability.

Experiment 3.1 sought to establish whether prototypicality gradients could occur in the absence of exposure, in order to control for this learning-at-test effect in subsequent experiments. This was important to confirm whether any incidental learning that resulted was actually due to the exposure phase and not to participants
explicitly learning about the category during the test phase. Using a mock-subliminal procedure (Palmeri & Flanery, 1999), whereby participants were misled into believing that they were presented with subliminal stimuli, Experiment 3.1 found ‘false’ prototypicality gradients despite no exposure to the stimuli prior to the test phase, and no differences between recognition and categorization tests. Steeper prototypicality gradients and higher ratings for old stimuli were found in both categorization and familiarity tests in a separate group of participants who were exposed to the stimuli. Thus, the stimuli created for these experiments appeared to be subject to the same learning-at-test effects that have been demonstrated using other sets of stimuli (Bozoki, Grossman, & Smith, 2006; Palmeri & Flanery, 1999). The size of the false prototypicality gradient after mock-subliminal exposure was then used as a point of comparison to determine whether incidental learning about the category occurred in the Search groups in subsequent experiments.

Experiments 3.2 and 3.3 used a familiarity test to assess both recognition and the prototypicality gradient. The choice to use a single measure was motivated by the suggestion that parameter differences between categorization and recognition tasks might be responsible for dissociations between amnesic patients and healthy controls despite both indexing the same memory trace (Nosofsky & Zaki, 1998; Zaki et al., 2003). Using the same familiarity test to assess both category learning (through the prototypicality gradient) and recognition (through comparison of ratings for old and new test stimuli) should minimize these parameter differences. Experiments 2 and 3 compared learning between two groups: a Memorize group who were required to memorize the stimuli for a subsequent memory test (which they were aware of), and a Search group who were required to search through the stimuli for an ‘odd one out’ and respond according to the identity of that target (whether the line width was
thicker or thinner, similar to contextual cueing, see Chun & Jiang, 1998). The visual search task was assumed to engage participants’ cognitive functions more extensively than previous incidental learning conditions in the prototype distortion task, where there was no way to assess whether participants were actively performing their given task (e.g. pointing to the center dot, Knowlton & Squire, 1993). Indeed, the high level of accuracy in the visual search task suggested that participants were sufficiently engaged in their task during the exposure phase.

Experiments 3.2 and 3.3 were identical except that the number of trials in the exposure phase was doubled in Experiment 3.3. In both experiments, explicit memorization of the stimuli led to a steeper prototypicality gradient, but did not result in a clear improvement of recognition. Surprisingly, the Search groups were found to show above-chance levels of discrimination between old and new exemplars, suggesting that they had learned incidentally about the stimuli during exposure, but their prototypicality gradient was found to be no greater than the false prototypicality gradient exhibited after mock-subliminal exposure in Experiment 3.1. Experiment 3.3 added another Search group where the stimulus disappeared after a response was made (Search-Terminate group), to test whether any incidental learning that occurred in the Search groups could be attributed to explicit learning that occurred in the residual exposure time after the target was found and a response was made. No differences were found between the Search and Search-Terminate groups in Experiment 3.3, suggesting that any incidental learning that occurred in the Search groups did in fact result from visual search.

The experiments in Chapter 3 found no support for the idea that learning about a prototype-centered category of stimuli could occur incidentally, since the prototypicality gradient in the Search groups was found to be no different to the
magnitude of the false prototypicality gradient exhibited in the absence of exposure. Mysteriously though, some learning must have occurred during visual search since the Search groups were able to recognize exemplars they had seen before. A tentative explanation was offered alluding to the possibility of differences in the stimulus features encoded as a result of explicit memorization or visual search (encoding of the configural or specific features of the stimuli respectively). In any case, it was clear that instructions inducing an intention to learn about the stimuli produced a stronger prototypicality gradient and therefore suggests that learning about prototype-centered categories of stimuli is enhanced by intentional memorization. This was the case even though the stimuli were complex, composed of multiple features and difficult to describe verbally or by using a simple rule. Thus, learning in the prototype distortion task is not implicit in the sense of resulting automatically from incidental exposure, and studies using the prototype distortion task should consider the nature of the encoding conditions carefully and account for learning-at-test effects by using a mock-subliminal control group (Palmeri & Flanery, 1999).

5.1.3 Post-Discrimination Generalization

Chapter 4 examined differences in the pattern of generalization as a function of instructed attention and rule applicability at test. The stimuli used in Chapter 4 were similar to the circle stimuli used in Chapter 3 and comprised of 10 circles of varying sizes, colors and locations on a black square background. Participants were required to solve a categorization task where the stimuli differed on two dimensions (color and size of each circle) that were perfectly correlated during training (e.g. category 1 had greener and smaller circles and category 2 had bluer and larger circles). There were two between-subjects variables of interest. The first was selective
attention to either the color or size of the circles. The attended and unattended dimension was determined post-hoc in Experiment 4.1 but was explicitly manipulated through instructions in Experiment 4.2. This manipulation produced large differences in the pattern of generalization on test in both experiments, with monotonic gradients of generalization resulting when the attended dimension was varied, but flat generalization gradients when the unattended dimension was varied. These generalization gradients indicated that participants formed a relational rule on the category dimension on which they were attending during training, no stimulus control was acquired by the unattended dimension.

However, a separate manipulation of rule applicability on test influenced the results in a way that was of greater theoretical interest. This manipulation involved presenting stimuli on test where a relational rule derived on the attended dimension could only be used on half of the test trials. This was achieved by introducing stimuli with dimension (color or size) values set at the midpoint of the two categories, such that attention to, and use of a relational rule on that dimension was not helpful for accurate categorization. Therefore the primary comparison in Chapter 4 was between a group where a rule could be used consistently (Consistent group), and a group where a rule could only be used inconsistently at test (Inconsistent group).

In Experiment 4.2, reducing the applicability of a rule at test significantly impaired overall accuracy in categorization and changed the pattern of generalization. The Consistent group produced a monotonic gradient of generalization, consistent with rule use, and the Inconsistent group produced a peak-shifted gradient, consistent with similarity-based generalization and analogous discrimination studies in animals (e.g. Hanson, 1957). Evidence of rule use in the Consistent group was provided not just by the generalization gradient in category judgements on test but also by the
typicality ratings. The Consistent group showed the best level of categorization accuracy for stimuli at the extreme ends of the dimension, while simultaneously acknowledging those same stimuli to be least typical of the category. Interestingly, the manipulation of rule application on test not only produced a peak-shifted generalization gradient but also a flatter typicality gradient, perhaps indicating some uncertainty of the category boundaries as a result of increased rule uncertainty or attention to other stimulus features that were irrelevant for accurate categorization.

None of these group differences were found in Experiment 4.1, although the pattern of results appeared to be quite similar between Experiments 4.1 and 4.2. The lack of significant group differences in Experiment 4.1 suggests that the effect of manipulating rule applicability at test was dependent on verbal instructions directing participants to attend to a particular category dimension. The instructions presumably facilitated formation of a relational rule on the attended dimension as participants would have been attending to that dimension from the beginning of training. This may have allowed participants in Experiment 4.2 to discover a rule sooner than participants in Experiment 4.1, increasing their confidence in that rule, and perhaps allowing for a greater effect of subsequently disrupting application of that rule on test.

Further, the results of a post-hoc sequential analysis showed in the Inconsistent group, a peak shift on trials where their primary rule could not be used on the previous trial, but a monotonic gradient on trials where their rule could be used on the previous trial. This within-participants demonstration of both peak shift and a monotonic gradient was interpreted as participants learning about a relational rule as well as the physical features of the stimuli, with uncertainty of rule application on the previous trial dictating the pattern of generalization that was expressed.
5.2 What are the Effects of Instructions on Learning?

The empirical questions posed in the General Introduction will now be addressed in turn, with the associated theoretical implications discussed for each. Verbal instructions were used within this thesis to induce an intention to learn, provide explicit knowledge about the content that participants were to learn about, or to direct attention towards a specific feature of the stimuli or task. Where possible, attempts were made to control for influences on behavior that were unrelated to the phenomena of interest (sequential effects in Chapter 2 and learning-at-test in Chapter 3). The instructions used in this thesis were varied and could have had multiple effects on the participants. These effects could include increasing motivation to learn, enhancing attention in a general or selective way, or engaging a separate learning mechanism that otherwise would not have been engaged. While it is difficult to specify exactly how the instructions influenced each participant in each paradigm, some inferences can be made based on the manner in which instructions induced changes in learning.

5.2.1 General Effects on Motivation and Attention

The simplest explanation of the results in this thesis is that instructions increase motivation to learn by informing the individual that the task contains regularities and/or contingencies. This increase in motivation to learn may have simply increased overall attention to the task, resulting in general improvements in learning. Throughout each chapter, there were results consistent with this interpretation. In the SRT task, participants in the Hint group showed a larger cueing effect than the No Hint group, and also performed at a higher level of accuracy on the explicit knowledge tests. Participants who memorized the stimuli in the prototype
distortion task showed steeper prototypicality gradients, suggesting that they had learned more about the similarity structure of the category, and participants in Experiments 4.1 and 4.2 were more accurate at categorizing stimuli that varied their attended dimension than the unattended dimension.

However, the benefit of the instructions also showed selectively in its effects. In Chapter 2, the hint produced larger cueing effects for one type of subsequence (XYZ) but not another (XYZ) when given in its subtle form. In Chapter 3, memorization of the stimuli in the prototype distortion task produced a steeper typicality gradient, but did not reliably improve recognition of seen exemplars. In Chapter 4, categorization accuracy was better for variations on the attended than the unattended dimension but this was modulated by the degree of similarity between the test and training stimuli. Thus, while the instructions in each chapter may have increased motivation to learn and enhanced overall attention to the task, this alone is not sufficient to explain all the results. The findings might be better explained in terms of changes in selective attention, or in terms of different learning mechanisms operating in each condition.

5.2.2 The Role of Selective Attention

In Experiments 2.3 and 2.4, a selective benefit of the hint was found on XYZ subsequences (trials in which the target direction moved in a consistent direction on two consecutive trials, e.g. clockwise – clockwise), while there was no evidence of a benefit for ZYZ subsequences (trials in which the target direction alternated, e.g. clockwise – anticlockwise). One explanation for this result concerns the relative salience of the two subsequences. XYZ subsequences may have been particularly salient to the Hint group if they were searching for the predominant direction. Trials
on which the target direction repeated may have been particularly salient in either reinforcing or disconfirming their hypotheses concerning the cued direction. This selective benefit for the most salient subsequences fits well with a capacity-limited explicit learning mechanism that searches for useful information and tests hypotheses.

However, it is possible that participants in the No Hint group also learned explicitly, and the effect of the hint simply increased attention (and therefore learning) of XYZ subsequences. The general improvement in cueing for both XYZ and ZYZ subsequences for the Hint group in Experiment 2.5 might be explained by the increased clarity of the strong version of the hint allowing participants to more easily attend to ZYZ subsequences. Perhaps in Experiments 2.3 and 2.4, the weak version of the hint was ambiguous and due to the probabilistic nature of the contingencies, participants entertained more complex rules or thought there was a possibility that the cued direction would switch, leading them to closely attend to the direction of the target for the whole task and in particular, to trials on which the target direction moved consistently. In contrast, clarifying that there was a correct answer that was either clockwise or anticlockwise simplified the task by effectively reducing the task to deciding on the cued direction. Participants in the Strong Hint group may have figured out the cued direction earlier than participants in the Weak Hint group, meaning that participants were not actively searching for the cued direction throughout the whole task and thus XYZ trials were no longer particularly salient or relevant to their hypotheses.

The group differences in the prototype distortion task can also be explained using selective attention. The Search groups may have learned incidentally as a result of visual search, but through encoding the individual features of the stimuli serially during their task. In contrast, the instructions given to the Memorize groups
specifically asked them to study the stimuli as a whole (see Appendix C), perhaps leading them to encode the configural features of the stimuli. Encoding the specific features of the stimuli may have allowed the Search groups to recognize seen exemplars and show a level of recognition equivalent to the Memorize groups because they were employing a similar search strategy at test to find individual features that they recognized. However, this strategy may have meant poorer ability to detect similarity between exemplars in the Search groups (because they have only encoded the specific features of the stimuli and the exemplars were created by distorting each feature of the stimuli independently), leading to an impaired ability to generalize to novel stimuli on test. Thus, the Search groups’ focus on specific stimulus features may have been the reason why the prototypicality gradients were similar in magnitude to that observed after mock-subliminal exposure. The speculative conclusion was made that previous prototype distortion studies employing incidental tasks may have actually encouraged configural encoding of the stimuli, and this was responsible for any learning that eventuated and not incidental learning per se.

It is easier to infer the role of selective attention in the experiments in Chapter 4, as attention was explicitly manipulated in Experiment 4.2, and participants were asked at the end of the experiment to report the phase in which they noticed differences in each dimension. This allowed for exclusion of participants who did not attend to the instructed dimension during training in Experiment 4.2. Across both Experiments 4.1 and 4.2, flat generalization gradients were found for test stimuli that varied the unattended dimension, indicating a lack of stimulus control and no learning of the unattended dimension. In the categorization task employed in this chapter, it seemed that attention was necessary for learning about a stimulus dimension. This may be partly due to the difficulty of the training phase and the complexity of the
stimuli ensuring that attention was maintained to a single dimension throughout training. One can easily imagine a situation in which participants master the discrimination using their attended dimension and then look for further information to assist them in their categorization judgements. The differences in the pattern of generalization between Consistent and Inconsistent groups however, are more difficult to explain in terms of selective attention to physical features. This result will be discussed in more detail shortly.

In summary, many of the results in this thesis can be explained by changes in selective attention as a result of the specific instructions given. This highlights the importance of ensuring that task instructions in learning paradigms are not misleading or ambiguous, as they can have large effects on what trials participants attend to (Chapter 2), what features of the stimuli participants attend to (Chapter 3), and what dimensions of the stimuli participants attend to (Chapter 4). The results in this thesis show that it is critical to ensure that instructions in a given learning task direct participants’ attention in a way that is deemed to be appropriate for the experimenter’s aims.

### 5.2.3 Dissociable Learning Processes?

An alternative way to explain the selective nature of the instructional manipulations is that changing learning orientation also changes the relative influence of dissociable learning processes. As mentioned in the General Introduction, if a dual-process framework is adopted, it is impossible to ensure that tasks are process-pure and tap into a single process (Merikle & Reingold, 1991; Reingold & Merikle, 1988). As such, it may be that both implicit and explicit processes operate under incidental

---

29 Indeed, there was evidence that participants learned about the unattended dimension in Experiment 4.2
learning conditions. While the task may not explicitly encourage participants to learn, they may be naturally suspicious about the aim of the experiment from the beginning or may attempt to search for additional information once they have mastered the task sufficiently to respond accurately. Despite this possibility, the assumption was made that the instructions in this thesis should at least encourage the use of explicit processes. This implies that the contribution of explicit processes would be greater in the instructed or intentional conditions than the uninstructed or incidental conditions. It was also assumed that the contribution from explicit processes in the incidental conditions would be minimal. The following discussion considers the plausibility of an account that follows from these assumptions.

Subtraction logic must also be assumed in order to attribute any group differences to the additional influence of explicit processes in the instructed conditions. That is, it must be assumed that implicit or associative processes operate equally in both incidental and intentional conditions. This assumption is more difficult to defend, since selective attention has just been proposed as an explanation of the obtained group differences, and some associative models contain attentional mechanisms that modulate learning (e.g. Mackintosh, 1975; Pearce & Hall, 1980). Yet, this argument has been made by some researchers (e.g. McLaren et al., 2014), and both associative and implicit mechanisms are typically regarded as automatic in their operation. One of the proposed criteria for automatic processes is their invariance under all circumstances (Hasher & Zacks, 1979), and neither associative nor implicit learning mechanisms, as currently described, require intention to initiate.
As such, to make interpretation from a dual-process perspective more tractable, it will be assumed that associative and implicit mechanisms operate equally in all groups\(^\text{30}\).

It has been argued that qualitative differences in learning (i.e. differences in the pattern of learning and not simply the overall amount) are indicative of dissociable learning processes that are differentially engaged by altering learning orientation (Jones & McLaren, 2009). The results in the SRT task in Chapter 2 can be taken to support this idea, especially when considering the remarkable consistency of the cueing effect across all 8 fourth-order subsequences in Experiment 2.2 (see Figure 2.4), in contrast to the selective effect of the hint in Experiments 2.3 and 2.4. An implicit mechanism might be expected to learn about all subsequences since in the absence of the hint, all target locations and subsequences are more or less equally salient.\(^\text{31}\) The results contained in Chapter 2 are similar to those found by Jones and McLaren (2009), who found that changing learning orientation from incidental to intentional resulted in participants switching from learning about subsequences that ended in an alternation in target location (i.e. XY, YX), to learning about subsequences that end in a repetition of target location (i.e. XX, YY). While their task was a 2-choice SRT task where repetitions and alternations referred to target location and not direction, the pattern of sequential effects found in Experiments 2.1 and 2.2 were reminiscent of the repetition and alternation effects found with target location in 2-choice SRT tasks. Both sets of results may be explained by intentional learning conditions engaging explicit learning mechanisms, which selectively favor the most salient subsequences (those consisting of repetitions of event statistics).

---

\(^{30}\) Note that if humans do learn using two mechanisms, but explicit mechanisms have primacy in influencing behavior, then it is empirically impossible to verify whether implicit mechanisms operate equally across all learning conditions, as any manipulation that affects the operation of the explicit mechanism will also affect behavior.

\(^{31}\) Perhaps making an exception for long runs of repetitions and alternations, which usually show the fastest reaction times in sequential effect experiments (e.g. Bertelson, 1961; Cho et al., 2002).
It is less clear whether the Search and Memorize groups in Chapter 3 were learning in different ways since both the prototypicality effect and recognition were assessed using the same test, and the complex nature of the stimuli meant that the categories could not be learned by forming a simple rule. As discussed above, the results might be better explained in terms of selective attention. It might be argued that learning in the Search groups was implicit because learning occurred incidentally during visual search and learning about complex visual stimuli presumably involves incremental accumulation of information. However, in Chapter 3, an explicit learning strategy would amount to accumulating information in a similar way since the nature of the stimuli meant that encoding each discrete stimulus in declarative memory separately would be difficult. Thus, it is difficult to conclude that different mechanisms were engaged in the Memorize and Search groups, although there was certainly a difference in intentionality. Further studies are needed to confirm whether the observed group differences remain after attention to configural or specific features of the stimuli is equated.

The result in this thesis that provides the strongest evidence of dissociable learning processes concerns the manipulation of rule applicability in generalization in Chapter 4. This was possibly because the nature of the test (a generalization test) provided the clearest means to observe patterns of responding that would either be consistent or inconsistent with a relational rule. While attention was clearly a critical factor in these experiments in determining the overall level and pattern of categorization accuracy on test, differences in generalization on the attended dimension also eventuated as a result of manipulating the degree of rule applicability. Although this manipulation does not concern an instructional manipulation like the other chapters, the group differences were certainly dependent on instructions to
attend to a particular dimension. In any case, this result is important because it seems to demonstrate a situation in which associative and rule-based processes might interact.

In Experiment 4.2, the Consistent group showed a monotonically increasing generalization gradient with the highest level of accuracy for stimuli at the extreme ends of the dimension (e.g. stimuli with the bluest and greenest circles), consistent with applying a relational rule (e.g. “left category has greener circles”). This is consistent with previous studies employing discrimination tasks with simple stimuli differing on a single dimension (e.g. Wills & Mackintosh, 1998). Interestingly, peak shift, a phenomenon arising from post-discrimination generalization in animals (e.g. Hanson, 1959), was found in a group who could only apply their rule on half of their test trials. The explanation proposed was that this reduced their certainty in their rule, leading them to revert to generalizing on the basis of similarity to the physical features of the stimuli, which they must have learned during training alongside their rule. This interpretation was supported by the sequential analysis, where participants expressed both peak-shifted and monotonic gradients of generalization, depending on the applicability of the rule (and presumably their uncertainty or confidence in their rule) on the previous trial.

As discussed, the descending gradients in the typicality ratings as well as the high levels of accuracy in identifying the category difference on the attended dimension are indicative of rule learning in the Inconsistent group. The more contentious question is whether the peak shift in the Inconsistent group is indicative of associative learning processes. Alternative explanations of the peak shift were discussed in Chapter 4 and included failure to extrapolate a relational rule to the ends of the dimension, or a conservative similarity rule (e.g. “It belongs in category 1 if the
circles are green, but not too green”). It is worth stating that if peak shift is the result of an explicit similarity rule held by the Inconsistent group, then it remains to be explained how and why participants learn and maintain two rules during training, and why the relational rule has precedence in responding on test but suffers more than the similarity rule when the reliability of the attended dimension (which they both depend on) is disrupted.

Instead, peak shift in this instance is more easily explained in associative terms. The few demonstrations of peak shift in the human generalization literature have been situations in which rule formation is difficult (e.g. Lewis & Johnston, 1999; Livesey & McLaren, 2009; Wills & Mackintosh, 1998), and are exactly the situations in which it has been claimed associative processes have the best chance to affect behavior (Mackintosh, 1995). Peak shift is well explained by connectionist models employing elemental representation of stimulus dimensions (e.g. Blough, 1975; Ghirlanda & Enquist, 1998), and there are multiple animal studies demonstrating this effect (e.g. Blough, 1973; Terrace, 1968; Hanson, 1957). If an associative explanation of peak shift is accepted, it can be concluded that rather than associative learning automatically influencing behavior and producing unconscious knowledge, associative learning produces conscious information that enters into reasoned judgements and responses. In other words, while the mechanism that produces the knowledge is still different to that of rule or propositional learning, once learnt, that information can then interact with knowledge of different forms. It would seem then, that in the absence of the manipulation of rule applicability, propositional learning processes have a dominating influence on behavior, perhaps due to the fact that the process of learning is itself conscious and therefore provides a stronger justification for behavior.
In summary, there are aspects of the current results that suggest the operation of dissociable learning processes. Changes in post-discrimination generalization gradients suggested use of rules, and advantages afforded by increased knowledge of the contingencies in the SRT task benefitted performance selectively. The results can be interpreted as arising from one process that formulates abstract rules, and another process that learns associations or instances. Even proponents of a single-system view of learning (e.g. Shanks & St. John, 1994) accept that instructions may invoke different ‘strategies’ that learn information that is qualitatively different in content. Importantly, Shanks and St. John claim that the instance-based strategy operates when participants are asked to memorize or observe instances, which presumably are situations in which participants still have the intention to learn. The fact that intentional memorization produces differences in learning when compared to incidental exposure (Chapter 3) suggests that the intentionality of learning has an effect even within instance-based learning strategies. Thus, instructional manipulations seem to be a reliable method of producing learning that is qualitatively different in content, suggesting the existence of multiple learning strategies or processes.

5.3 Incidental Learning

There should be little doubt that the instructions used in this thesis were successful in engaging explicit learning processes, but it may be more contentious whether separate learning mechanisms were employed under incidental conditions. One of the empirical questions addressed in this thesis was the extent to which learning could occur incidentally, in the absence of conscious intention. There is no doubt that humans can learn incidentally. The efficiency in which learning is
accomplished under incidental conditions has led Logan (1988) to claim that incidental learning is closer to intentional learning than no learning at all. The experiments in Chapters 2 and 3 are consistent with the vast amount of experimental evidence showing that the intention to learn is not necessary for learning (e.g. Nissen & Bullemer, 1987; Chun & Jiang, 1998). Large cueing effects were found in the No Hint groups in Chapter 2, and the Search groups in Chapter 3 showed evidence of discriminating between old and new stimuli despite performing a visual search task with complex stimuli. The role of intention in learning is thus not a necessary one, suggesting that there is a degree to which learning can be automatic since it does not require conscious effort or intention to initiate. Learning thus satisfies one of the criteria proposed for an automatic process, (e.g. see Hasher & Zacks, 1979), and suggests that there are situations in which learning occurs as a by-product of processing information (see Jiménez, 2003). Although there are other criteria for automatic processes (Hasher & Zacks, 1979; Shiffrin & Schneider, 1977) that learning may not satisfy, at the very least, it would suggest that not all learning is the result of deliberate, effortful reasoning. This would suggest that single-process theories that postulate that learning is always of this nature (e.g. Mitchell et al., 2009) might need to be revised. In contrast, associative models of learning can easily accommodate the fact that incidental learning occurs since they do not require intention to operate.

However, the possibility cannot be ruled out that some participants in the incidental conditions throughout Chapters 2 and 3 did in fact possess an intention to learn despite the incidental nature of their instructions or task. If this were the case, it would seem that the proportion of participants who did so in Chapter 2 was not large, since the No Hint groups showed lower levels of explicit sequence knowledge than
the Hint groups. The paradigms themselves also help to ensure that learning is incidental. For instance, both the SRT task and visual search task required speeded responses, and are easy to perform, discouraging the need for participants to seek out additional information to aid them in responding. In Chapter 3, the addition of the Search-Terminate group in Experiment 3.3 ensured that participants had no residual exposure time to study the stimulus. These aspects of the results strongly suggest that any learning that occurred in the No Hint groups, and Search groups, was in fact incidental.

5.3.1 Requirements of Incidental Learning

One might question why incidental learning occurred in the No Hint group in Chapter 2 and the Search group in Chapter 3, but not for the unattended category dimension in Chapter 4. In Chapter 4, there was no stimulus control by the unattended dimension in category judgements, despite the color and size dimensions of the stimuli being perfectly correlated during training. This may be because the color and size dimensions were separable, rather than integral dimensions (Garner, 1978), meaning that attention to one dimension does not entail attention to the other dimension. In fact, when attention was directed at a particular dimension, participants would not need to process the unattended dimension at all in order to accurately categorize all stimuli. This stands in contrast to the SRT task, where participants are required to process the location of each target since their responses need to correspond to the target location on each trial (and it is the sequence of target locations that participants learn about).³²

³² But note that processing of the target locations does not entail processing of target direction.
What is intriguing is that the stimulus dimensions of the stimuli used in the prototype distortion task are also separable (e.g. color, line thickness and location in the circle stimuli), and yet incidental learning of these stimuli did occur, as seen by an ability of the Search groups to recognize old stimuli. One way to explain this discrepancy between incidental learning in the prototype distortion task with the lack of incidental learning in the categorization task is through differences in selective attention deployed as a result of visual search. It may be the case that participants’ attention was focused on a single dimension in the categorization task while attention was more diffuse during visual search. Despite the fact that participants used a single dimension (line width) to distinguish the singleton, perhaps the active process of searching for a singleton is particularly conducive to incidental learning of complex stimuli. This certainly accords with the ease with which contextual cueing effects are found (e.g. Chun & Jiang, 1998).

It seems that what is necessary (but perhaps not sufficient, see Weidemann, Satkunarajah, & Lovibond, 2016) for incidental learning to occur is for participants to actively process the relevant information in making their responses. This conclusion is similar to Logan’s (1988) instance theory, which states that attention to a stimulus necessarily results in encoding in memory, as well as Jiménez’s (2003) characterization of implicit learning as the by-product of attention. Thus, to provide the best chances of observing incidental learning, the responses that participants make in a task must rely on attention to the relevant information to be learned. However, incidental learning does seem to be limited to learning of instances or associative information. Across all three chapters, there was no evidence of rule learning under incidental conditions.
5.3.2 Implications for Implicit Learning

The implicit status of learning in two different paradigms was assessed in this thesis. Chapter 2 showed that sequence learning emerged reliably under incidental learning conditions but was not impervious to instructions inducing an intention to learn. Chapter 3 showed that a prototypicality gradient that was significantly greater than a learning-at-test effect did not emerge under incidental exposure conditions (although recognition did) and was also strengthened by verbal instructions to memorize the stimuli. In both paradigms, it seems that verbal instructions and intention to learn can affect the strength of implicit learning effects.

Questions remain about whether learning in the SRT and the prototype distortion tasks are in fact implicit. In fact, it is debatable whether “implicit learning” as a unitary construct is valid or useful at all. As discussed in the General Introduction, there is disagreement over the functional characteristics of implicit learning (Frensch, 1998), how to assess awareness (Shanks & St. John, 1994), and whether implicit and explicit learning processes share information with one another (Lewicki, 1986; Stadler & Frensch, 1994). If implicit learning is conceived as a non-intentional mechanism that accumulates statistical regularities or instances as a result of processing, it would seem that sequence learning in the No Hint groups in Chapter 2, and the ability for the Search groups in Chapter 3 to recognize seen exemplars after visual search, would count as implicit. However, in both paradigms, explicit processes were found to improve learning, and there was some evidence that participants in the No Hint group could explicitly report their knowledge. Thus, the results in Chapters 2 and 3 suggest that the notion of implicit learning being an independent process that is unaffected by explicit processes (e.g. Song, Howard, & Howard, 2007; Willingham, Nissen, & Bullemer, 1987) may need to be revised.
5.4 Implications for the Single-Process View

The current results clearly implicate a role for explicit and propositional processes in learning, but simultaneously show that it is possible for incidental learning to occur. A single-process account might be able to accommodate these seemingly conflicting results if it can explain how an effortful propositional mechanism can come to function so efficiently under incidental conditions, and specify how differences in learning result as a consequence of manipulating the verbal instructions given to participants. One answer that may be proposed is that propositional processes can function under a small degree of effort and conscious intention, but under those conditions is unlikely to produce sensible rules or strong propositional knowledge, resulting in a small amount of learning and knowledge that is fragile and not easily verbalized. Instead, what propositional processes learn under incidental conditions is functionally equivalent to simple associations that are instance or feature-based, and are not abstract or relational.

In other words, if it is assumed that both incidental and intentional conditions engage propositional mechanisms to a different degree, and this produces qualitative changes in the output of learning, then the results in this thesis can be explained. Adopting this view would imply that the conditions under which participants learn are important because they can shift the qualitative properties of the content of learning. Investigating these qualitative shifts is arguably more theoretically interesting than deciding on the number of learning processes we possess (see Cleeremans & Dienes, 2008). Indeed, the distinction between single- and dual-process theories starts to become meaningless in this context, as both theories have the same task of explaining how learning can vary as a result of changing learning orientation. One of the aims of this thesis was to attempt to characterize the nature of interaction between associative
and cognitive processes assuming that they both exist, and as such the single-process view will not be considered further\textsuperscript{33}.

5.5 Implications for the Dual-Process View

In the next section, a tentative conclusion will be drawn from the dual-process perspective about how implicit and explicit, or associative and propositional learning might interact to determine behavior. This is necessarily complex and may be subject to multiple parameters concerning the nature of the task at hand as well as other cognitive processes such as working memory or attention. However, many researchers have acknowledged the theoretical importance of this challenge (e.g. McLaren et al., 2014; Mitchell et al., 2009; Sloman, 1996) and the experiments contained in this thesis hopefully will provide a starting point in this endeavor. For simplicity, the debate will simply be framed in terms of associative and cognitive processes, where the two are treated as qualitatively distinct in their functional properties and output.\textsuperscript{34}

5.5.1 Learning vs. Performance

Firstly, it is important to make a distinction between learning and performance. A dual-process theory should specify not only the conditions under which each learning process will operate, but also how the outputs of such processes interact to produce behavior. This is critical because there are two results within this thesis that demonstrate a situation in which learning may have occurred, but was simply not expressed. The first instance concerns the fact that the Hint group in Chapter 2 did not

\textsuperscript{33} Note that the ensuing discussion is still relevant to single-process theories of learning if it is framed in terms of an associative/instance-based strategy and a rule-based strategy that are both conscious and exist within a single system (e.g. Shanks & St. John, 1994).

\textsuperscript{34} Note that some authors regard associative and cognitive processes as resulting from the same computational hardware (e.g. McLaren et al., 2014; McClelland & Rumelhart, 1985).
show a larger cueing effect for ZYZ subsequences in Experiments 2.3 and 2.4. One might claim that if the aim of the hint was to encourage participants to derive an abstract, relational rule, then that knowledge should have affected all subsequences equally since the rule is applicable to every trial. The fact that it did not might be taken as evidence that the hint was somehow unhelpful or that participants did not derive the desired relational rule.

An alternative explanation discussed in Chapter 2 is that the Hint group used the hint successfully to derive a relational rule, but the expression of their knowledge on ZYZ trials was hindered due to the alternation in target direction making it difficult to respond. In other words, the deficit was in performance, rather than learning. The sequential effects in Experiments 2.1 and 2.2 both support this idea, with slower and less accurate responding on ZYZ trials compared to XYZ trials. It may be that explicit sequence knowledge is difficult to exhibit on trials where responding is difficult. The perfect and near-perfect performance of the Hint group in selecting the cued direction of motion in Experiments 2.4 and 2.5 respectively also supports the fact that the Hint groups did in fact acquire the relevant rule and therefore knowledge of all contingencies.

The second instance concerns the sequential analysis of generalization in the Inconsistent group in Chapter 4. Here, two different patterns of generalization were expressed within the same group of participants, one consistent with rule use and another consistent with peak shift. Participants in the Inconsistent group generalized on the basis of a rule when their rule was applicable on the previous trial, and generalized on the basis of similarity when their rule was not applicable on the previous trial at test. This sequential analysis suggests that participants had learned about a relational rule, and about the physical features of the stimuli during their
categorization training. Since the Consistent group exhibited generalization consistent with rule use throughout the whole test phase, it can be concluded that the presence of a rule usually obscures any evidence of similarity-based generalization if that rule is clearly relevant on test. Therefore, if a dual-process view is adopted, there may be instances in which people learn associatively, but this does not translate into performance if participants also derive a relational rule (see Livesey & McLaren, 2009; McLaren et al., 2014, for similar ideas). This may be due to the conscious nature of cognitive processes and their associated output being more likely to dominate behavior. If the cognitive system usually dominates behavior, it becomes an important task for future research to specify the conditions under which associative learning can be observed.

5.5.2 Factors determining Interaction

What follows, is a speculative account of two such parameters that may determine interaction between associative and cognitive processes. It will be assumed that separate factors govern the relative influence of cognitive and associative processes in learning, and then in determining behavior. This thesis has already suggested one factor that might determine interaction in the relative contribution of each process during learning – whether the individual has the intention to learn. Since cognitive processes are conscious and effortful in their initiation and operation, it should not be controversial to claim that they are more likely to operate (or operate more effectively) under intentional than incidental conditions. Intentional learning conditions can be induced through instructions as in this thesis, or may arise due to intrinsic motivation in the individual to ‘figure out’ the aims of the experiment or search for helpful rules to simplify the task at hand.
The influence of intention (as induced by instructions) can be subsumed into the broader factor of cognitive control. This factor has already been suggested by McLaren et al. (2014) as a critical parameter dictating whether learning is associative or cognitive in nature. Cognitive control might affect the relative contribution of each learning process, as well as the expression of learning. For example, manipulations or tasks which give the participant little opportunity to think and reason during the learning phase may allow associative learning to be the primary mode of learning (see Mackintosh, 1995). Implicit learning tasks that employ complex stimuli that are not easily memorized or speeded responses where the participant does not have time to think carefully may mean that learning is better accomplished incidentally than intentionally memorizing or using abstract rules. Similarly, expression of rules or explicit knowledge on more direct tests of learning may not be possible if the task is demanding or difficult. Thus one reason why some implicit learning tasks seem impervious to explicit knowledge is because the nature of the task (involving speeded responses, complex stimuli or sequences etc.) ensures that the degree of cognitive control is low, meaning that the relevant explicit knowledge cannot be easily applied to the task (e.g. Jiménez, Méndez, & Cleeremans, 1996).

The second factor determining whether or not explicit processes impact learning, or whether or not explicit knowledge impacts performance, is the utility of that process (or knowledge) in a given task. That is, explicit learning, or explicit knowledge, will be used to the extent that it is useful or practical in a task. One of the reasons why tasks such as the SRT task are considered to be good examples of implicit learning is because the task itself (responding to a target) is easy to perform without needing to derive rules or make explicit predictions. The same can be said of

---

35 Although note that McLaren et al. (2014) claim that cognition is simply controlled association, which is something that not all dual-process theorists might agree on.
the visual search task employed in Chapter 3. Under these situations cognitive processes are unlikely to contribute to learning since there is no obvious reason for their initiation.

However, if cognitive processes are initiated, the utility of the resultant knowledge might then determine whether behavior will be affected. Chapter 2 demonstrated that contrary to previous studies demonstrating no effect of explicit knowledge in probabilistic sequence learning (e.g. Jiménez, Méndez, & Cleeremans, 1996), explicit sequence knowledge can benefit sequence learning as long as that knowledge is sufficiently simple to hold in working memory and apply to the task. Chapter 4 showed that while the presence of a relational category rule was easily used on test when that rule continued to be clearly applicable, generalization in line with similarity to shared perceptual features emerged once application of that rule was disrupted. One interpretation of the results in Experiment 4.2 was that the effect of the test manipulation reduced the utility of the rule in the Inconsistent group, leading participants to revert to generalizing on the basis of similarity. This may have been a deliberate switch between two conscious strategies, or an effect of uncertainty resulting in a change to more intuitive category judgements based on the perceptual features of the stimuli. Similarly, it may be the case that once the cognitive system extracts a useful rule, the utility of associative information is reduced and thus has little influence on behavior, since that rule should adequately summarize all the necessary (associative) information and therefore be more useful for the participant.

These two parameters represent highly speculative attempts to integrate the findings from the varied paradigms within this thesis. Due to the difficulty in devising decisive tests to determine whether associative or propositional learning has occurred (and indeed many of the results in this thesis can be interpreted from multiple
perspectives), it is obviously very difficult to begin specifying their manner of interaction. Still, the importance of examining the interaction between processes has been noted from both single- and dual-process advocates (McLaren et al., 2014; Mitchell et al., 2009; Sloman, 1996), and investigating the parameters that influence the interaction may also be informative in describing the operational characteristics of different learning strategies or processes. Regardless of whether a single- or dual-process stance is adopted, the results in this thesis highlight the importance of considering how verbal instructions and learning orientation can change learning and behavior.

5.6 Concluding Remarks

In conclusion, human learning proceeds incidentally but is also altered by verbal instructions inducing an intention to learn. Verbal instructions can change what participants attend to in a task, determine whether participants derive abstract rules or learn about associations, and influence the ability of participants to generalize their knowledge to novel stimuli. The results in this thesis suggest that human learning is highly sensitive to the exact verbal instructions given, and highlight the importance of considering learning orientation in tasks that are presumed to assess implicit or associative learning. Manipulating verbal instructions and intention may provide a reliable means to change the content of learning, allowing for a better understanding of whether we possess separable learning processes and how they might interact.
References


Beesley, T., Vadillo, M. A., Pearson, D., & Shanks, D. R. (2015). Pre-exposure of repeated search configurations facilitates subsequent contextual cuing of...


Hanson, H. M. (1959). Discrimination training effect on stimulus generalization gradient for spectrum stimuli. Science, 125, 888-889.


253


Appendix A

Results from the Training Phase in Experiment 2.2

Table A1.
ANOVA results for RTs in the training phase of Experiment 2.

<table>
<thead>
<tr>
<th></th>
<th>$F$</th>
<th>$p$</th>
<th>$\eta^2_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>cueing</td>
<td>1.71</td>
<td>.212</td>
<td>.109</td>
</tr>
<tr>
<td>fourth</td>
<td>.264</td>
<td>.615</td>
<td>.019</td>
</tr>
<tr>
<td>third</td>
<td>5.03</td>
<td>.042</td>
<td>.264</td>
</tr>
<tr>
<td>second</td>
<td>31.41</td>
<td>&lt;.001</td>
<td>.692</td>
</tr>
<tr>
<td>cueing * fourth</td>
<td>2.19</td>
<td>.161</td>
<td>.136</td>
</tr>
<tr>
<td>cueing * third</td>
<td>.026</td>
<td>.875</td>
<td>.002</td>
</tr>
<tr>
<td>fourth * third</td>
<td>1.42</td>
<td>.254</td>
<td>.092</td>
</tr>
<tr>
<td>cueing * fourth * third</td>
<td>.110</td>
<td>.745</td>
<td>.008</td>
</tr>
<tr>
<td>cueing * second</td>
<td>6.63</td>
<td>.022</td>
<td>.322</td>
</tr>
<tr>
<td>fourth * second</td>
<td>7.29</td>
<td>.017</td>
<td>.342</td>
</tr>
<tr>
<td>cueing * fourth * second</td>
<td>4.68</td>
<td>.048</td>
<td>.251</td>
</tr>
<tr>
<td>third * second</td>
<td>.125</td>
<td>.729</td>
<td>.009</td>
</tr>
<tr>
<td>cueing * third * second</td>
<td>.032</td>
<td>.860</td>
<td>.002</td>
</tr>
<tr>
<td>fourth * third * second</td>
<td>5.10</td>
<td>.040</td>
<td>.267</td>
</tr>
<tr>
<td>cueing * fourth * third * second</td>
<td>.014</td>
<td>.906</td>
<td>.001</td>
</tr>
</tbody>
</table>

Table A2.
ANOVA results for errors in the training phase of Experiment 2.

<table>
<thead>
<tr>
<th></th>
<th>$F$</th>
<th>$p$</th>
<th>$\eta^2_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>cueing</td>
<td>10.56</td>
<td>.006</td>
<td>.430</td>
</tr>
<tr>
<td>fourth</td>
<td>1.97</td>
<td>.183</td>
<td>.123</td>
</tr>
<tr>
<td>third</td>
<td>.031</td>
<td>.863</td>
<td>.002</td>
</tr>
<tr>
<td>second</td>
<td>.049</td>
<td>.828</td>
<td>.004</td>
</tr>
<tr>
<td>cueing * fourth</td>
<td>6.46</td>
<td>.024</td>
<td>.316</td>
</tr>
<tr>
<td>cueing * third</td>
<td>.043</td>
<td>.839</td>
<td>.003</td>
</tr>
<tr>
<td>fourth * third</td>
<td>.001</td>
<td>.974</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>cueing * fourth * third</td>
<td>.024</td>
<td>.879</td>
<td>.002</td>
</tr>
<tr>
<td>cueing * second</td>
<td>4.20</td>
<td>.060</td>
<td>.231</td>
</tr>
<tr>
<td>fourth * second</td>
<td>7.03</td>
<td>.019</td>
<td>.334</td>
</tr>
<tr>
<td>cueing * fourth * second</td>
<td>4.23</td>
<td>.059</td>
<td>.232</td>
</tr>
<tr>
<td>third * second</td>
<td>.304</td>
<td>.590</td>
<td>.021</td>
</tr>
<tr>
<td>cueing * third * second</td>
<td>.165</td>
<td>.690</td>
<td>.012</td>
</tr>
<tr>
<td>fourth * third * second</td>
<td>.225</td>
<td>.643</td>
<td>.016</td>
</tr>
<tr>
<td>cueing * fourth * third * second</td>
<td>.178</td>
<td>.680</td>
<td>.013</td>
</tr>
</tbody>
</table>
Figure A1. RTs (a) and proportion of errors (b) for cued and miscued trials for each fourth-order subsequence in the training phase of Experiment 2. Subsequences are divided according to whether transitions at the second-order (XYZ vs. ZYZ, shown as separate lines), third-order (left vs. right points connected by lines), and fourth-order (left vs. right side of the figures) level were the same or different direction to the first-order transition (YZ). Cued and miscued trials are shown as separate lines.
Appendix B

Additional Analyses for the Transfer Phase in Experiment 2.4

In the Hint group, the cueing effects displayed in the transfer phase were found to be significant in RT for both XYZ and ZYZ subsequences and XYZ errors, smallest $t(38) = 3.52$, $SEM = .005$, $p = .001$, but the ZYZ cueing effect was marginally non-significant in errors, $t(38) = 1.84$, $SEM = .013$, $p = .073$. In the No Hint group, the cueing effect for XYZ subsequences was marginally non-significant in RTs, $t(36) = 1.88$, $SEM = .005$, $p = .068$, significant in XYZ, $t(36) = 3.18$, $SEM = .008$, $p = .003$, significant in ZYZ RT cueing, $t(36) = 4.38$, $SEM = .003$, $p < .001$, and not significant in ZYZ errors, $t < 1$. 
Appendix C

Instructions used in Experiments 3.2 and 3.3

Instructions given to Memorize group:
This experiment will test your VISUAL MEMORY abilities. In this experiment, you will be presented with various visual stimuli. In the first phase, the stimuli will appear on screen for a short time and then disappear. While they are on screen, please pay attention to the stimuli and try and MEMORIZE them. Try not to focus on specific features, but rather memorize the stimulus as a whole. Your memory of the stimuli will be tested later. You will be told when the stimuli have stopped displaying. Press spacebar to begin.

Instructions given to Search group:
This experiment will test your VISUAL SEARCH abilities. In this experiment, you will be presented with different stimuli. The stimuli will appear on screen for a short time and then disappear. The stimuli will either contain 10 circles, or 10 lines. On each trial one of the circles or lines will be different to the rest. In other words, there is an ODD ONE OUT. The odd one out will have a line width that is either THICKER or THINNER than the rest (see example). If you think that the odd one out is THICKER than the rest press A (on the left). If you think the odd one out is THINNER than the rest, press L (on the right). You only need to press A or L once. Try and respond as fast as you can, while still being accurate. If the stimulus disappears and you still haven’t responded, just guess. You will be timed out otherwise and the next trial will begin. If you have any questions, please ask the experimenter now. Otherwise, press spacebar to begin.
Note that there will be no feedback given, so you will not know whether you are correct or not. So just try your best. Remember, you can still respond quickly after the stimulus has disappeared. If you see the words ‘Too slow’, it means that you have been timed out. If you don’t, it means that your response has been recorded. Press the spacebar to begin.
Appendix D

Unfolded Dimension for Experiments 4.1 and 4.2

Figure D1. Categorization accuracy and typicality ratings for the unfolded attended (upper panels) and unattended (lower panels) dimensions in Experiments 4.1 and 4.2.
Appendix E

Additional Statistical Results from Experiment 4.2

Training

The Consistent group performed significantly better overall in training, $F(1,179) = 5.62, p = .024, \eta^2_p = .028$, and participants in Experiment 4.2B outperformed those in Experiment 4.2A, $F(1,179) = 23.36, p < .001, \eta^2_p = .115$, but there was no main effect of attention group, $F < 1$. No other interactions were significant, largest $F(1,179) = 1.78, p = .183, \eta^2_p = .010$. Experiment 2A and 2B were run at different times but the main effect of group is difficult to explain due to participants in all groups receiving the same training. In any case, there was no evidence that participants in different groups or experiments learned at different rates.

Category Judgements

Attended Dimension. The Attend Size group (figure E1c) had higher categorization accuracy than the Attend Color group (figure 8a), $F(1,179) = 4.47, p = .036, \eta^2_p = .024$. A significant interaction between the linear trend and attention group was also found, $F(1,179) = 4.62, p = .033, \eta^2_p = .025$. Inspection of Figures E1a (Attend Color) and E1c (Attend Size) reveals that this interaction is due to participants in both Consistent and Inconsistent groups showing a monotonic gradient when told to attend to size, but more pronounced group differences for those told to attend to color.

Unattended Dimension. There was a significant 4-way interaction between linear trend, attention group, test group, and experiment, $F(1,179) = 4.80, p = .030,$
\( \eta_p^2 = .026 \), and a significant interaction between the main effect of attention group and experiment, \( F(1,179) = 5.93, p = .016, \eta_p^2 = .032 \).

Typicality Ratings

**Attended Dimension.** The 4-way interaction with linear trend, attention group, test group and experiment was significant, \( F(1,179) = 6.45, p = .012, \eta_p^2 = .035 \). Typicality ratings were higher overall for the Attend Size group (figure E1d) compared to the Attend Color group (figure 8b), \( F(1,179) = 9.76, p = .002 \eta_p^2 = .052 \), and also higher for the Inconsistent group than the Consistent group (figure 6b), \( F(1,179) = 5.64, p = .019, \eta_p^2 = .031 \). There was also a significant interaction between attention group, test group, and experiment, \( F(1,179) = 14.64, p < .001, \eta_p^2 = .076 \).

**Unattended Dimension.** There were steeper gradients in the Attend Color group than the Attend Size group, \( F(1,179) = 6.28, p = .013, \eta_p^2 = .034 \). There was a significant main effect of attention group, with significantly higher typicality ratings overall in the Attend Size group, \( F(1,179) = 5.42, p = .021, \eta_p^2 = .029 \), and a significant interaction between attention group, test group, and experiment, \( F(1,179) = 8.88, p = .003, \eta_p^2 = .047 \).
Figure E1. Categorization accuracy and typicality judgements for variations on the attended dimension split by attention group (a and b: attend color, c and d: attend size) and experiment (e and f: Experiment 4.2A, g and h: Experiment 4.2B). Error bars represent the standard error of the mean.
### Table E1.

**Full ANOVA results for category judgements for variations on the attended dimension.**

<table>
<thead>
<tr>
<th></th>
<th>$F$</th>
<th>$p$</th>
<th>$\eta^2_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>test stim (linear)</td>
<td>24.46</td>
<td>&lt;.001</td>
<td>.120</td>
</tr>
<tr>
<td>test stim (quadratic)</td>
<td>27.29</td>
<td>&lt;.001</td>
<td>.132</td>
</tr>
<tr>
<td>test stim * attention group</td>
<td>4.62</td>
<td>.033</td>
<td>.025</td>
</tr>
<tr>
<td>test stim * test group (linear)</td>
<td>1.03</td>
<td>.312</td>
<td>.006</td>
</tr>
<tr>
<td>test stim * test group (quadratic)</td>
<td>5.00</td>
<td>.027</td>
<td>.027</td>
</tr>
<tr>
<td>test stim * experiment</td>
<td>.348</td>
<td>.556</td>
<td>.002</td>
</tr>
<tr>
<td>test stim * attention group * test group</td>
<td>.466</td>
<td>.496</td>
<td>.003</td>
</tr>
<tr>
<td>test stim * attention group * experiment</td>
<td>1.07</td>
<td>.303</td>
<td>.006</td>
</tr>
<tr>
<td>test stim * test group * experiment</td>
<td>.002</td>
<td>.965</td>
<td>.000</td>
</tr>
<tr>
<td>test stim * attention group * test group * experiment</td>
<td>1.53</td>
<td>.217</td>
<td>.008</td>
</tr>
<tr>
<td>attention group</td>
<td>4.47</td>
<td>.036</td>
<td>.024</td>
</tr>
<tr>
<td>test group</td>
<td>10.59</td>
<td>.001</td>
<td>.056</td>
</tr>
<tr>
<td>experiment</td>
<td>1.02</td>
<td>.315</td>
<td>.006</td>
</tr>
<tr>
<td>attention group * test group</td>
<td>.823</td>
<td>.366</td>
<td>.005</td>
</tr>
<tr>
<td>attention group * experiment</td>
<td>.543</td>
<td>.462</td>
<td>.003</td>
</tr>
<tr>
<td>test group * experiment</td>
<td>1.43</td>
<td>.234</td>
<td>.008</td>
</tr>
<tr>
<td>attention group * test group * experiment</td>
<td>3.23</td>
<td>.074</td>
<td>.018</td>
</tr>
</tbody>
</table>

* Note that the results for test stim refer to the linear trend (unless otherwise specified) and not the main effect.

### Table E2.

**Full ANOVA results for category judgements for variations on the unattended dimension.**

<table>
<thead>
<tr>
<th></th>
<th>$F$</th>
<th>$p$</th>
<th>$\eta^2_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>test stim</td>
<td>.641</td>
<td>.424</td>
<td>.004</td>
</tr>
<tr>
<td>test stim * attention group</td>
<td>.735</td>
<td>.392</td>
<td>.004</td>
</tr>
<tr>
<td>test stim * test group</td>
<td>1.14</td>
<td>.287</td>
<td>.006</td>
</tr>
<tr>
<td>test stim * experiment</td>
<td>.001</td>
<td>.975</td>
<td>.000</td>
</tr>
<tr>
<td>test stim * attention group * test group</td>
<td>.012</td>
<td>.912</td>
<td>.000</td>
</tr>
<tr>
<td>test stim * attention group * experiment</td>
<td>.772</td>
<td>.381</td>
<td>.004</td>
</tr>
<tr>
<td>test stim * test group * experiment</td>
<td>2.25</td>
<td>.136</td>
<td>.012</td>
</tr>
<tr>
<td>test stim * attention group * test group * experiment</td>
<td>4.80</td>
<td>.030</td>
<td>.026</td>
</tr>
<tr>
<td>attention group</td>
<td>.323</td>
<td>.570</td>
<td>.002</td>
</tr>
<tr>
<td>test group</td>
<td>395.3</td>
<td>&lt;.001</td>
<td>.688</td>
</tr>
<tr>
<td>experiment</td>
<td>.653</td>
<td>.420</td>
<td>.004</td>
</tr>
<tr>
<td>attention group * test group</td>
<td>.001</td>
<td>.976</td>
<td>.000</td>
</tr>
<tr>
<td>attention group * experiment</td>
<td>5.93</td>
<td>.016</td>
<td>.032</td>
</tr>
<tr>
<td>test group * experiment</td>
<td>.002</td>
<td>.966</td>
<td>.000</td>
</tr>
<tr>
<td>attention group * test group * experiment</td>
<td>1.39</td>
<td>.240</td>
<td>.008</td>
</tr>
</tbody>
</table>

* Note that the results for test stim refer to the linear trend and not the main effect.
Table E3.
Full ANOVA results for typicality ratings for variations on the attended dimension.

<table>
<thead>
<tr>
<th></th>
<th>$F$</th>
<th>$p$</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>test stim</td>
<td>98.57</td>
<td>&lt;.001</td>
<td>.355</td>
</tr>
<tr>
<td>test stim * attention group</td>
<td>3.14</td>
<td>.078</td>
<td>.017</td>
</tr>
<tr>
<td>test stim * test group</td>
<td>22.08</td>
<td>&lt;.001</td>
<td>.110</td>
</tr>
<tr>
<td>test stim * experiment</td>
<td>.617</td>
<td>.433</td>
<td>.003</td>
</tr>
<tr>
<td>test stim * attention group * test group</td>
<td>.621</td>
<td>.432</td>
<td>.003</td>
</tr>
<tr>
<td>test stim * attention group * experiment</td>
<td>.565</td>
<td>.453</td>
<td>.003</td>
</tr>
<tr>
<td>test stim * test group * experiment</td>
<td>3.07</td>
<td>.082</td>
<td>.017</td>
</tr>
<tr>
<td>test stim * attention group * test group * experiment</td>
<td>6.45</td>
<td>.012</td>
<td>.035</td>
</tr>
<tr>
<td>attention group</td>
<td>9.76</td>
<td>.002</td>
<td>.052</td>
</tr>
<tr>
<td>test group</td>
<td>5.64</td>
<td>.019</td>
<td>.031</td>
</tr>
<tr>
<td>experiment</td>
<td>2.18</td>
<td>.142</td>
<td>.012</td>
</tr>
<tr>
<td>attention group * test group</td>
<td>.082</td>
<td>.775</td>
<td>.000</td>
</tr>
<tr>
<td>attention group * experiment</td>
<td>.216</td>
<td>.642</td>
<td>.001</td>
</tr>
<tr>
<td>test group * experiment</td>
<td>1.01</td>
<td>.317</td>
<td>.006</td>
</tr>
<tr>
<td>attention group * test group * experiment</td>
<td>14.64</td>
<td>&lt;.001</td>
<td>.076</td>
</tr>
</tbody>
</table>

* Note that the results for test stim refer to the linear trend and not the main effect.

Table E4.
Full ANOVA results for typicality ratings for variations on the unattended dimension.

<table>
<thead>
<tr>
<th></th>
<th>$F$</th>
<th>$p$</th>
<th>$\eta^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>test stim</td>
<td>76.57</td>
<td>&lt;.001</td>
<td>.300</td>
</tr>
<tr>
<td>test stim * attention group</td>
<td>6.28</td>
<td>.013</td>
<td>.034</td>
</tr>
<tr>
<td>test stim * test group</td>
<td>9.56</td>
<td>.002</td>
<td>.051</td>
</tr>
<tr>
<td>test stim * experiment</td>
<td>.025</td>
<td>.875</td>
<td>.000</td>
</tr>
<tr>
<td>test stim * attention group * test group</td>
<td>.284</td>
<td>.595</td>
<td>.002</td>
</tr>
<tr>
<td>test stim * attention group * experiment</td>
<td>&lt;.001</td>
<td>.990</td>
<td>.000</td>
</tr>
<tr>
<td>test stim * test group * experiment</td>
<td>.582</td>
<td>.447</td>
<td>.003</td>
</tr>
<tr>
<td>test stim * attention group * test group * experiment</td>
<td>1.61</td>
<td>.206</td>
<td>.009</td>
</tr>
<tr>
<td>attention group</td>
<td>5.42</td>
<td>.021</td>
<td>.029</td>
</tr>
<tr>
<td>test group</td>
<td>.013</td>
<td>.908</td>
<td>.000</td>
</tr>
<tr>
<td>experiment</td>
<td>1.19</td>
<td>.277</td>
<td>.007</td>
</tr>
<tr>
<td>attention group * test group</td>
<td>.037</td>
<td>.848</td>
<td>.000</td>
</tr>
<tr>
<td>attention group * experiment</td>
<td>.133</td>
<td>.716</td>
<td>.001</td>
</tr>
<tr>
<td>test group * experiment</td>
<td>2.91</td>
<td>.090</td>
<td>.016</td>
</tr>
<tr>
<td>attention group * test group * experiment</td>
<td>8.88</td>
<td>.003</td>
<td>.047</td>
</tr>
</tbody>
</table>

* Note that the results for test stim refer to the linear trend and not the main effect.