View metadata, citation and similar papers at core.ac.uk brought to you by CORE

provided by Sydney eScholarship

THE UNIVERSITY OF
SYDNEY

COPYRIGHT AND USE OF THIS THESIS

This thesis must be used in accordance with the provisions of the Copyright Act 1968.

Reproduction of material protected by copyright may be an infringement of copyright and copyright owners may be entitled to take legal action against persons who infringe their copyright.

Section 51 (2) of the Copyright Act permits an authorized officer of a university library or archives to provide a copy (by communication or otherwise) of an unpublished thesis kept in the library or archives, to a person who satisfies the authorized officer that he or she requires the reproduction for the purposes of research or study.

The Copyright Act grants the creator of a work a number of moral rights, specifically the right of attribution, the right against false attribution and the right of integrity.

You may infringe the author's moral rights if you:

- fail to acknowledge the author of this thesis if you quote sections from the work
- attribute this thesis to another author
- subject this thesis to derogatory treatment which may prejudice the author's reputation

For further information contact the University's Director of Copyright Services

sydney.edu.au/copyright

LEARNING IN COMPLEX TASKS: A COMPARISON OF COGNITVE LOAD AND DUAL SPACE THEORIES

Andrew Howe

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

School of Psychology

Faculty of Science

University of Sydney

December 2013

ACKNOWLEDGEMENTS

This thesis would not have been possible without the support of a great number of people to whom I am extremely grateful. To my supervisor, Dr Bruce Burns, thank you for your patience, insight, and guidance. I would not have been able to finish this thesis without your considered input. To my associate supervisor, Dr Damian Birney, thank you for your support, positivity and understanding. Without your open door and kind words, my thesis would have been a far less pleasant experience.

Thank you to Professor Robert Wood, and all the Accelerated Learning Laboratory team for giving me the opportunity to undertake this thesis. Your financial and research support was greatly appreciated. It was a wonderful experience to work with you all.

This research could not have been achieved without the programming skills and generosity of Adam Pryor. Thank you Adam for your kind assistance. I would also like to thanks Yohans Bastian for his capable and unflustered IT assistance throughout this project.

Thank you to my PhD mates (you know who you are)- this would have been a lot less fun without you!

To my family, dad, mum and James. Thank you for the unfaltering love and support you have given me through what I'm sure has seemed an unending process. It has been a great relief to have you behind me every step of the way. Dad, I couldn't have done this without you these past few weeks.

Finally, to my partner Jen, I cannot thank you enough for supporting me though the highs and lows of this thesis. Thank you for your patience, understanding, for always being there for me, and, most of all, for always believing in me.

ABSTRACT

This thesis aimed to compare two theoretical accounts of learning in complex settings: Cognitive Load Theory (CLT) and Dual Space Theory (DST). CLT proposes that learning is fundamentally limited by the information processing capacity of working memory. Tasks that impose higher processing demands on working memory (cognitive load) are argued to produce poorer learning. DST conceives of learning as search through two internal task representations that comprise of either task rules (rule space) or task states (instance space). Tasks that encourage lower rule space search and higher instance space search are argued to produce poorer learning. The theories differ most prominently in their explanations of the *goal free effect*, where specific goals have been found to elicit poorer learning than non-specific goals. CLT explains this effect by suggesting that specific goals elicit higher cognitive load than non-specific goals whilst DST argues that specific goals reduce rule, and increase instance, space search. To reconcile these different explanations, CLT researchers have proposed that the theories are complementary, suggesting that cognitive load determines the extent of rule space search. They suggest that higher cognitive load prevents rule space search whilst lower cognitive load encourages it. However, empirical evidence for this relationship is mixed. This thesis therefore aimed to investigate the relationship between cognitive load and rule space search to determine their independence.

To examine whether rule space search is influenced by cognitive load, three studies were conducted. Each attempted to manipulate rule space search independently of cognitive load. Study 1 (N=63) trained participants to perform a complex skill acquisition task under conditions that either encouraged or discouraged rule space search. Cognitive load was held constant between the conditions. Results indicated that the participants encouraged to search rule space search acquired more knowledge despite equivalently high cognitive load across the conditions. Whilst this suggested cognitive load did not influence rule space search, results may have been confounded by motivational differences between the groups. To remedy these issues, a second study (N=75) was conducted that manipulated both rule space search and cognitive load in a 2 (goal assignment) x 2 (information level) between-subjects design. Manipulations were intended to create conditions where cognitive load and rule space search were both high or both low, in opposition to their proposed interdependency. Results however were mixed. Whilst cognitive load and rule space seemed to vary independently between the groups, they exhibited a negative relationship overall, consistent with their proposed relationship. Study 3 (N=107) addressed these issues by better controlling the influence of task manipulations. Using the same 2 x 2 design as Study 2, results indicated that groups encouraged to search rule space did so independently of any influence of cognitive load. However, results were not entirely consistent with either CLT or DST.

Taken together the findings of this thesis tentatively indicate that cognitive load does not influence rule space search in all situations. Rule space search may be sufficient to account for the goal free effect, but in more complex settings, recourse to cognitive load may be necessary. It is argued that further research should examine whether cognitive load is a necessary variable to propose in explaining all instances of the goal free effect.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Cognitive Load Theory (CLT) and Dual Space Theory (DST) offer competing explanations of the cognitive processes underpinning learning in complex settings. CLT proposes that learning depends fundamentally on the information processing capacity of working memory. Task characteristics that impose excessive processing demands on working memory (i.e. cognitive load) are therefore likely to impair learning. For CLT, minimising cognitive load is the primary means of facilitating learning. DST proposes that learning can be conceived as a search through two internal problem representations called spaces. One space represents the actual states of a task whilst the other, the rules and principles that govern its operation. DST proposes that acquiring comprehensive task knowledge requires learners to search rule space. Encouraging rule space search is therefore the primary means of facilitating learning.

Despite their different explanations, CLT researchers have suggested that CLT and DST are complementary. They propose that the extent to which learners search rule space is dependent on cognitive load. Under this approach, higher cognitive load effectively prevents rule space search and vice versa. This suggestion does not accord with a number of empirical studies that have indirectly suggested that cognitive load and rule space search may vary independently. However, since this has never been tested directly, the relationship between cognitive load and rule space search remains ambiguous.

This thesis aimed to investigate the relationship between cognitive load and rule space search to determine their independence. Three studies report on this aim. To introduce this research, Chapter 1 provides an historical overview of the theories and research that have informed explanations of learning in complex tasks. Chapter 2 outlines CLT and DST and contrasts their accounts of learning in complex settings. The chapter also outlines the empirical basis for expecting cognitive load and rule space search to vary independently. Chapter 3 details the rationale and proposed methods of the empirical studies and proposes the hypotheses that will be tested in the subsequent studies. Chapter 4 describes Study 1. In this study 63 participants were trained to perform a complex skill acquisition task under conditions that either encouraged or discouraged rule space search under conditions of equivalent cognitive load. Chapter 5 presents a revised experimental design based on the findings of Study 1 and describes Study 2. In this study 75 participants were trained in a 2x2 between-subjects design that manipulated goal type and level of information to better examine the independence of cognitive load and rule space search. Chapter 6 presents study 3. This study trained 107 participants under the same design as Study 2 but with minor revisions to more effectively distinguish between the influences of cognitive load and rule space search than in the previous study. Chapter 7 summarises the results of the three studies and discusses their implications. Conclusions and future directions are presented at the end of this chapter.

CHAPTER ONE:

COGNITIVE FOUNDATIONS OF COMPLEX TASK PERFORMANCE

This chapter provides an historical overview of cognitive research on complex task performance. It focuses on the major theoretical developments from the domains of problem solving and expertise, both of which have strongly influenced the two theories that are central to this thesis. The chapter aims, in particular, to elucidate the fundamental cognitive structures and processes identified by the fields of problem solving and expertise that underpin performance in complex tasks.

Problem solving and the relevance of task knowledge.

Investigations of problem solving were perhaps the genesis of research on complex task performance. Their focus was on how one could achieve a desired outcome when not initially knowing how to do so (Duncker, 1945). The earliest researchers in the field were the Gestalt psychologists who emphasised the relevance of mental representations and, in particular, how grouping and reorganising components of a representation could facilitate problem solving (Wertheimer, 1959). For example, Duncker (1945) in his now famous candle problem found that presenting objects in a way that was typical of their use (a box used as a receptacle) prompted a use-typical representation that made it difficult to find a solution that involved atypical use of the objects. Presenting the same objects separately, independent of their typical function, increased problem solving performance by allowing a less 'fixed' mental representation to be developed. Whilst the influence of this research faded with the expansion of behaviourism in the 1950s, its legacy was to make the hitherto inscrutable process of problem solving amenable to explanation by cognitive processes. This was the basis upon which cognitive explanations of complex task performance were founded.

The cognitive revolution in psychology in the late 1950s brought renewed attention to cognitive processes, particularly those underpinning problem solving. This research was lead by Allen Newell & Herbert Simon. Newell & Simon were early pioneers of the information processing approach to cognition that likened human cognitive processing to the operations of a computer program (Newell & Simon, 1961). They posited that the cognitive system underpinning problem solving comprised of a set of elementary processes that would operate on information held in a set of memory stores to generate a problem solution (Newell, Shaw, & Simon, 1958). To examine the nature of these stores and processes, they focused on identifying general problem solving strategies that could be used in a broad range of domains. They first did so in the domain of proving logic theorems in their highly influential 'elements' paper (Newell et al., 1958), subsequently extending this research to demonstrate that similar strategies were used in a range of different problem solving settings (Newell & Simon, 1972; Simon & Newell, 1971). Problem solvers were, for example, often found to use a means-end strategy whereby they would work backwards from a goal in an attempt to find a solution or a hill-climbing strategy where they would continually try, in every successive step, to move in the direction of a goal. The identification of such domain-general strategies not only permitted insight into how humans solved complex problems but also the nature of the memory stores upon which these strategies depended.

Newell & Simon had noted that the general strategies used by many problem solvers were almost invariably performed in a slow, step-by-step, serial manner. This suggested that the capacity of the memory system supporting these strategies was highly limited since it did not permit parallel processing of any problem solving steps (Newell et al., 1958). Newell & Simon therefore suggested that problem solving, at least in novel settings, relied on short-term (working) memory (Newell et al., 1958; Simon & Newell, 1971; Simon, 1970), a store that had recently been discovered to hold only seven (plus or minus two) pieces of information at any given time (Miller, 1956). Subsequent studies supported this conception leading to the characterisation of the cognitive system as a serial information processor dependent on a limited capacity short-term memory store (Newell & Simon, 1972). However, whilst this was an undoubtedly informative account of the cognitive processes supporting problem solving, a further component of memory was required to more comprehensively account for the range of problem solving behaviour evident in everyday life.

Domain general, or so-called ‘weak method’, strategies were slow and highly effortful and could not therefore adequately represent performance in everyday life (Newell & Simon, 1972). The use of the general strategies also had been found to decline with practice as problem solvers switched to more domain-specific and effective ‘strong method’ problem solving strategies (Anzai & Simon, 1979; Newell et al., 1958; Newell & Simon, 1972). The proposed cognitive system therefore required a further memory store that could hold information gained from experience and use it to facilitate the problem solving processes operating within short-term memory. This store was long-term memory and it became greater focus as research progressed through the late 1960s and 1970s.

By the late 1960s, there was a growing realisation of the importance of task knowledge held in long-term memory for problem solving. Previously, existing knowledge had been excluded in an attempt to ensure that observations of general problem solving strategies were not confounded by differences in task knowledge.

However, there was growing evidence that even in novel tasks, differences in knowledge could influence problem solving. For example, performance on abstract problem solving tasks could be dramatically improved by likening them to real world examples where individuals could make use of their existing knowledge (Simon, 1970; Wason & Johnson-Laird, 1972). Even in abstract tasks, if problem solvers possessed some domain relevant knowledge, performance could be improved (Chi, 1978). When existing knowledge could be employed, problem solving was faster, less effortful and less error-prone than the general problem solving strategies identified by Newell & Simon. Research focus therefore turned to explaining how the knowledge stored in long-term memory could facilitate problem solving and complex task performance. The natural place to begin such investigations was to compare those whose knowledge differed most: experts and novices.

Expertise.

Developments in the understanding of how differences in knowledge influenced the performance of complex tasks derived initially from comparisons between novice and expert performers (Feltovich, Prietula, & Ericsson, 2006). This research derived from the seminal work of de Groot who had sought to explain why in chess, highly skilled players would almost invariably beat those who were less skilled (de Groot, 1978/1946). At the time of de Groot's research, the prevailing understanding of chess expertise was based on innate abilities. Highly skilled players simply possessed better reasoning skills or could think more moves ahead and could thus 'out-think' players of lesser abilities. de Groot's investigations however suggested that it was not relative strength in innate cognitive capacities, but a superior memory of chess board configurations built up through years of experience. de Groot presented a series of chess board configurations to both novice and expert players for a short, 5 second, duration and then asked them, following a delay, to reconstruct the board configurations they had viewed. Surprisingly, even after such a short exposure, expert players were able to recreate almost the entire board consisting of 25 pieces whilst novices could manage only around five. Memory therefore seemed an important source of chess experts' superiority.

Chase and Simon (Chase & Simon, 1973a, 1973b) extended DeGroot's work by presenting novice and expert chess players with chess board configurations that were either legal, where the positioning of the pieces conformed to the rules of chess, or illegal, where the pieces were placed randomly with some in violation of the rules of chess (e.g. bishops placed on the same colour squares, pawns placed on the rearmost row of the board, kings placed well in advance of other pieces, etc). Recall of each board configuration was then tested in a similar way to de Groot (1946). Like de Groot's studies, experts were vastly superior to novices when configurations were legal. However, when configurations were illegal, experts' recall was only a few pieces higher than novices. The source of experts' superior performance in de Groot's studies could not therefore be attributed the experts' superior short-term memory or reasoning abilities because their superiority was only evident for legal board configurations. It therefore seemed likely that the experts' experience playing chess had somehow contributed to their superior memory performance.

Chase & Simon (1973a, 1973b) suggested that the reason for experts' superiority in recalling legal board configurations was because they possessed a greater, and better structured, store of chess-related knowledge. Rather than memorising every individual piece position, Chase & Simon argued that experts' years of playing chess had enabled them to develop a substantial memory of board configurations that were grouped and organised into well-structured patterns called *chunks*. These chunks enabled experts, when presented with a legal board configuration, to perceive a configuration rather than a number of individual pieces that could then be rapidly matched to a chunk already stored in long-term memory. Experts consequently had to recall only a single, already memorised, chunk to correctly reconstruct an entire board. Since novices lacked the experience to develop chunks of knowledge, they could only attempt to remember board configurations piece-by-piece, and were therefore subject to the limits of short-term memory (Miller, 1956). When pieces were positioned illegally, experts could not benefit from their chunked knowledge since it was relevant only to recognising familiar, legal, board configurations. For illegally positioned pieces, experts were therefore subject to the same limitations as novices, explaining their similar performance. The exceptional recall of chess experts for legal board configurations therefore appeared due to their superior and better-structured knowledge in long-term memory. Further, since chess experts almost invariably beat novices in games, differences in knowledge seemed to explain differences in complex task performance. In complex task performance it appeared that “novices need to use thinking skills [while] experts use knowledge” (Sweller, Ayres, & Kalyuga, 2011, p21).

Chase and Simon (1973a, 1973b) also noted however, that experts appeared to organise their knowledge of board configurations according to the more abstract and strategic aspects of the game of chess, such as configurations’ suitability for attack or defence. This suggested that not only did experts hold chunks of information in long-term memory, but that they organised these chunks in a way that facilitated performance. This insight was more fully articulated by Chi, Feltovich, & Glaser (1981) in their seminal publication concerning expert-novice differences in categorisation. Chi and colleagues asked experts (academics and advanced graduate students) and novices (first year university students) to sort physics problems into groups they would solve in a similar way. Experts sorted the problems according to their deep underlying structure (the theorems relevant to solving each) whereas novices did so only according to superficial features (e.g., whether the problem involved an incline plane). This suggested that experts represented their knowledge at a deeper, more structural, and abstract level than novices. Further, since such representations likely facilitated performance by helping to identify the elements of a problem most relevant to its solution, experts' knowledge appeared to be structured in a way that actively facilitated their performance (Chi, Glaser, & Rees, 1982; Chi et al., 1981).

Chi and colleagues referred to the organised knowledge structures of experts as *schemas* (based, originally, on the theories of Piaget, (1928), and Bartlett, (1932), They were defined as cognitive constructs that allowed the many elements of a task to be chunked into a unitary structure in a way that was consistent with their common use (Chi et al., 1982; Chi, Glaser, & Farr, 1988). The benefits of schemas therefore were not only to reduce the load on short-term memory by grouping familiar elements together but also to guide responses based on how like tasks had been performed in the past. In other words, once a task was recognised as matching a schema stored in long-term memory, one could simply apply schemas to produce an almost pro forma response. In a review of the research concerning the influence of schemas, particularly on categorisation tasks, Zeitz (1997) found that schemas also facilitated integration and retrieval of relevant information from memory, aided the filtering of relevant from irrelevant task information during task performance, and provided abstract representations that could aid reasoning when situations were uncertain. The way in which knowledge benefited complex task performance was therefore not simply through its content. More abstract and representative knowledge organisation also benefited complex task performance.

With growing understanding of schemas and their influence on complex task performance, more research began to investigate how they were acquired. As implied by Chi's definition (Chi et al., 1982; Chi et al., 1988), one fundamental requirement was practice (Ericsson, 2006; Ericsson, Krampe, & Tesch-Romer, 1993). To begin to acquire a schema learners first had to practice performing a number of similar tasks to begin to extract their common elements (Chi et al., 1982; Chi et al., 1981). These elements could be integrated into schemas and, with further practice, refined to a level where they could be employed almost without conscious effort (Shiffrin & Schneider, 1977). However, whilst practice was undoubtedly necessary for schema acquisition, it was not sufficient (Ericsson et al., 1993). Learners could for example, fail to acquire abstract and representative schemas even after substantial task experience, especially if instructional design was poor (e.g. Glaser, 1987; Klahr & Nigam, 2004). The conditions under which learning, or schema acquisition, occurred therefore appeared highly relevant to its success.

How different conditions influence the acquisition of the rule-based knowledge that form schemas is the focus of the present thesis. Whilst there are numerous theories of learning that attempt to explain why and how certain conditions facilitate or impede learning, the present focus is on two: Cognitive Load Theory (CLT) and Dual Space Theory (DST). These theories were selected because they offer competing, but potentially complementary, accounts of how effective schema acquisition occurs in complex settings. The following chapter will outline and compare the accounts of each.

Summary.

The present chapter aimed to provide an historical overview of cognitive research on complex task performance. Problem solving research was perhaps the first area to examine the role of cognitive processes in complex task performance and provided significant insights into the nature of memory stores and the fundamental cognitive processes used to perform novel, complex tasks. This research gave rise to investigations of the role of knowledge in complex task performance. By comparing experts and novices in complex settings, researchers were able to elucidate how the content and organisation of knowledge stored in long-term memory (i.e. schemas) could facilitate performance. Explorations of how such schemas could most effectively be acquired in complex settings revealed the necessity but insufficiency of task practice. Learning conditions also needed to be conducive to learning to facilitate effective schema acquisition. The focus of the present thesis is on how different learning conditions affect the successful acquisition of schemas. It focuses on two theories: CLT and DST, that offer competing explanations of the conditions that facilitate and impede schema acquisition in the domain of complex task performance. Chapter 2 provides an overview and comparison of each.

CHAPTER TWO:

COGNITIVE LOAD AND DUAL SPACE THEORIES

Cognitive Load Theory (CLT) and Dual Space Theory (DST) are cognitive theories of learning that differ in their explanations of how knowledge is acquired. On the one hand, CLT argues that knowledge acquisition is limited by the information processing capacity of working memory. This theory therefore contends that minimising processing demands on working memory will benefit learning. On the other hand, DST argues that knowledge acquisition requires an extensive search for the rules and principles that govern a task's operation. This theory therefore suggests that encouraging greater rule-search behaviour will improve learning. This chapter outlines both CLT's and DST's explanations of knowledge acquisition in complex settings. It compares the accounts of each theory before evaluating the claim made by CLT researchers that the theories' are not independent but complementary. The chapter finally presents a rationale for investigating the independence of each theory's account of knowledge acquisition.

Cognitive Load Theory

CLT is, in essence, a theory of instructional design. Its aim is to explain why different instructional procedures do and do not work. The theory is based on a cognitive architecture that provides a comprehensive explanation for how knowledge is acquired, particularly in complex settings. The architecture derives from established cognitive psychological research, particularly in the fields of problem solving and expertise (outlined in Chapter 1), and specifies how the cognitive components that comprise human cognition, long-term and working (short-term) memory, interact to determine how knowledge is acquired. For CLT, it is the characteristics of these two memory stores that are fundamental to the theory's account of learning and provide a clear account of how learning is facilitated or impaired.

Long-term memory in CLT, like most theories of cognition, is an effectively unlimited information store where information has been organised into broad categorical and functional themes called schemas. As explained in Chapter 1, schemas are vital to complex task performance because they permit information to be grouped into more easily processed chunks of information (Chase & Simon, 1973a, 1973b; de Groot, 1978) that very efficiently guide the processing of incoming information (Chi et al., 1982; Chi et al., 1981). Schemas therefore make performing the complex tasks for which they are relevant far more straightforward than they would be without such knowledge. Given their benefit, CLT defines learning as an increase in the functional schemas stored in long-term memory (Sweller et al., 2011). As a theory of learning, the theory's aim is therefore to explain how schemas are acquired.

For CLT, schema acquisition depends fundamentally on the characteristics of the second memory store: working memory. This store serves as the intermediary between incoming information from the environment and long-term memory. Its role is to hold and process incoming information, either combining it with relevant schemas from long-term memory to guide responses, or integrating it into schemas to store in long-term memory. The information-processing capacity of working memory is however extremely limited (Cowan, 2001; Miller, 1956; Peterson & Peterson, 1959). Whilst some researchers have suggested working memory can hold up to nine elements at any given time (Miller, 1956), in particularly complex settings it may be as low as four (Cowan, 2001). Crucially for CLT, this limited capacity restricts the rate that new information can be integrated into long-term memory. The theory therefore argues that if processing demands exceed working memory capacity, learning will be impaired. Minimising the processing demands on working memory, (i.e. *cognitive load*) will therefore facilitate learning.

Minimising cognitive load is particularly relevant to knowledge acquisition because processing novel information places greater demands on working memory than familiar information. When information incoming to working memory is familiar it activates schemas in long-term memory that efficiently determine how it should be processed. This places minimal demands on working memory because schemas require minimal processing (Chase & Simon, 1973a, 1973b) as there is no need to determine how the information should be processed. So minimal is this cognitive load that some researchers have proposed a separate working memory for familiar information (Ericsson & Kintsch, 1995). By contrast, novel information cannot, by definition, activate schemas in long-term memory. It must instead be held in working memory and processed to determine how it can be used to achieve a desired outcome. CLT proposes that this can be achieved by either borrowing the schemas of others to guide information processing, by either using instructions or imitating more experienced performers (Bandura, 1986), or, if this is not possible, by randomly generating a response to the information and evaluating its effectiveness against a desired outcome (Sweller et al., 2011). Whilst the former method produces lower cognitive load than the latter, both produce higher cognitive load than processing familiar information. Learning to perform a complex task that contains a large amount of novel information may thus easily produce levels of cognitive load that impair learning. Minimising cognitive load during learning is therefore of particular importance for CLT.

Cognitive load is not caused only by the type of processing incoming information requires. It also depends on the nature of the information itself. CLT defines the load caused by incoming information according to the number of elements it contains that must be processed together for a complete understanding to be achieved. This is referred to as the information's *element interactivity* (where an element is defined as component of the information that must be processed). Elements must be processed together when they are logically related. For example, learning the elements of the periodic table does not require simultaneous processing because each element can be processed independently. Learning simply requires that they be memorised. Conversely, developing knowledge about a chemical equation, such as for the combustion of hydrogen in air ($2H_2 + O_2 = 2H_2O$ + heat) requires substantial simultaneous processing of the elements that cannot be learnt in isolation. All elements need to be processed together in working memory to permit development of a complete and accurate schema for describing how hydrogen and oxygen combine. For CLT, achieving this constitutes a complete understanding of the information. Tasks or instructional designs characterised by high levels of element interactivity are likely therefore to produce high levels of cognitive load and commensurately slow or impaired learning. Since learning necessarily involves processing high-load novel information, minimising the element interactivity of a task is, for CLT, the primary means by which cognitive load can be minimised. One of the primary aims of the theory is therefore to explain how this can be achieved.

Minimising the element interactivity of a task is of course only possible to the extent that incoming information can be simplified without losing its meaning. It is difficult to imagine for example how the aforementioned combustion equation could be simplified further whilst still conveying how oxygen and hydrogen combine. CLT refers to the load caused by the unalterable characteristics of a task as *intrinsic load* (Sweller et al., 2011). Since this load is intrinsic to the information it cannot therefore be reduced. Whilst it may seem from the above exposition that tasks whose intrinsic load is beyond the capacity of working memory may never be fully understood, learning can overcome such situations. Initial processing may integrate some of the interacting elements into a schema thereby reducing the element interactivity of the task, eventually allowing all elements to be processed together. Numerous studies have for example demonstrated the efficacy of pre-training (Mayer, Mautone, & Prothero, 2002), constructing learning sub goals (Catrambone, 1998; Catrambone & Holyoak, 1990), or teaching decomposed parts of a complex task (Pollock, Chandler, & Sweller, 2002) for constructing preliminary schema to facilitate learning of highly complex tasks. Intrinsic load cannot then be minimised, but only overcome with learning. The focus of CLT is therefore not on reducing intrinsic cognitive load but on the load caused by non intrinsic factors[1].

Cognitive load caused by task characteristics that are not central to understanding are referred to by CLT as *extraneous load*. This load derives not from the nature of the task itself but from the manner in which it is presented. Extrinsic load adds to the cognitive load caused by intrinsic factors by adding, unnecessarily, to the element interactivity of incoming information. If intrinsic load is already high, additional extrinsic load may overwhelm working memory capacity and consequently impair learning. As a theory of instructional design, CLT focuses on instructions as the primary source of extrinsic load. For example, instructions for the same complex

---

[1] Some CLT researchers differentiate the intrinsic load of incoming information from that required to develop schema, i.e. *germane load* (e.g. Hilbert & Renkl, 2009; Paas & Van Merrienboer, 1994b). Since germane load is caused by schema development, increasing germane load has been argued to facilitate rather than impair learning (Hilbert & Renkl, 2009; Schnotz & Kurschner, 2007). The concept of germane load has however been discounted by Sweller and colleagues (Sweller et al., 2011). Since differentiating germane from intrinsic load is not the focus of the present dissertation, and since there is debate concerning its existence, germane load will not be considered further.

task may be either minimal or extensive and thereby produce different levels of extraneous cognitive load. As explained previously, minimal conditions are unlikely to provide sufficient information to guide task responses. A learner would therefore to have to randomly generate and test possible responses to the task, gradually building up schema for how to, and not to, perform the task. This approach produces high cognitive load since the learner must generate and keep track of each attempt in working memory. Conversely, providing extensive guidance would reduce the need to randomly generate and test possible response options by conveying which responses were likely to be effective. This would allow a learner to more readily acquire the schema for task performance under conditions of lower cognitive load. The higher cognitive load caused by the minimal instructions would therefore be extraneous since it is unrelated to learning to perform the task. In an analogous argument, in the 13th Century, Roger Bacon suggested that it would take between 30 and 40 years to master mathematics using self study (Singer, 1958, cited Ericsson, 2006). Using today's teaching methods, roughly equivalent knowledge can be acquired in under 12 years by the majority of high school students (Ericsson, 2006). Since extrinsic load adds unnecessarily to the cognitive load experienced during learning, the principal focus of CLT is to suggest ways of minimising extraneous load in instructional designs.

In summary, CLT is a theory of instructional design based on a cognitive architecture that specifies how learning is achieved. Fundamental to the theory is the limited capacity of working memory, the intermediary between incoming information from the environment and long-term memory. This limited capacity, particularly when processing novel information, means it can easily be overwhelmed, impairing the integration of the novel information into memory. The cognitive load on working memory derives from the element interactivity of information it must process to achieve a desired outcome. High element interactivity that is intrinsic to a task cannot be reduced (without losing important information) but, if it produces excessive cognitive load, can be overcome with learning. Element interactivity that is extrinsic to a task, i.e that which derives from the task's presentation rather than the task itself, places additional and unnecessary load on working memory. If intrinsic load is already high, this can overwhelm working memory and impair learning. Extrinsic load can however be reduced. To maximise learning, CLT therefore advocates minimising the extraneous cognitive load of a task's presentation. This is the primary means by which the theory suggests to facilitate learning.

Application: the goal free effect.

The principles of CLT have been applied to a number of effects in the learning literature. For instance, providing a worked example during learning has been shown to reduce cognitive load and improve learning by providing learners with a schema that guides their processing of the novel problem information (Cooper & Sweller, 1987; Sweller & Cooper, 1985). The negative influences of redundant information on learning can also easily be understood as introducing extraneous element interactivity that increases cognitive load and consequently impairs learning (Chandler & Sweller, 1991). Most relevant to the present thesis however is CLT's explanation of how goal assignment can affect learning.

Early motivational researchers had found that assigning specific and highly challenging performance goals lead to superior performance outcomes across a broad range of tasks (Locke & Latham, 1990, 2002; Mace, 1935). Such goals were believed to operate by effectively directing attention to goal-relevant aspects of the task, then increasing effort and persistence to achieve the goal. Cognitive researchers had however found that specific goals produced poorer outcomes when assigned during learning, particularly if the task was complex (Sweller, 1988; Sweller & Levine, 1982; Wood, Mento, & Locke, 1987). CLT argues that this was due to specific goals eliciting a means-end strategy that created extraneous cognitive load and consequently impaired learning.

Newell & Simon (Newell, 1973; Simon & Simon, 1978) had identified means-end strategies as a general weak-method strategy learners often adopted when performing novel tasks. Learners using this strategy would repeatedly compare their current task state to the goal state and consistently act to reduce the distance between the two. Sweller & colleagues however demonstrated the use of means-end strategies lead to poorer learning outcomes than strategies that did not involve a current-goal state comparison (Mawer & Sweller, 1982; Sweller & Levine, 1982; Sweller, Mawer, & Howe, 1982; Sweller, Mawer, & Ward, 1983). Since specific goals provided a point of comparison, they also appeared to encourage the use of means-end strategies, thus impairing learning (Owen & Sweller, 1985; Sweller, 1988).

The reason means-end strategies produced poorer learning outcomes was due, according to CLT, to their effect on extraneous cognitive load. As demonstrated by Sweller (1988), means-end strategies involved greater element interactivity than simply using previous task experience because learners had to keep both the goal state and current state active in working memory whilst also trying to determine how to move closer to the goal. This resulted in a high level of cognitive load on working memory and consequently poor learning. Comparatively, when no goal was provided, learners could focus predominately on the current state and on any way of moving from that state, thus removing the need to make continual comparisons to the goal. Specific goals therefore elicited a task strategy that produced greater cognitive load than the provision of no goal. When cognitive load was high, such as the case when learning to perform a complex task, this additional load produced poorer learning[2].

The ‘goal specificity effect’, or ‘goal-free effect’ as described by CLT, was the genesis of CLT. It was the first effect to be formally explained by the theory and served as the basis for investigations of other CLT effects such as the influence of worked examples or informational redundancy. It is therefore of fundamental importance to the theory. It is also the effect in which the explanations of CLT and DST most clearly differ. In comparing CLT and DST, the present thesis’ focus is on comparing each theory’s account of the goal-free effect in learning.

The following section outlines DST and its explanation of the goal-free effect before the theories are compared in the section thereafter.

Dual Space Theory

DST is primarily a theory of problem solving. Its central focus is on describing the cognitive processes involved when learning to solve novel problems. In describing these processes however, the theory also presents an explanation for how learning outcomes may differ depending on the nature of the task. DST conceives of learning as a search within two internal problem representations referred to as *spaces*. Each space represents different task aspects with the content of learning dependent on the space searched. DST therefore proposes that it is the type of processing (i.e. space search) elicited by a task, rather than the amount of processing (i.e. load), which influences learning. Further, since it is predominately the nature of the task that determines search space, the theory proposes that aspects of a task’s presentation are

---

[2] Sweller (1988) also argued that specific goals also impaired learning by reducing the amount of task exploration in which learners engaged. Assigning specific goals therefore impaired the extent of knowledge gained about a task, subsequently affecting transfer performance. This explanation has not however been included in the most recent formulation of CLT (Sweller et al, 2011) and has not therefore been included in the current description.

important to learning outcomes. For DST, the influence of task characteristics on search space is fundamental to learning outcomes.

DST's conception of problem solving as a search derives from Newell & Simon's (1972) general theory of problem solving. Under this conception, problem solving was regarded as a search through a space of possible problem states described as *problem space* to find a state corresponding to a solution. Each state in problem space represented a level of knowledge that had been acquired about a problem. To develop knowledge, learners would use general, 'weak method', strategies to draw inferences from their current knowledge to 'discover' new problem states that were closer to the goal. Once knowledge was sufficient to identify the goal state, a solution would be achieved. Since Newell & Simon's aim was to elucidate the general cognitive processes underpinning problem solving, they suggested that problem space search represented the common means by which individuals solved a variety of novel problems. However, whilst providing a convincing theory for problem solving tasks, where learners sought to reach a goal, it did not readily account for some other types of problem solving, in particular rule induction (e.g. discovering the rule governing a number sequence). Simon & Lea (1974) therefore proposed DST to overcome this limitation.

Newell & Simon's (1972) theory did not appear an adequate explanation of rule induction tasks for two reasons. First, such tasks initially provide all relevant task states thereby obviating the need to search problem space to 'discover' states corresponding to a solution. Second, induction tasks required discovery of a rule rather than a goal state so matching a problem to a solution state could not have been an accurate description of the cognitive processes involved (Simon & Lea, 1974). Despite these apparent differences, Simon & Lea (1974) noted that similar cognitive processes could apply to both general problem solving and rule induction tasks. In both, problem solvers needed to construct an internal problem representation and then use general search strategies to develop and verify their knowledge until it was sufficient to reach a solution. The primary difference between the task types then was not in the cognitive processes by which a solution was achieved but in the type of information being searched. Simon & Lea therefore proposed that problem solving involved a search for problem states whilst rule induction, a search for task rules or concepts.

To accommodate the different types of information involved with problem solving and rule induction tasks, Simon and Lea (1974) proposed DST, a theory of two problem spaces, each corresponding to a different type of information. *Instance space* was similar to Newell & Simon's problem space and consisted of all possible problem states (i.e. instances) including the operators and processes required to transform the states to find a solution. *Rule space*, on the other hand, consisted of all possible rules governing a task's operation as well as the operators and processes required to generate, modify and test the rules in developing a solution. The nature of the task largely determined the space searched during learning with standard problem solving tasks involving greater search of instance space and rule induction tasks a greater search of rule space. However, whilst the cognitive processes supporting learning in each space were similar to those proposed by Newell & Simon (1972), for rule space search, the process of verifying knowledge was somewhat unique.

Knowledge generated in any search space needs to be tested to determine whether it is effective for bringing a learner closer to a solution (Newell et al., 1958; Newell & Simon, 1972; Simon & Lea, 1974). This requires applying knowledge developed through search to the task. Since all tasks are comprised of instances, knowledge testing must therefore be performed in instance space. For knowledge generated in instance space then, both the generation and testing of knowledge occurs in instance space only. For knowledge generated in rule space however, testing must also occur in instance space. DST therefore proposes that rule and instance spaces, whilst conceptually distinct, interact, particularly when searching rule space. Rule space search generates rules that are applied to task instances in instance space with results informing further search and development of knowledge in rule space. Rule induction tasks therefore involve both rule and instance space search, while problem solving tasks, predominately instance space search.

Whilst Simon & Lea (1974) suggested that search space was dictated largely by task type, they also noted that it could be influenced by the way a task was presented. Encouraging greater rule space search on an ostensibly problem solving task could, for instance, produce greater rule space search and consequently greater rule learning. Instructing learners in the Tower of Hanoi task to identify rules for moving disks from one peg to another (rather than the usual goal of simply moving the disks from left to right), would, for example, change the task from a problem solving to a rule induction task. As a result, learners would likely acquire a greater knowledge of the task rules that, under normal conditions, would have been unlikely. DST does not therefore simply describe the cognitive processes prompted by different task types. It also presents an explanation for how task variations can influence learning outcomes. Variations that encourage greater search of rule space are argued to produce greater rule learning.

Simon and Lea (1974) distinguished between the knowledge generated by rule and instance space search only on the basis of content, not benefit. Rule and instance knowledge were considered according to their appropriateness to the task rather than any inherent advantage/disadvantage of either type. However, the implication of DST is that rule and instance knowledge are not of equivalent value. The characteristics of rule knowledge as constituting understanding of the rules governing a task's operation mean it is likely to be more beneficial to the creation of organised and abstracted knowledge representations (i.e. schemas) than knowledge of specific task instances. Knowledge of task instances is likely to be sufficient to reach a specific goal in a specific circumstance whilst an understanding of how a task operates is likely to assist in reaching multiple goals within the same task or even across tasks sharing similar rules (Gentner, 1983; Gick & Holyoak, 1983; Novick & Holyoak, 1991). Although beyond the expressed bounds of Simon and Lea's theory, DST therefore represents an explanation for why the level of learning may differ according to a task's presentation, or between different people. Task conditions that elicit greater rule space search, or individuals who engage in greater rule space search, are likely to produce better learning.

Klahr & Dunbar (1988) extended DST to the study of scientific reasoning and, in doing so, provided the first clear evidence for the superiority of rule over instance space search. Based on Simon & Lea's (1974) explanation of rule space search, Klahr & Dunbar proposed that scientific reasoning involved development of hypotheses about how a task functioned in hypothesis space (an equivalent to rule space) followed by the testing of these hypotheses in experiment space (an equivalent to instance space). Reasoning was therefore proposed to involve an interactive search between rule and instances spaces in accordance with DST. To test their model, Klahr and Dunbar asked participants to think aloud while attempting to work out the function of a novel computer programming command in a simple programming system. Both task instructions and the command's label led participants to initially construct an erroneous belief about how the command operated so an extensive search was required to overcome the mistake and determine the correct function. Results showed that participants indeed searched both rule and instance space in coming to the correct understanding of the command. However, they also indicated that participants who demonstrated a greater search of rule space, by testing specific hypotheses about the function, did so more quickly and efficiently. To explore this result further, Klahr & Dunbar attempted to increase rule space search in another group of participants by asking them to formulate as many hypotheses about the command as possible before commencing the task. These participants were found to reach the solution faster and to use more effective strategies than the first group. Encouraging an initial search of rule space through hypothesis generation and testing therefore appeared to benefit both the speed and systematicity of problem solving. Klahr & Dunbar's findings strongly suggested then that the benefits of searching rule and instance space were not equivalent. Rule space search, as manifested by hypothesis testing, appeared to produce faster and more effective learning than instance space search.

Application: the goal free effect

The apparent superiority of rule space search demonstrated by Klahr & Dunbar (1988) represented an alternative means of interpreting a number of known learning effects, in particular the goal free effect identified by Sweller and colleagues (e.g. Owen & Sweller, 1985; Sweller, 1988). Vollmeyer and colleagues (Vollmeyer, Burns, & Holyoak, 1996) proposed that specific and non specific goals produced different learning outcomes not due to differences in cognitve load but because they encouraged different levels of rule space search. Specific goals were argued to promote instance space search at the expense of rule space search because they elicited means-end strategies in which learners would compare their current state to the goal, both of which were represented in instance space. By comparison, non-specific goals provided no direction for how to search instance space and thus encouraged learners to use rule space search to direct their search of the problem (a suggestion supported by Hagmayer, Meder, Osman, Mangold, & Lagnado, (2010). Learners provided a non-specific goal were therefore more likely to focus on learning how a task worked (i.e. learning the rules governing its operation) whilst those provided a specific goal, only on achieving a goal. Non-specific goal learners were therefore anticipated to develop greater rule knowledge that would facilitate their performance on a greater range of task goals than those assigned specific goals.

To test their dual-space account of the goal free effect, Vollmeyer et al (1996) examined participants' use of strategies when learning to perform a complex dynamic control (CDC) task. Participants were trained under either a specific or non-specific goal condition with half also instructed on the use of a highly systematic task strategy ('vary one thing at a time' – VOTAT) previously shown to facilitate learning. Task knowledge, strategy use during training, and success at reaching post training and transfer goals were assessed. In addition to demonstrating a clear goal-free effect, with non-specific goal participants developing more extensive knowledge and showing greater ability to reach the transfer goal, strategy analyses revealed that specific goal conditions showed higher goal-directed (i.e. instance space search) strategies during training. Additionally, whilst almost all participants given VOTAT instruction used it initially, specific goal participants quickly switched to means-end/goal-directed strategies whilst those given a non-specific goal continued to use VOTAT throughout training. Specific goals therefore appeared to strongly encourage instance space search, despite the ready availability of a more effective strategy, and therefore impaired learning. Vollmeyer et al’s (1996) findings therefore suggested DST as a viable explanation of the goal-free effect.

Whilst Vollmeyer et al’s (1996) findings showed that specific goals elicited instance space search, they did not demonstrate that non-specific goals elicited rule space search. Their findings indicated only that non-specific goal participants used more systematic strategies, not whether such strategies constituted a greater search of rule space. To more directly test the effects of goal specificity on rule space search, Burns & Vollmeyer (2002) conducted a further study utilising verbal protocols to examine the nature of the strategies elicited by the different goal types. Participants were trained on a simplified version of their previous CDC task under identical specific or non-specific goal manipulations. In this study however, all participants received VOTAT instruction prior to training and half were provided an incorrect hypothesis about the task to encourage rule space search. The strategies identified from verbal protocols collected during training were classified into three categories: ‘goal-oriented’, where participants would explicitly attempt to bring task outputs closer to the goal, ‘hypothesis testing’, where participants would manipulate a system input with a specific expectation about its effect on outputs, or ‘nonpredictive testing’ where participants would change an input with no expectation of the effect. Hypothesis testing was considered indicative of rule space search since it required a participant to have formed an hypothesis about how the task operated in rule space before testing it using the strategy. Whilst results again showed that non-specific goal participants learnt more about the task and could more successfully reach the transfer goal than specific goal participants (although provision of the incorrect hypothesis to specific goal participants mitigated the differences), strategy use between the groups appeared in direct support of the theory. After the first round of training, where all participants used similar strategies, the vast majority of specific goal participants switched to goal-oriented strategies whilst the majority of non-specific goal groups used hypothesis testing. Moreover, those provided with the incorrect hypothesis also appeared to engage in greater hypothesis testing overall. Results therefore provided clear evidence that non-specific goals encouraged rule space search and, furthermore, that rule space search facilitated learning. In direct support of DST then, differences in learning outcomes elicited by specific and non-specific goals appeared to be due to the extent to which participants searched rule and instance space during training.

Other researchers have also attributed goal specificity effects to differences in rule and instance space search. Geddes and Stevenson (1997) for example examined the influence of specific and non-specific goals on implicit learning. They proposed that learners failed to acquire explicit, verbalisable, rule knowledge in implicit learning tasks because they were typically assigned specific goals, which likely impaired learning. To test this, Geddes and Stevenson trained participants on an implicit learning task, based closely on that used by Berry & Broadbent (1984), under specific or non-specific goal conditions. Consistent with previous results, non-specific goal participants again demonstrated superior performance to specific goal groups after training. However, results also indicated that non-specific goal learners could accurately verbalise the majority of task rules and could effectively apply them to predict the outcomes of novel (i.e. previously unseen) task inputs. By contrast, specific goal participants developed almost no verbalisable rule knowledge and could only predict the outcomes of the task inputs they had previously experienced during training. Non-specific goals therefore appeared to encouraged rule learning, whilst specific goals, instance learning. Geddes & Stevenson concluded that specific goals appeared to encourage instance space search at the expense of rule space search, consequently impairing rule acquisition, whereas non-specific goals encouraged the “free exploration of rule space unimpeded by the need to find a route to a specific goal” (p761), that was necessary for the acquisition of rule knowledge. Findings were therefore in direct support of DST.

Osman (2008), also invoked a DST explanation for her investigation into the effects of goal specificity and observational learning on knowledge acquisition. Utilising the same task as Burns & Vollmeyer (2002), Osman trained participants under a specific or non-specific goal condition where they either manipulated task inputs themselves (action) or watched a demonstration of another participant manipulating task inputs (observation). Replicating previous findings, Osman found that non-specific goal participants again performed better on both training and transfer tasks, demonstrating superior procedural and declarative knowledge for specific goal participants. Notably, no differences were observed between the action and observation groups suggesting that observation produced similar learning to performing the task. Although not directly tested, Osman concluded that the superior knowledge acquired by both non-specific goal groups was due to their greater testing of hypotheses generated in rule space, in further support of DST.

To summarise, DST proposes that the goal free effect can be explained by differences in rule and instance space search. Specific goals encourage instance space search by eliciting goal-directed or means-end strategies whereby learners compare their current and goal states in instance space, and act to reduce their difference. By comparison, non-specific goals encourage rule space search by providing no direction on how to search instance space. This leads learners to develop hypotheses about a task’s operation in rule space before testing them in instance space. Non-specific goals therefore facilitate the development of rule knowledge whilst specific goals, knowledge about the task instances leading to a specific goal. Since rule knowledge can facilitate performance across a greater range of tasks than knowledge of how to achieve a specific goal, non-specific goals encourage better learning outcomes than specific goals. Numerous studies have found support for this interpretation of the goal free effect. Specific goals have repeatedly been found to elicit more goal-directed strategies and poorer learning whilst non-specific goals with hypothesis testing strategies and superior learning. DST therefore represents a viable explanation for the goal free effect.

Comparison of CLT and DST

As demonstrated above, CLT and DST represent alternative explanations for how task conditions can influence learning outcomes in complex tasks. For CLT, conditions high in element interactivity place excessive cognitive load on working memory, consequently impairing the processing of information into long-term memory. Minimising element interactivity is therefore the prime means CLT suggests for facilitating learning. For DST, conditions that encourage learners to focus only on learning task instances are argued to impair learning by preventing adequate focus on acquiring rule knowledge. Encouraging learners to search rule space by developing and testing hypotheses about task rules is therefore the approach DST proposes to facilitate learning. The differences between the theories are most apparent in their respective explanations of the goal free effect.

Whilst both theories agree that specific goals produce poorer learning outcomes than non-specific goals in complex learning tasks, they disagree as to why. CLT argues that specific goals impose a higher cognitive load than non-specific goals and therefore produce poorer learning. This is because specific goals elicit means-end strategies that require a learner to simultaneously compare their current and goal states whilst also trying to work out how to minimise the difference between them. This represents a large number of interacting task elements that consequently produce a high level of cognitive load, potentially overloading working memory. Non-specific goals do not create this load because they provide no point of comparison and therefore preclude use of means-end strategies. Non-specific goals therefore permit available resources to be used for working out potential ways of simply manipulating the task, rather than doing so to achieve a goal (Sweller, 1988; Sweller et al., 2011).

DST proposes that specific goals encourage learners to focus on task instances or states, which prevents their acquisition of rule knowledge. Whilst agreeing that specific goals elicit means-end or goal directed strategies, DST argues that it is the strategy's focus on task instances that impair learning. Means-end strategies involve comparison between task instances only, and therefore largely ignore task rules. They therefore permit learning of how to achieve a specific goal from a set starting point through a path of task instances but not the acquisition task rules governing this path. Since non-specific goals provide no instances for which to aim, they instead encourage learners to focus on acquiring knowledge about how the task works. This involves constructing and testing hypotheses about task rules in rule space to develop accurate and flexible rule knowledge that facilitates performance in a greater range of situations.

Despite these differences however, there is substantial overlap in the explanations of CLT and DST. In particular, both theories propose similar mechanisms for how specific and non-specific goals affect learning. Both for example suggest that specific goals elicit means-end strategies and, in doing so, impair learning. This has been clearly demonstrated in research from both perspectives (Burns & Vollmeyer, 2002; Mawer & Sweller, 1982; Owen & Sweller, 1985; Sweller, 1988; Vollmeyer & Burns, 1996). Further both theories agree that non-specific goals encourage greater task exploration than specific goals, which likely facilitates learning (Burns & Vollmeyer, 2002; Sweller, 1988). As both theories note, the absence of a specific goal means learners are able to focus on learning how to move from their current task state. The effects that specific and non-specific goals have on behaviour are therefore largely consistent between the theories. They differ primarily then in their explanations of how these behaviours influence learning. For CLT, it is the strategies' effect on cognitive load that affects learning, whilst for DST it is their effect on rule versus instance space search, and therefore the acquisition of rule versus instance knowledge.

The similarity of the explanations of CLT and DST for the goal free effect has notably lead CLT researchers to propose that DST is consistent with CLT (Sweller et al., 2011). Under this conception, cognitive load influences the extent of rule and instance space search. High cognitive load prevents rule space search whilst low load encourages it. This is because rule space search produces higher cognitive load than instance space search since it involves the search of both rule and instance spaces. As Sweller et al (2011) note: "limited working memory … prevents us from attending to both an instance and a rule space simultaneously" (p96). The interactive search of instance and rule spaces that characterise rule space search is therefore only possible under conditions of low cognitive load. Low cognitive load effectively encourages rule space search since learners have more cognitive resources available to focus on learning both task rules and instances. Since specific goals produce higher cognitive load than non-specific goals, they therefore inhibit rule space search and allow search only of task instances, consequently impairing rule learning. CLT and DST are therefore complementary with the former effectively subsuming the latter. Variations in rule space search are caused by differences in cognitive load.

Given the similarities between the theories, the suggestion that CLT and DST are in fact complementary is certainly plausible. Rule space search may be possible only to the extent that cognitive load is low. Whilst this has not been tested directly, there is some evidence from existing studies that provide some support for this interpretation. Geddes & Stevenson (1997) for example, found that a group they had assigned dual specific and non-specific goals during learning acquired less rule knowledge than other groups (who were assigned the goals separately). They attributed this to the higher cognitive load imposed by having two competing goals. Importantly however, the dual-group developed a comparable level of instance knowledge to the other groups, suggesting, consistent with CLT's combined approach, that only instance space search was possible under the high load conditions. Osman (2008) also suggested that high cognitive load may have prevented rule space search in explaining why those assigned specific goal had performed poorly during training. Typically in CDC tasks, performance differences between specific and non-specific goal groups emerge only in transfer since training goals are provided to specific goal groups throughout training (e.g. Burns & Vollmeyer, 2002; Vollmeyer & Burns, 1996). Since Osman's specific goal manipulations had required more memory and recall than previous studies, she argued that this greater cognitive load had prevented specific goal groups from acquiring the usual level of rule knowledge found in other studies. Cognitive load therefore again appeared to impair rule acquisition, presumably by preventing rule space search. Although neither of these studies assessed cognitive load, their interpretations are consistent with a contingent relationship between cognitive load and rule space search. Cognitive load may therefore influence the extent of rule space search.

There are however a number of counter examples that do not support the proposed interdependence of CLT and DST. Both Vollmeyer et al (1996) and Vollmeyer & Burns (2002) found that rule space search differed between specific and non-specific goal groups only after practice with both groups performing a moderate level of rule space search in early training trials. Since cognitive load should have been higher for the specific goal group, particularly early in training, the result suggests that high cognitive load did not impair rule space search during the initial stages of training. Moreover, specific goal participants in both studies were found to switch to means-end strategies after the initial training trials. Since it is unlikely cognitive load increased between early and late training, the switch to instance space search strategies seems therefore to have been independent of any changes in cognitive load. Also, as the authors noted, choosing a high load strategy over a low one seems implausible if a CLT interpretation is correct. Although cognitive load was not tested directly, the studies therefore suggest that rule space search was unrelated to cognitive load. Rule space search was high when cognitive load was likely to have been high, and fell for specific goal groups when there was no reason to suspect load was increasing. Whilst indirect, the findings suggest that cognitive load and rule space search vary independently.

Geddes & Stevenson (1997) also presented some findings that were inconsistent with cognitive load influencing rule space search in that specific goal participants were found to have developed some rule knowledge during training. Whilst this may be explained as cognitive load not being high enough to completely prevent rule space search (as it was for the dual goal group mentioned earlier), rule learning was specific only to the assigned goal. In other words, consistent with Vollmeyer et al (1996) and Burns & Vollmeyer (2002), specific goals appeared to allow rule space search sufficient only to reach the assigned goal rather than suppressing rule space search generally. The authors therefore concluded that "it is not cognitive load that accounts for poor rule learning with specific goals, but the minimal use of rule space" (p761). These findings therefore further suggest that cognitive load may not determine rule space search in all circumstances.

Wirth, Kunsting, & Leutner (2009) also reported findings inconsistent with the interdependence of rule space search and cognitive load, but from the perspective of CLT. They investigated the influence of both goal specificity (specific versus non-specific) and goal type (performance versus learning) on knowledge acquisition and cognitive load outcomes in a computer-based learning environment. Consistent with previous research, and CLT, specific goal groups were anticipated to experience higher cognitive load and develop lower knowledge than non-specific goal groups. However, these differences were anticipated to be smaller for the learning goal groups since they were expected to use learning, rather than performance, strategies that relied less heavily on working memory resources. Notably however, the authors described these learning strategies almost identically to rule space search. Results were consistent with predictions for performance goal groups but not for those assigned learning goals. For performance goal groups, cognitive load was higher and knowledge lower for the specific, compared to non-specific, performance goal group, consistent with predictions. For learning goal groups however, knowledge was high in both specific and non-specific goal conditions despite the specific group demonstrating markedly higher cognitive load. The specific learning goal group therefore appeared to develop a high level of knowledge despite also experiencing high cognitive load, contrary to the predictions of CLT. The authors proposed that it was due to learning goals eliciting ‘learning strategies’ that permitted acquisition of rule knowledge despite high cognitive load. Since learning strategies were described similarly to rule space search, the results strongly suggest that rule space search was, for the specific learning goal group, independent of cognitive load. The result therefore suggests that cognitive load does not necessarily affect rule space search. This raises some doubt about the proposed consistency of CLT and DST accounts of learning.

In summary, whilst there is some support for CLT’s proposal that CLT and DST provide consistent accounts of the goal free effect, with cognitive load determining the level of rule space search, there is also compelling evidence to the contrary. Whilst in some cases rule space search appears low as a result of high cognitive load (Geddes & Stevenson, 1997; Osman, 2008), in others they appear to vary independently (Burns & Vollmeyer, 2002; Geddes & Stevenson, 1997; Vollmeyer et al., 1996; Wirth et al., 2009). Since no study has yet directly examined the relationship between cognitive load and rule space search, their independence, and, by extension, that of CLT and DST, remains unknown. The focus of the present thesis is to directly examine whether cognitive load and rule space search are independent and, in doing so, establish whether CLT and DST are independent or complementary theories of learning.

Summary

This chapter presented descriptions of CLT and DST and explained their application to the goal free effect in which specific goals have been found to produce poorer learning outcomes than non-specific goals. Whilst CLT proposes that the means-end strategies elicited by specific goals create extraneous cognitive load, and therefore impair learning, DST argues that the strategies instead focus attention on learning task instances, rather than task rules. This consequently inhibits acquisition of the rules and structures governing the task's operation and thereby impairs performance. For CLT, non-specific goals do not elicit means-end strategies and so do not impose a high cognitive load whilst for DST, non-specific goals allow learners to focus on developing rule knowledge rather than focusing on achieving a specified goal. Despite their different explanations however, the mechanisms by which both CLT and DST explain the goal free effect are similar. Both agree that specific goals elicit means-end strategies and that non-specific goals permit a greater level of task exploration. This has lead CLT to propose that the theories are in fact complementary with cognitive load determining the level of rule space search possible under a given set of conditions. Despite the plausibility of this argument, and some evidence in support, there is also significant evidence that cognitive load and rule space search are independent. Since the independence of the theories has not been tested directly, this remains an open question. The focus of the present thesis therefore is to directly examine the independence of cognitive load and rule space search. In doing so, the thesis aims to establish whether CLT and DST are independent or complementary theories of learning.

CHAPTER THREE:
RESEARCH DESIGN

As outlined in Chapter 2, the present thesis aimed to examine whether CLT and DST were complementary or independent explanations of learning. Of particular focus was the claim by CLT researchers that rule space search is dependent on low levels of cognitive load and that only instance space search is possible when cognitive load is high. To test this claim, the proposed research aimed to manipulate rule space search independently of cognitive load. If successful, results would therefore suggest some separation between CLT and DST explanations of learning. This chapter details the rationale for the proposed research and presents the hypotheses to be tested in subsequent chapters. It also describes and justifies the method selected to test the independence of cognitive load and rule space search.

Rationale & Hypotheses

The present research focused on CLT's proposition that rule space search is contingent on the level of cognitive load experienced during learning. Whilst there have been no direct investigations of this relationship, some studies have suggested that rule space can be performed in conjunction with high cognitive load (Burns & Vollmeyer, 2002; Vollmeyer et al., 1996; Wirth et al., 2009) contrary to CLT's proposition. Other research has, in part, suggested some association between higher rule space search and lower cognitive load (Geddes & Stevenson, 1997; Osman, 2008). Given the absence of direct investigations and the ambiguity of existing evidence, there is therefore a need to empirically examine the relationship between cognitive load and rule space search during learning.

To test whether cognitive load and rule space search can vary independently, the present research attempted to experimentally manipulate rule space search independently of cognitive load during learning of a complex task. If variations in rule space search could be demonstrated without the anticipated changes in cognitive load, this would suggest, contrary to the proposition of CLT, that rule space search does not depend on the level of cognitive load experienced during learning. If independent variability in cognitive load and rule space could not be demonstrated, this would support the contingency proposed by CLT researchers[3].

Rule space search was manipulated in two ways. First, it was varied whilst holding cognitive load constant. If no clear changes in cognitive load were observed despite pronounced changes in rule space search (and therefore rule learning), results would suggest that both were independent. Second, rule space search was manipulated in the opposite direction to cognitive load such that high rule space search was anticipated under conditions of high cognitive load and low rule space search manipulations under conditions of low cognitive load. This was a stronger test of the independence of cognitive load and rule space search since it involved manipulation of both variables in the opposite directions to those predicted by CLT.

Hypotheses were based on the indirect evidence, presented in Chapter 2, that rule space search and cognitive load vary independently. In particular, Vollmeyer, Burns, and colleagues' (Burns & Vollmeyer, 2002; Vollmeyer et al., 1996) findings that despite likely dissimilar levels of cognitive load, both specific and non-specific goal groups demonstrated similarly high levels of rule space search during early training, suggest that rule space search was possible regardless of cognitive load. Further, their findings that specific goal participants switched from rule space search

---

[3] Notwithstanding potentially problematic interpretation of the null hypothesis.

strategies to higher load, means-end, strategies with practice, despite little reason to suspect a change in load, also suggests that cognitive load and rule space search can vary independently. Wirth and colleagues' (Wirth et al., 2009) findings that participants under high cognitive load developed high levels of rule knowledge, likely due to rule space search, also suggest independence. By comparison, counter evidence that rule space search only occurred under conditions of low cognitive load (Geddes & Stevenson, 1997; Osman, 2008) appeared less convincing since both papers suggested a DST interpretation over CLT overall. The general hypotheses in the present investigation were therefore based on previous, albeit indirect, evidence that rule space search and cognitive load were independent influences on rule learning in complex settings.

The general research hypothesis was therefore that rule space search would vary independently from cognitive load. More specifically it was anticipated that:

1. *Rule space search will predict learning outcomes independently of the level of cognitive load and (relatedly)*
2. *Rule space search will vary independently of the level of cognitive load such that high (low) levels of rule space search will be observed under conditions of high (low) cognitive load.*

Method

Justification

Since both CLT and DST derive from early investigations of complex problem solving, complex problem-solving tasks have generally been employed to test their predictions. Such tasks consist of an opaque system where learners have to conduct some analysis, exploration, or translation of information presented to determine how to perform the task and/or reach a defined goal. Participants are generally provided a short training period to allow development of the necessary knowledge to perform the task before knowledge or performance tests are administered. The amount of knowledge acquired in this period is the dependent measure in analyses. The goal free effect has generally been demonstrated in these types of tasks.

One limitation of problem-solving tasks and designs however is that they focus on the amount of task knowledge that can be acquired in a set period of time rather than the maximal level of knowledge a learner could acquire given sufficient practice. Whilst this limitation is largely irrelevant for educational settings where the time allocated for instruction or learning is fixed (or where faster knowledge acquisition is substantially more advantageous), this methodology does not address the influence of training manipulations for long term performance. Differences in training conditions may for example, persist or, conversely, diminish with increasing task practice.

For the present investigation, examining the maximal level of knowledge acquisition possible given an extended period of practice offered an additional means of distinguishing between the CLT and DST. CLT suggests that cognitive load diminishes with practice since novel, high element-interactivity incoming information will be converted into schemas stored in long term memory with increasing task experience, progressively reducing the processing demands on working memory. Commensurately, the cognitive load imposed by provision of a specific goal should therefore ameliorate with continued task practice and so too any knowledge differences caused by differences in load. For DST however, if task conditions continue to encourage instance over rule space search, then differences in rule knowledge should not diminish, regardless of the amount of task practice. DST therefore implies that knowledge differences caused by the provision of specific and non-specific goals will persist over time. Adopting a task in which participants are trained to a maximal level of knowledge was therefore adopted for the present research to provide an additional means to discriminate between the predictions of CLT and DST.

Since problem-solving tasks are rarely used to train research participants to asymptotic levels of performance, a skill acquisition task was selected for the present study. Skill acquisition research is concerned primarily not with the initial acquisition of knowledge but the gradual increase in the accuracy and speed of performance with practice (e.g. Anderson, 1982; Anderson, 1987; Logan, 1988). Participants are therefore typically provided extended periods of task practice to ensure they reach highly proficient levels of performance. The advantage of using a task from this domain is that both the process and duration of training required to reach asymptote is known in advance. Training protocols can therefore be designed around these characteristics without first having to determine how practice is likely to progress.

The task selected for the current research was the Kanfer-Ackerman Air Traffic Control (KA-ATC) task. This task was developed by Ackerman (1988) but employed most prominently by Kanfer & Ackerman (1989). This task was selected above other skill acquisition tasks since it had previously been used to examine the goal free effect and was likely to exhibit clear associations with measures of both cognitive load and rule space search. It was therefore a suitable medium for testing the research hypotheses.

The following section describes previous research that has used the KA-ATC task to demonstrate the task's suitability for the present research. The section thereafter provides a detailed description of the task.

Application to research questions

Since CLT and DST differ most clearly in their explanations of the goal free effect, the task selected to address hypotheses had to demonstrate both susceptibility to the effect and permit measurement of the influence of cognitive load and rule space search. Whilst these requirements are met by a number of the problem solving tasks employed by CLT and DST researchers (e.g. Burns & Vollmeyer, 2002; Miller, Lehman, & Koedinger, 1999; Paas, Camp, & Rikers, 2001; Wirth et al., 2009), these were not considered optimal since they were problem solving as opposed to skill acquisition tasks. The KA-ATC task however met each of these requirements as well as being a skill acquisition task.

The goal free effect has been demonstrated using the KA-ATC task in two previous studies. In the first, Kanfer & Ackerman (1989) assigned participants specific or non-specific goals either early or late in their task training. Specific goals were found to elicit poorer learning outcomes when assigned early in training, particularly for low cognitive ability participants[4]. In the second, Kanfer & colleagues (Kanfer & Ackerman, 1996; Kanfer, Ackerman, Murtha, & Dugdale, 1994) assigned specific or non-specific goals in combination with either massed or spaced practice conditions (differentiated by either a very short or very long inter-trial interval respectively). A goal free effect was also reported for participants trained under

---

[4] Kanfer & Ackerman's (1988) first experiment found no effect of early goal assignment. Early goal assignment was found only to disrupt the otherwise anticipated correlations between performance and ability measures. In their third experiment however, a clear goal free effect was reported.

massed practice conditions. The results therefore demonstrate that the goal free effect is demonstrable in the KA-ATC task.

The KA-ATC task also provides a clear means of measuring the influence of cognitive load on the goal free effect. Working memory capacity has been consistently associated with performance on the KA-ATC (Ackerman, 1988; Kanfer & Ackerman, 1989; Schunn & Reder, 2001). Further, these associations are largely consistent with those anticipated of cognitive load by CLT. For example, higher correlations between working memory and performance have been observed in early, as opposed to late, training (Ackerman, 1988) and the negative influence of specific goals has been observed predominately when working memory–performance correlations have been high (Kanfer & Ackerman, 1989). Since cognitive load is synonymous with the processing demands placed on working memory (Sweller et al., 2011; Sweller, van Merrienboer, & Paas, 1998), and since the pattern of correlations between working memory and performance have been consistent with those predicted of cognitive load, working memory therefore represents a clear means of assessing cognitive load on the KA-ATC task[5]. The task therefore permits investigation of the influence of cognitive load on the goal free effect via measures of working memory capacity.

Whilst the precise way in which the KA-ATC task permits observation of rule space search will be described in the following section, one previous task

---

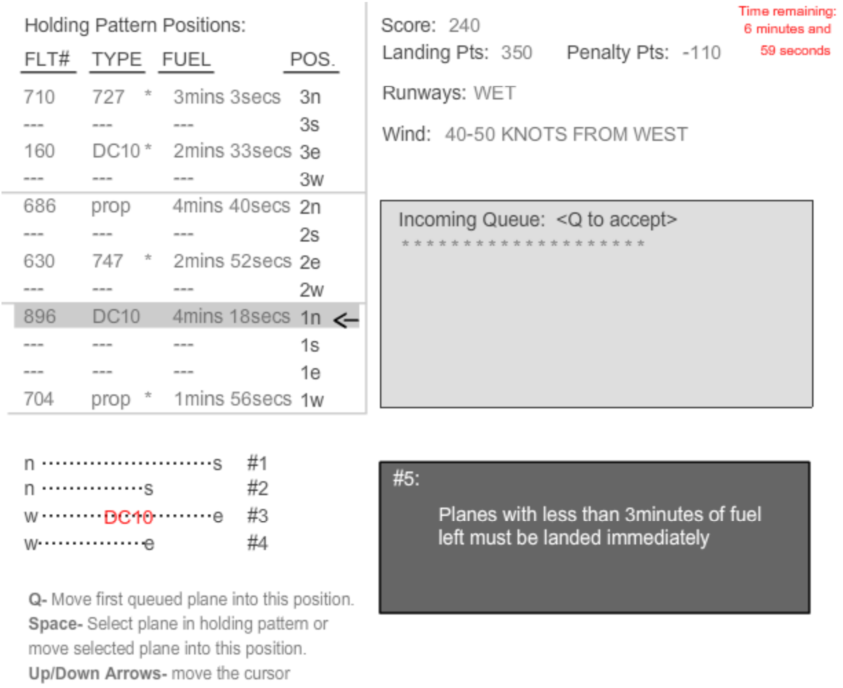
[5] Cognitive load has traditionally been measured using self-rating scales of mental effort (Paas, 1992; Paas, Tuovinen, Tabbers, & Van Gerven, 2003) and/or mental efficiency (Paas & Van Merrienboer, 1993). Although ostensibly measuring perceived cognitive load, such scales are considered a valid proxy for the construct (Sweller et al, 2011). However, since working memory measures are task independent and objective, unlike perceived measures, they were considered preferable. A secondary task could also have been used to assess load through given the high attentional demands of the KA-ATC task, doing so may have impaired learning and performance (e.g. Halford, Maybery, & Bain, 1986). Working memory capacity therefore appeared a more suitable means of assessing cognitive load for the present purposes.

manipulation has suggested an influence of rule space search on the goal free effect. Kanfer & Ackerman (1989) pre trained participants using either ‘declarative’ or ‘procedural’ part-task training conditions to influence their rule knowledge prior to commencing task performance. ‘Declarative training’ conditions instructed participants to “learn the rules of the…task” (p679) by developing and testing hypotheses. ‘Procedural training’ conditions instead instructed participants to learn only the key press sequences required to perform the task by following a list of key presses and observing the outcome. The declarative condition therefore strongly encouraged rule space search by explicitly instructing participants to focus on rule learning and hypothesis testing whilst the procedural condition encouraged instance space search by directing attention to task states and outcomes only. Participants were then asked to perform the KA-ATC task under either specific or non-specific goal conditions. Consistent with the predictions of both the authors and DST, the procedural-trained group demonstrated persistently poor performance under specific goal conditions whilst declarative-trained participants did not. That is, consistent with DST, the greater rule knowledge acquired by the declarative-trained group appeared to facilitate performance under specific goal conditions in a similar way to previous DST studies (Burns & Vollmeyer, 2002; Vollmeyer et al., 1996). Whilst the result is confounded by the reduction in cognitive load caused by an increase in task knowledge, the manipulation was almost certainly one of rule and instance space search. This result indicates that rule space search likely informs the development of knowledge and performance in the KA-ATC task. The task is therefore suitable for investigating the relative influence of both rule space search and cognitive load on learning.

Description: Kanfer-Ackerman Air Traffic Control (KA-ATC) Task

The KA-ATC task is a rule-based, real-time, computer simulation of some of the tasks carried out by an air traffic controller. The three main sub-tasks required are accepting incoming planes from other airports into the holding pattern, moving planes within this holding pattern, and clearing planes for landing on one of the runways. The aim of the task is to accumulate as many points as possible by landing as many planes as possible and by making as few mistakes (i.e. violating task rules) as possible within each 10 minute trial. A typical task display is shown in *Figure 1* below.

As can be seen from *Figure 1*, there are multiple sections in the display (from bottom left, clockwise): response keys needed to play the game; four runways, two long, two short, two facing North-South, two facing East-West; 12 holding pattern positions divided into three levels corresponding to altitude (level three is highest, level one is lowest); current score including both penalty and landing points; time remaining in the trial (min/sec); current weather conditions, including condition of runways (dry, wet, or icy) and wind speed (0-20, 25-35, above 40-50 knots) and wind direction (N, E, S, W); incoming queue of planes waiting to be accepted into the holding pattern (each plane in the queue is represented by an asterisk); error message box that appears if a mistake is made or if rules are requested by the participant.



Holding Pattern Positions:

FLT#	TYPE	FUEL	POS.
710	727 *	3mins 3secs	3n
---	---	---	3s
160	DC10 *	2mins 33secs	3e
---	---	---	3w
686	prop	4mins 40secs	2n
---	---	---	2s
630	747 *	2mins 52secs	2e
---	---	---	2w
896	DC10	4mins 18secs	1n ←
---	---	---	1s
---	---	---	1e
704	prop *	1mins 56secs	1w

Score: 240
Landing Pts: 350 Penalty Pts: -110
Time remaining: 6 minutes and 59 seconds
Runways: WET
Wind: 40-50 KNOTS FROM WEST

Incoming Queue: <Q to accept>
* * * * * * * * * * * * * * * * * * * *

n ························s #1
n ···············s #2
w ·········DC10·········e #3
w···············e #4

#5:
Planes with less than 3minutes of fuel left must be landed immediately

Q- Move first queued plane into this position.
Space- Select plane in holding pattern or move selected plane into this position.
Up/Down Arrows- move the cursor

Figure 1: The KA-ATC task display.

To successfully perform the task, participants must continuously accept planes from the Queue into the Holding Pattern, move planes in the Holding Pattern to level 1, and land the planes from level 1 on one of the four Runways. To accept a plane from the Queue into the Holding Pattern, participants move the cursor arrow (shown next to Holding Pattern Level ‘1n’ in Figure 1), by pressing the ‘Up’ or ‘Down’ arrows, next to a vacant Holding Pattern position and then press ‘Q’. A plane will then appear in this position and one asterisk will be removed from the Queue box. To move a plane, participants move the cursor next to a plane in the holding pattern and then press ‘Space’ to select it. Once selected, a plane is highlighted grey as shown in Figure 1. Participants then move the cursor adjacent to a vacant position and press ‘Space’ to move the selected plane into this position. To land a plane, participants perform the same sequence as moving a plane but instead move the selected plane to a vacant runway. There is a one-to-one mapping of key responses to cursor movements and changes on screen such that for every key press a participant makes, a corresponding change occurs onscreen.

Planes were added to the Queue every 7 sec and remained in the queue until accepted into the Holding Pattern. Once accepted, additional detail about the plane was shown including plane type and fuel remaining (see Figure 1). There were four types of planes in the task: 747s, 727s, DC10s and Props (propeller aircraft). Minutes of fuel remaining was calculated randomly once a plane was added to the holding pattern and ranged from four to six minutes. Fuel then counted down in real time. Weather conditions changed every 15 seconds with one condition (runway conditions, wind direction, wind speed) changing randomly each time.

Rules. Six rules governed performance on the task (see Table 1 below). Any violation of the rules resulted in an error message being displayed (10 sec) and a 10-point deduction from a participant's score. Error messages are shown in the Rule column of Table 1. Rule details (shown in the Detail column) were presented to participants during instructions and again at the commencement of each trial. Rule-4 details were only shown to the two Full Rule groups.

Table 1: Rules governing performance on the KA-ATC task.

The rule column displays the message shown when a mistake was made; the detail column displays the rule information presented during task instructions.

Number	**Rule**	**Detail**
1	Planes must land into the wind	If the wind is coming from the East or West, the East-West facing runways (#3 and #4) must be used.
2	Planes can only land from hold level 1	Planes can only be cleared for landing once they are in the lowest hold level, level 1. Planes in levels 2 or 3 cannot be cleared for landing.
3	Planes in the holding pattern can only move one level at a time	Planes can only move from level 3 to 2 or from 2 to 1. Planes cannot be moved directly from level 3 to 1.
4	Weather conditions determine the runway length required by different plane types (error message: 'can use short runways when: 747- Never, Prop- Always, DC10- Not icy & not 40-50 knots, 727- dry or 0-20 knots')	- 747s: can only use long runways. - 727s: can use short runways when dry or when wind is below 40 knots. - DC10s: can use short runways when dry or wet and wind below 40 knots. - Props: can always use short runways.
5	Planes with less than 3 minutes of fuel must be landed immediately	When fuel remaining is 3 minutes or less, plane must be landed immediately.
6	Only one plane at a time can occupy a runway	A runway must be clear before another plane can be cleared for landing.

Rule-4 was of particular importance for the present thesis. The other five rules described relatively simple, non-contingent, constraints on performance. Rule-4 however was contingent on plane type and, in the cases of 727s and DC10s, also on weather conditions. Landing a 747 or Prop was relatively straightforward: regardless of weather conditions, 747s could only use long runways and Props could use either runway. Similarly, landing a 727 or DC10 on a long runway could be accomplished regardless of weather conditions. However, landing either a 727 or a DC10 on a short runway required participants to consider both the wind speed and runway conditions. 727s were governed by a disjunctive rule whereby they could use the short runway when winds were 0-20 knots or when runways were not icy. DC10s were governed by a conjunctive rule whereby they could use the short runways when they were not icy and winds were below 40 knots. Since performance was primarily dependent on the number of planes landed, it was advantageous to use both long and short runways simultaneously. Knowledge about when 727s and DC10s could use the short runways would therefore benefit performance.

Error correction. Participants were able to view each of the rules of the game during the trial by pressing the number key corresponding to the rule (e.g. by pressing key “1” Rule-1 would appear in the error messages box). No points were deducted for calling up rules during play.

Scoring. Participants received 50 points for every plane landed, -10 points for every error made and -100 points for every plane that ran out of fuel and crashed. Cumulative landing, penalty and total point scores were displayed to participants during play as shown in Figure 1. At the completion of each trial, a screen displaying total, landing and penalty score was also displayed. Whilst participants’ score and key press data was recorded, only three measures, assessed at the conclusion of each trial were included in the analysis.

Dependent Measures. Although multiple measures of performance were recorded for each trial, of particular importance were plane landings (Landings), Rule-4 errors, and short runway landings of mid-sized 727s and DC10s (OpShort). These measures have been demonstrated to be valid indicators of performance in previous studies (Kanfer & Ackerman, 1989; Kanfer et al., 1994; Schunn & Reder, 2001) and represented a specific aspect of task performance.

Landings represent the number of planes landed in any given trial. They are the standard dependent measure of KA-ATC task performance (Kanfer & Ackerman, 1989; Kanfer et al., 1994). Since landing planes is both the expressed aim of the task and the only means of increasing score, this outcome is likely to be the predominant focus of participants when performing the task.

Successful landings depend on knowledge of task rules. The more comprehensive and automated rule knowledge is, the more planes a participant can land each trial. For example, if a participant has comprehensive knowledge of Rule-4, they will be able to make greater use of the short runways to land planes and consequently increase the overall number of planes landed each trial. Plane landings are therefore taken to indicate overall task proficiency. Higher Landing scores will be taken to suggest a more complete and automated knowledge of task rules.

Rule-4 errors indicate an incorrect attempted to land a plane on a short runway. In other words, they represent a violation of Rule-4. As described above, Rule-4 is the most complex of the six rules governing task performance because it involves contingent relationships between plane types and weather. Any differences in task rule knowledge between participants will therefore likely manifest most strongly in Rule-4 knowledge. Rule-4 errors are therefore posited to be a more sensitive index of rule knowledge than overall landings performance.

Rule-4 errors will be interpreted in two ways depending on the task conditions. Under some task manipulations presented in this thesis, participants were required to develop their own rule knowledge from direct experience with the task. In other words, these conditions encouraged rule space search. In these cases, attempting short runway landings and observing the consequences was encouraged to assist participants develop Rule-4 knowledge. Rule-4 errors were not penalised the usual -10 points under these conditions to encourage rule space search. For these conditions then, Rule-4 errors were considered indicative of rule space search. This interpretation is relevant only to Studies 2 and 3 reported in Chapters 5 and 6.

In all other conditions, Rule-4 information was provided to participants during instructions and sometimes also between each trial. In these cases, Rule-4 errors were indicative of participants not having learnt the rule from the presented information. In these cases, Rule-4 errors were penalised the usual -10 points. Under these conditions, Rule-4 errors were considered indicative of a lack of Rule-4 knowledge.

OpShort indicate the number of mid-sized planes (727s and DC10s) landed on short runways. The measure was originally developed by Schunn & Reder (Schunn & Reder, 2001) as a measure of strategy use during task performance. Schunn & Reders' version however counted only OpShort landings when both long and short runways were available. The measure as used here considers all OpShort landings regardless of the availability of the long runway at the time of the short runway landing.

OpShort landings are inherently advantageous in the KA-ATC task since they increase the availability of the long runways for the planes unable to use the short runways (i.e. 747s or DC10s/727s depending on the weather). Increasing OpShort is therefore a means of increasing the use of both long and short runways and therefore overall performance. Successful OpShort also depend exclusively on Rule-4 knowledge (or possibly luck). As described above, this is the rule where knowledge is most likely to differ between participants. Since OpShort landings are advantageous and dependent on rule knowledge, they are taken to be a strong indicator of rule knowledge. Higher OpShort landings were considered indicative of better Rule-4 knowledge.

Summary

This chapter presented the rationale, hypotheses, and method for the proposed research. Since CLT claims that rule space search is contingent on the level of cognitive load, such that rule space search decreases as cognitive load increases, the general research approach was to demonstrate variation in rule space search independent of cognitive load. Specifically, experimental conditions were predicted to elicit different levels of rule space search, and consequently rule learning, under conditions of equivalent cognitive load as well as high (low) levels of rule space search under high (low) levels of cognitive load. The task selected to test these predictions was the Kanfer-Ackerman Air Traffic Control (KA-ATC) task (Kanfer & Ackerman, 1989). This task has demonstrated sensitivity to the goal free effect and to the influences of cognitive load and rule space search manipulations but, unlike many problem-solving tasks, also permits examination of the maximal level of knowledge acquired over an extended period of practice. The task was therefore suitable to the proposed investigation. In summary then, the present research aims to demonstrate that cognitive load and rule space search exert independent influences on learning and performance outcomes in the KA-ATC task.

CHAPTER FOUR

STUDY 1: PELIMINARY INVESTIGATION OF THE RELATIONSHIP BETWEEN RULE SPACE SEARCH AND COGNITIVE LOAD

As explained in the preceding chapters, CLT proposes that rule space search is only possible under conditions of low cognitive load. This is because acquiring rule knowledge in conjunction with learning to perform a task places additional processing demands on working memory that may, if cognitive load is already high, impair overall learning. Any attempts to increase rule learning under high load conditions are therefore anticipated to be either ineffective or counterproductive. The present research sought to test this proposition by administering a basic manipulation of rule space search under conditions of equivalently high cognitive load. Learners were either encouraged to, or discouraged from, acquiring rule knowledge whilst learning to perform the high-load experimental task to the best of their ability. Task element interactivity was carefully matched between the conditions to ensure both groups experienced equivalent cognitive load whilst learning. The study therefore addressed the first hypothesis of the proposed research: that differences in rule space search, and consequently rule learning, can be achieved under conditions of equivalent cognitive load. This was the first step in examining the independence of CLT and DST.

Encouraging rule space search under conditions of high cognitive load

CLT implies that learning necessarily imposes cognitive load. This is because integrating incoming information from the environment into schemas stored in long-term memory requires processing in working memory (Sweller et al., 2011). Since cognitive load depends on the number of interacting elements of information to be processed in working memory at any given time, encouraging learners to attend to task rules in conjunction with learning to perform a task may elevate cognitive load. It was for this reason that CLT proposes that rule space search is possible only under conditions of low cognitive load. Simply encouraging greater rule learning may then, if cognitive load is already high, impair overall learning. Since the proposed manipulations (outlined in Chapter 3) are designed to increase rule learning by increasing rule space search under high-load task conditions, simply encouraging greater attention to task rules may impair overall learning outcomes. In other words, encouraging rule learning may actually impair learning because it imposes additional processing requirements on a learner when they are already operating at the limit of their processing capacity. Since the aim of the present thesis was to investigate whether rule space search and cognitive load were independent, the present study presented a simple manipulation of rule space search under conditions of high cognitive load to determine whether encouraging or discouraging rule learning affected learning outcomes.

As outlined in the previous chapter, learning to perform the KA-ATC task involves a high level of cognitive load. This is evidenced by repeated demonstrations of strong correlations between task performance and working memory measures in the early trials (Ackerman, 1988; Kanfer & Ackerman, 1989; Keil & Cortina, 2001). Further, specific goal manipulations have been shown to impair learning when assigned in early training, particularly for learners lower in cognitive ability (Kanfer & Ackerman, 1989). This load is imposed because the task is relatively complex and time pressured so, in early trials, learners have to convert and integrate a large amount of novel declarative knowledge into productions stored in long-term memory (Ackerman, 1987, 1988; Lee & Anderson, 2001). It is therefore possible that encouraging learners at this early stage of training to focus also on acquiring detailed knowledge of the task's rules may overload working memory and impair learning. It is therefore possible that encouraging greater rule space search during the initial stages of performance on the KA-ATC task will either impair learning, if learners attempt to search rule space, or have no effect, if their working memory limitations do not permit rule space search to be performed at all. However, as outlined in Chapters 2 and 3, studies using tasks other than the KA-ATC task have suggested that rule space search is both possible and beneficial under conditions where cognitive load is likely to have been high (Burns & Vollmeyer, 2002; Wirth et al., 2009). These findings suggest that in the KA-ATC task, encouraging rule space search in the early stages of performance may also elicit superior rule knowledge. Overall then, encouraging rule space search in the KA-ATC task may or may not facilitate rule learning. However, if it does, results will suggest that rule space search is both possible and beneficial under conditions of high cognitive load. This would be contrary to the claim by CLT researchers that rule space search depends on the level of cognitive load.

Study overview

Training

To establish the independence of cognitive load and rule space search on learning outcomes in the KA-ATC task, the present study sought to vary rule space search between two conditions whilst holding cognitive load constant. The conditions presented participants with one of two simplified versions of the KA-ATC task designed to elicit equivalently high levels of cognitive load. In one version, rule space search was encouraged whilst in the other it was discouraged. Rule learning was anticipated to be higher in the high rule space search condition despite both conditions showing equivalently high levels of cognitive load.

The two task versions differed only in the type of weather conditions presented during performance of the KA-ATC task. Each presented only four of the possible nine wind and runway-condition combinations available. The weather conditions selected were designed to encourage or discourage rule space search by limiting or permitting OpShort landings (i.e. landings of mid-sized planes, 727s and DS10s, on short runways). As explained in Chapter 3, OpShort landings are advantageous because they facilitate more efficient runway use and greater performance overall. All participants were therefore anticipated to attempt a high number of OpShort landings each trial. Limiting OpShort landings meant attempts would be largely unsuccessful (i.e. Rule-4 error) thus prompting a revision of task rules to improve subsequent chances of success. Not limiting OpShort landings would produce high levels of success and thus prompt no revision of task rules to improve performance. Limiting OpShort therefore encouraged rule space search whilst permitting OpShort discouraged rule space search.

These particular manipulations of rule space search were devised specifically to elicit equivalent cognitive load whilst also manipulating rule space search. According to CLT, cognitive load is determined by the element interactivity of the information required to be processed in working memory for understanding to be achieved (Sweller et al., 2011). Equating the cognitive load between two task conditions therefore requires that the element interactivity of each are the same. For example one manipulation cannot contain more information in instructions, additional plane types, or more complex rules relative to the other if cognitive load is to be held constant. To achieve this, both high and low rule space search versions contained the same number of weather conditions and the same interactivity between the weather conditions and plane landings. More specifically, in the high rule space search condition, OpShort landings were permitted in only one of the four weather conditions presented, and only for DC10 aircraft. In the low rule space search condition, OpShort landings were prevented in the same single weather condition and also only for DC10 aircraft. The element interactivity of both conditions was therefore consistent. Moreover, rule knowledge was equally irrelevant to both high and low rule space conditions, being relevant only to DC10 aircraft and only in one of the four weather conditions presented. Cognitive load between the conditions was therefore anticipated to be equal whilst rule space search differed.

To summarise, the high rule space search group was anticipated to develop greater rule knowledge than the low rule space search group despite both demonstrating equivalent cognitive load.

Transfer

To assess the level of knowledge learners had acquired during training under high and low rule space search conditions, a transfer task was administered following training. This approach is commonly used to assess rule knowledge since it is more applicable to a broader range of situations than knowledge of only task instances (e.g. Gick & Holyoak, 1983; Holyoak, 2005; Kimball & Holyoak, 2000; Vollmeyer et al., 1996). Whilst typically a single transfer task is used for all groups to provide a consistent basis to compare groups, in the present research groups performed separate transfer tasks. This was done to provide a more sensitive test of knowledge differences between the groups than was possible under a single transfer condition[6].

Separate transfer conditions provided a better opportunity to detect knowledge differences between the groups due to the way in which rule knowledge could be demonstrated in transfer. Participants were anticipated to differ predominately in their knowledge of Rule-4 since training conditions mainly affected the application of this rule (i.e. OpShort). To demonstrate this knowledge, rather than simply knowledge of the instances in which the rule applied, participants had to use it in a novel situation. Due to the different training conditions however, novel situations were different for each group. A novel situation for the high rule search group was one in which OpShort landings were permissible since they were largely non-permissible for the group during training. However, for the low rule space search group, a novel situation was one in which OpShort landings were not permissible, since they had been permissible for the group during training. In other words, demonstrating transferable rule-4 knowledge for the high rule space search group required performing OpShort landings whilst for the low rule space search group it required avoiding OpShort landings (or, more specifically, avoiding Rule-4 errors). Demonstrating Rule-4 knowledge therefore depended, for either group, on the number of opportunities the transfer task provided for either performing or avoiding OpShort.

If a single transfer task was provided to both groups during transfer, there was concern that there would be either too few or an unequal number of opportunities to demonstrate rule knowledge. If for example, the standard 9-weather condition version of the KA-ATC task was provided to participants, only half of the conditions in any trial would provide an opportunity to demonstrate rule knowledge (as shown in *Table*

---

[6] In hindsight, this was an unnecessary step that overcomplicated the transfer manipulation and interpretation of results. A single transfer task would have been simpler and would, most likely, have found similar results. A single transfer task was used in subsequent studies.

3 below). Since participants were unlikely to make use of every one of these opportunities, the task may not have provided a sufficiently sensitive test of rule knowledge between the groups. Alternatively, if a different simplified version of the task was used as a single transfer task (for example the upper-right or lower left quadrant of Table 3), opportunities for performing or avoiding OpShort landings would not have been equal between the groups. This would not have permitted equal measurement of rule knowledge between the groups.

The proposed solution to these concerns was to administer the task opposite to that which they performed during training. That is, the high rule space search group was provided the task version in which OpShort landings were readily permitted and the low rule focus group was provided the task version in which OpShort landings were rarely permitted. In both cases, three out of the four weather conditions presented to each of the groups were novel and therefore involved a novel application of rule knowledge to the task. The two separate tasks therefore provided a sufficiently sensitive and equivalent test of rule knowledge in transfer. Moreover, as an additional benefit, both tasks should elicit the same cognitive load as they had in training.

Although the transfer conditions were more suitable to test knowledge between the conditions because each task was different, knowledge could not be directly compared between them. Since the transfer tasks were the same as the opposite group's training task, transfer performance was compared to the opposite group's performance at the end of training. Transfer performance of one group was therefore compared to the end-training performance of the other under like task conditions. Since knowledge was anticipated to be at a maximum at the end of training, immediately preceding transfer, knowledge levels between the two groups should have also been similar.

Summary

The present research sought to distinguish between the explanations of DST and CLT in learning a skill acquisition task. By training participants under conditions of equivalent cognitive load, but differing in levels of rule space search, it was anticipated that rule space search could be shown to influence rule learning independently of cognitive load. Specifically, it was hypothesised that those who had experienced high rule space search manipulations during training would acquire better task knowledge than those who had experienced low rule space search manipulations despite both experiencing similarly high levels of cognitive load.

To assess knowledge, two separate transfer tasks were administered following training. These tasks were the same as those administered to the opposite group during training. This was done to provide a more sensitive and unbiased assessment of rule knowledge in transfer. However, since the transfer tasks were different, knowledge differences between the groups were assessed by comparing transfer performance to the end-training performance of the opposing group. Comparisons were therefore made between like-task conditions.

Method

Participants

67 first year Psychology students (74.2% female) aged between 18 and 37 from the University of Sydney participated in the study for course credit. Three participants were excluded from analysis for not completing the study and one was excluded for obtaining task performance scores lower than 2.5 standard deviations below their group mean in every task trial, suggesting they were not taking performance seriously. A further 12 participants did not complete either the final

training or final transfer trial (or both) because they proceeded through trials too slowly, exceeding time allocated for the experiment. These participants were included in all analyses except those that involved Trial 7 and/or Trial 10 data. A final sample of 55 was considered for analyses involving Trial 7 or 10 data and a sample of 63 was considered for all other analyses.

Apparatus

Participants performed all tasks on Intel Core 2 Duo PCs with 17-inch VGA monitors in groups of no more than 8. Participants were not able to view each other's screens during the experiment. The tasks were hosted on a university server that participants accessed using the Mozilla Firefox web browser. Instructions were presented to all participants using a projected Microsoft PowerPoint presentation. Verbal instructions were read from a script. The experimental task was programmed in Adobe Flash CS3.

Measures

Dependent Measures. As outlined in Chapter 3, multiple measures of performance were recorded for each trial. Landings were considered indicative of overall task proficiency, Rule-4 errors a lack of Rule-4 knowledge, and OpShort, a direct indicator of Rule-4 knowledge. Since participants were anticipated to differ most markedly in their understanding of when to land planes on the short runways (i.e. Rule-4), both Rule-4 errors and OpShort were of particular relevance. In particular, since Rule-4 was made continuously available to participants both before and during task performance, any Rule-4 errors were considered a strong indication that participants had failed to acquire rule knowledge due to a lack of focus on acquiring rule knowledge.

Cognitive Ability. To assess the cognitive demands of the task as participants progressed though the phases of skill acquisition, three measures of cognitive ability were administered. Consistent with Ackerman's research on the role of general cognitive abilities in early performance (Ackerman, 1986, 1990, 1992), two measures of general ability were first administered. First, prior to completing any other task, participants completed a computer-based 36-item complete form Raven's Advanced Progressive Matrices (APM) test (Raven, Raven, & Court, 1994). The APM is a commonly used measure of fluid abilities and general intelligence (Bors & Stokes, 1998; Carroll, 1993; Kane & Engle, 2002; Winfred & Woehr, 1993). A 45-minute time limit was imposed in order to keep testing time to a minimum and limit the possibility ceiling effects (Jaeggi, Studer-Luethi, Buschkuehl, Su, Jonides, & Perrig, 2010). The test possesses minimal cultural bias (Murphy & Davidshofer, 1998) and was designed to test the top 25% of the population, so is appropriate for use on undergraduate university students (Bors & Stokes, 1998).

The second general ability measure, the Noughts and Crosses working memory task (Mackintosh & Bennett, 2003), was administered at the beginning of the second testing session. The task is an entirely visuo-spatial complex working memory span measure and was selected to match the modality of the ATC task to improve correlations relative to a standard, non visuo-spatial, measure (Ackerman & Cianciolo, 2002; Carroll, 1993). The task presents participants with a game of noughts and crosses, shown one round at a time. Each trial, participants are first shown a slide depicting a blank 3 x 3 cell grid. Following this, participants are shown three to four slides each depicting the same grid but with the addition of one nought (O) and one cross (X), representing moves of the game. Each slide is shown for 1 second with a .2 second interval between each slide. At the completion of each trial, the preceding slides will have shown a winning ‘three-in-a-row” combination for either the noughts or crosses. Participants are asked to mark the winning combination (both location and whether winning line was for noughts or crosses) using their mouse on an on-screen response grid. A complete description of the task can be found in (Mackintosh & Bennett, 2003).

Design

As stated, the study employed a two-group design in which groups were trained under conditions of either high or low rule space search. Rule space search was manipulated by presenting each group with a different set of weather conditions. For the training phase, the high rule search group experienced only poor weather conditions, as shown in Table 2. These conditions comprised of the two wind and runway conditions that were least permissive of short runway landings to limit the group’s ability to land a high number of planes. In fact the only OpShort landing that was possible under these conditions was that of a DC10 when conditions were wet and moderately windy, as shown in Table 3. These conditions would comprise on average only 25% of a trial. By limiting the frequency with which the high rule search group could utilise the short runway, and thus land planes overall, it was anticipated that the group would turn their attention to learning the task rules to attempt to overcome this limitation. Limiting the performance of the high rule search group was therefore anticipated to increase their focus on learning task rules, and therefore their task and rule knowledge.

Table 2: Weather conditions presented to the low and high rule search groups during the training phase.
For the low rule search group, winds were always below 35 knots and runways never icy. For the high rule search group, winds were always above 25 knots and runways were never dry.

Wind/Runway conditions	Dry	Wet	Icy
0-20 knots	Low Rule Search	Low Rule Search	
25-35 knots	Low Rule Search	Low & High Rule Search	High Rule Search
Above 40 knots		High Rule Search	High Rule Search

For the low rule search group, weather conditions were the two most permissive wind and runway conditions for OpShort landings. As shown in Table 3, OpShort landings could be performed under all four wind and runway condition combinations with the only exception being that DC10s could be landed on short runways when the winds were moderate and runways wet (again a combination likely for only 25% of any trial). By allowing frequent OpShort landings, overall landings for the group were anticipated to be high, thus providing little incentive for the group to invest the cognitive resources required to learn the task rules. Under such permissive conditions, task rules would have appeared largely irrelevant to continued high and improving performance. The low rule search group was therefore anticipated to develop only a limited knowledge of task rules during training because training conditions presented little incentive to acquire rule knowledge.

Table 3: Weather conditions under which 727 and DC10 landings were possible on the short runways (OpShort).

OpShort landings of both plane types were possible in three of the four good weather conditions, and not possible in three of the four poor weather conditions. In both weather sets, when runways were wet and wind speeds between 25 and 35 knots, DC10s could be landed on the short runways.

Wind/Runway conditions	Dry	Wet	Icy
0-20 knots	727 DC10	727 DC10	727
25-35 knots	727 DC10	DC10	-
Above 40 knots	727	-	-

░ Low Rule Search
▓ High Rule Search

Differences between the conditions could not influence cognitive load if load was to be equivalent between the groups. Had one weather set elicited lower load than the other, any differences in knowledge could be attributed to cognitive load rather than rule space search differences. The two weather condition sets were therefore designed to be equivalent, as far as possible, in element interactivity. This was achieved in two ways. First, the number of interacting task elements was held constant between the two conditions. Each group received identical task instructions and tasks with the exception of the weather conditions and the number of weather conditions was the same for each group. The number of interacting task elements was therefore equivalent between the groups.

Second, the actual relevance of rule knowledge to both conditions was equivalent and low. In both conditions, task knowledge could only assist performance when conditions were moderately windy and runways were wet, a combination likely during only 25% of any trial. Under other weather combinations, plane landings were not contingent on the weather. Further, possessing complete rule knowledge would, under this weather condition, permit only the avoidance of a Rule-4 error (in the low rule search group) or the landing of one additional plane (in the high rule search group), neither of which would have contributed greatly to overall performance. The actual relevance of rule knowledge to each group was therefore consistent and fairly low. Notwithstanding the subtle difference between making a Rule-4 error and performing an OpShort landing, the number of interacting elements that needed to be processed to perform the task well was anticipated to be equivalent.

Although a simpler manipulation could have been to remove the moderately windy, wet weather condition from both sets such that OpShort landings were either always or never possible, the inclusion of one weather combination where rule knowledge was relevant to performance was deemed necessary. If rule knowledge were never relevant to performance during training, participants would have had no incentive for acquiring rule knowledge. Investing effort to learn task rules would have, under such conditions, been counterproductive. Some, albeit limited, opportunity to use rule knowledge was therefore required to ensure participants in both conditions had at least some incentive for learning the task rules.

The transfer tasks administered following training were the same tasks provided in training but for the opposing group. Comparisons of rule knowledge between the groups were therefore made between transfer trials and the final training trial of the opposite group since performance of both were under equivalent conditions.

Procedure

Participants were randomly allocated to experimental conditions according to experiment session. Participation was completed in two 1.5 hour sessions completed over two days, no more than two days apart, in groups of between three and seven. Participants began the first session with the working memory followed by task instructions. Instructions were presented in a Microsoft PowerPoint presentation projected on an in-room screen and accompanying instructions were read from a script. The instructions outlined the aims of the game, parts of the display, how to play, the six rules, and response keys. Instruction duration was approximately 5min.

Following instructions, participants were shown a demonstration of the game in which each sub-task was performed and explained in real time. Specifically, a plane (747) was added to the queue, moved down the through the holding pattern, and then landed on one of the long runways. Participants then logged into the KA-ATC task and completed 7 x 10 minute trials of the task in succession. Participants were permitted to take short breaks between trials if needed.

At the commencement of each trial an information screen was shown that displayed the response keys needed to play the game and the set of six task rules. At the completion of the 7 trials, participants left and were reminded to return at their allocated time.

Participants began the second session by completing the RPM measure and then completed the remaining three task trials. Before beginning trial 8, participants were told that the upcoming trials were "slightly different" to preceding trials but were not told of any specific changes to the task. Participants were debriefed following completion of the final task trial.

Results

Analyses

Repeated measures ANOVA and trend analyses were first conducted on each of the dependent measures in the training phase to determine whether the two sets had elicited the anticipated differences in performance and whether the groups had developed some level of rule knowledge. To determine whether groups differed in the level of rule knowledge each had acquired during training, comparisons were made between dependent measures in the final trial of training and the transfer phase overall. If both groups had acquired a good level of knowledge during training, no differences between training and transfer performance would be observed but if, as anticipated, the low rule search group acquired a lower level knowledge than the high rule search group, their transfer performance was anticipated to be lower. Comparisons were made only between like weather sets and were therefore between group rather than repeated measures comparisons. Two independent samples t-tests were planned for each dependent measure: one for the comparison between the good weather sets and the other between the poor weather sets.

To ensure that the level of cognitive load did not differ between the conditions during training or transfer, Fisher's z test comparisons were performed on the strength of correlations between ability and dependent measures. If ability measures were highly correlated as anticipated, both would be first combined to form a single fluid ability composite to simplify analyses.

Manipulation check

Correlations between each of the dependent measures were first assessed to ensure each was representative of the intended aspects of task performance. Table 3 displays the correlations between each dependent measure as well as means and standard deviations. Overall, correlations were within anticipated ranges and directions. Landings were correlated moderately and negatively (r = -.30 to -.58) with Rule-4 errors throughout training suggesting that those who made fewer errors, and had a better understanding of Rule-4 errors, also landed more planes. Correlation between Landings and OpShort were moderately to strongly correlated (r = .52 to

.74), reflecting the extent to which OpShort was a component of overall landings. These correlations were expected to differ somewhat between each task version given the relative ease with which participants could use the short runways in the high versus low rule space search conditions. Rule-4 errors and OpShort showed initial moderate (trial 1: $r = -.41$) but declining correlations (trial 7: $r = .18$) with practice indicating that initially, those who made better use of the short runways were likely to make fewer errors (suggesting a high level of Rule-4 knowledge) but as performance progressed short runway landings became less related to errors, possibly due to declining error and increasing OpShort rates with practice. The correlations between OpShort and Rule-4 errors also suggest that both measures were largely independent and their correlation with overall landings performance suggest both were valid indicators of Rule-4 knowledge.

Measures of cognitive ability, as (indirect) proxies for cognitive load, were anticipated to be highly correlated given the considerable overlap in the constructs measured by each (Ackerman, Beier, & Boyle, 2005; Carroll, 1993; Colom, Rebollo, Palacios, Juan-Espinosa, & Kyllonen, 2004; Kyllonen & Christal, 1990). It was therefore intended that APM and Noughts and Crosses measures could be combined to form a composite fluid ability measure. However, the correlation between the measures was only small to moderate (r = .25), so each measure was analysed separately. Comparisons of cognitive ability between the groups revealed that the groups were equivalent in their abilities (APM: $t(61) = -.09$, *ns*; Noughts and Crosses: $t(59) = -.25$, *ns*) meaning working memory capacity was equivalent between the groups.

Training phase.

To establish that the different weather conditions experienced by each group during training had produced differences in performance, three repeated measures ANOVAs, including associated trend analyses, were conducted on training phase dependent measures. Results of the analyses are shown in Table 4 and mean performance for each dependent variable for all 10 trials are shown in Figure 2.

Differences between the two weather sets were observed for all three dependent measures over the training phase with those trained under the good weather set showing higher landings, fewer Rule-4 errors, and higher OpShort landings overall. These results were directly consistent with the intended ease with which the low rule search group could perform landings during training. Tests of linear trends revealed that although both groups improved in all measures over the course of training, low rule space search participants improved more than high rule space search participants. Significant quadratic trends for all dependent measures indicated that performance improvements occurred at a decreasing rate for both groups during training, consistent with the power law of practice, but significant interactions for Rule-4 errors and OpShort suggested that the rates of plateau differed between the groups. Low rule space search participants demonstrated a rapid increase and plateau of OpShort whereas high rule space search participants demonstrated slow but steadily increasing rate of OpShort landings, as shown in *Figure 2*. For Rule-4 errors, low rule space search participants committed a consistently low number of errors across training whilst the high rule space search participants demonstrated an initially higher number of Rule-4 errors that decreased with practice, at an apparently decreasing rate, as shown in *Figure 2*. Overall, training phase results reflected the relative ease of landing planes and relative difficulty of committing Rule-4 errors in the low compared to high rule search conditions, consistent with expectations. Importantly, the results also indicated that both groups had reached performance plateau suggesting that learning was at or nearing a maximum by the end of training.

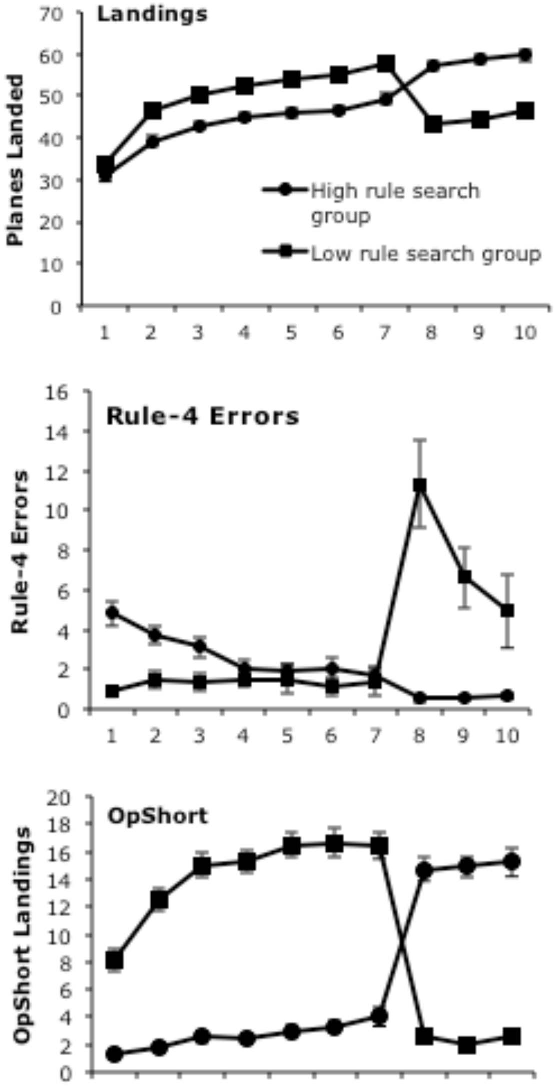


Figure 2: Mean performance during training (trials 1 to 7) and transfer (trials 8 to 10).

Landings are displayed in top panel, Rule-4 errors in the second panel, and OpShort in the bottom panel. Error bars represent +/- 1 SE from mean

Table 3: Correlations between each of the dependent measures for each trial of the study

	M	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
Landings																															
Trial 1	32.24	9.61																													
Trial 2	42.41	9.34	.81**																												
Trial 3	46.30	8.97	.75**	.90**																											
Trial 4	48.29	8.90	.79**	.87**	.91**																										
Trial 5	49.35	9.58	.78**	.89**	.90**	.92**																									
Trial 6	50.46	9.30	.72**	.86**	.89**	.88**	.90**																								
Trial 7	53.25	9.97	.71**	.85**	.85**	.86**	.86**	.89**																							
Trial 8	50.81	10.24	.37**	.26*	.22	.24	.24	.19	.23																						
Trial 9	52.15	10.10	.38**	.22	.18	.20	.20	.18	.21	.91**																					
Trial 10	53.09	10.77	.50**	.36**	.30*	.34*	.33*	.27*	.31*	.91**	.87**																				
Type-4 Errors																															
Trial 1	3.05	3.46	-.30*	-.49**	-.48**	-.41**	-.53**	-.49**	-.43**	.19	.22	.11																			
Trial 2	2.73	2.84	-.44**	-.55**	-.50**	-.48**	-.56**	-.51**	-.46**	-.03	.01	-.08	.70**																		
Trial 3	2.33	2.91	-.43**	-.50**	-.45**	-.41**	-.48**	-.45**	-.46**	-.12	-.09	-.18	.50**	.67**																	
Trial 4	1.78	2.20	-.22	-.31*	-.30*	-.31*	-.33**	-.38**	-.43**	-.17	-.23	-.24	.36**	.33**	.58**																
Trial 5	1.71	3.06	-.34**	-.35**	-.32*	-.29*	-.35**	-.26*	-.33*	-.23	-.18	-.26	.43**	.54**	.62**	.38**															
Trial 6	1.62	2.92	-.32**	-.32*	-.33**	-.38**	-.34**	-.37**	-.42**	-.07	-.15	-.27	.21	.42**	.55**	.48**	.43**														
Trial 7	1.51	2.99	-.46**	-.39**	-.38**	-.41**	-.37**	-.35**	-.35**	-.22	-.19	-.28*	.33*	.60**	.60**	.26	.73**	.75**													
Trial 8	5.30	9.44	-.19	-.01	.03	.02	.06	.11	.05	-.58**	-.52**	-.57**	-.19	.13	.23	.04	.46**	.25*	.47**												
Trial 9	3.31	6.20	-.31*	-.11	-.13	-.14	-.09	-.04	-.04	-.53**	-.54**	-.56**	-.15	.18	.26*	.04	.51**	.41**	.54**	.87**											
Trial 10	2.80	7.11	-.38**	-.24	-.21	-.26	-.22	-.15	-.18	-.43**	-.39**	-.50**	.00	.35**	.40**	.11	.70**	.60**	.72**	.82**	.88**										
OpShort																															
Trial 1	4.29	4.70	.52**	.62**	.63**	.65**	.66**	.65**	.60**	-.34**	-.34**	-.25	-.41**	-.38**	-.35**	-.14	-.13	-.24	-.20	.33**	.25*	-.02									
Trial 2	6.56	6.31	.32*	.59**	.59**	.56**	.57**	.61**	.57**	-.45**	-.49**	-.37**	-.52**	-.43**	-.36**	-.22	-.17	-.28*	-.18	.38**	.31*	.11	.79**								
Trial 3	8.10	7.25	.37**	.55**	.63**	.61**	.62**	.60**	.54**	-.46**	-.46**	-.36**	-.53**	-.42**	-.31*	-.19	-.12	-.18	-.15	.44**	.29*	.10	.84**	.88**							
Trial 4	8.14	7.29	.40**	.59**	.64**	.65**	.64**	.68**	.61**	-.43**	-.46**	-.32*	-.55**	-.47**	-.34**	-.19	-.11	-.22	-.18	.46**	.35**	.14	.83**	.89**	.89**						
Trial 5	8.98	7.83	.45**	.60**	.62**	.64**	.67**	.66**	.61**	-.40**	-.43**	-.31*	-.56**	-.48**	-.40**	-.26*	-.21	-.27*	-.23	.44**	.32*	.07	.83**	.86**	.90**	.94**					
Trial 6	9.24	7.97	.46**	.64**	.68**	.66**	.68**	.74**	.67**	-.33**	-.38**	-.26	-.52**	-.45**	-.35**	-.20	-.16	-.19	-.20	.42**	.31*	.09	.83**	.85**	.85**	.92**	.92**				
Trial 7	9.89	7.61	.39**	.55**	.60**	.54**	.57**	.62**	.59**	-.38**	-.41**	-.30*	-.54**	-.41**	-.32*	-.09	-.19	-.14	-.18	.40**	.32*	.09	.78**	.86**	.86**	.89**	.88**	.91**			
Trial 8	9.33	7.32	.10	-.09	-.09	-.13	-.08	-.16	-.08	.81**	.84**	.75**	.33**	.15	.09	.00	-.08	.13	-.04	-.48**	-.43**	-.29*	-.52**	-.66**	-.63**	-.64**	-.58**	-.56**	-.52**		
Trial 9	9.08	7.40	.02	-.18	-.21	-.23	-.21	-.26*	-.19	.76**	.82**	.71**	.47**	.24	.16	.06	.03	.05	.00	-.49**	-.42**	-.25	-.59**	-.70**	-.70**	-.72**	-.70**	-.64**	-.61**	.89**	
Trial 10	8.94	7.67	.05	-.15	-.15	-.16	-.14	-.19	-.13	.74**	.77**	.72**	.42**	.23	.20	-.03	.03	.09	.07	-.44**	-.39**	-.25	-.59**	-.66**	-.65**	-.65**	-.66**	-.57**	-.54**	.88**	.92*

Table 4: Repeated measures ANOVA and contrast (trend) analyses for each of the three dependent measures, Landings, Rule-4 errors, and OpShort, over the first seven training trials.

Variable	*df*	*MS*	*MSE*	*F*
Landings				
Within-Subjects				
Trial	6, 53	2556.93	13.01	196.50**
Trial x Condition	6, 53	44.95	13.01	3.46**
Between-Subjects				
Condition	1, 53	5233.14	480.51	10.89**
Trend Analyses				
Linear: Trial	1, 53	12593.32	24.53	513.42**
Quadratic: Trial	1, 53	1981.39	15.00	132.11**
Linear: Trial x Condition	1, 53	140.33	24.53	5.72*
Quadratic: Trial x Condition	1, 53	53.15	15.00	3.54
Type-4 Errors				
Within-Subjects				
Trial	6, 53	16.96	3.60	4.72**
Trial x Condition	6, 53	22.70	3.60	6.31**
Between-Subjects				
Condition	1, 53	1472.37	29.34	4.89*
Trend Analyses				
Linear: Trial	1, 53	86.12	6.66	12.94**
Quadratic: Trial	1, 53	7.37	2.89	2.55
Linear: Trial x Condition	1, 53	98.53	6.66	14.80**
Quadratic: Trial x Condition	1, 53	32.47	2.89	11.25**
OpShort				
Within-Subjects				
Trial	6, 53	211.70	6.42	32.98**
Trial x Condition	6, 53	71.78	6.42	11.18**
Between-Subjects				
Condition	1, 53	13864.62	66.15	209.59**
Trend Analyses				
Linear: Trial	1, 53	1044.51	10.53	99.23**
Quadratic: Trial	1, 53	189.19	5.60	33.81**
Linear: Trial x Condition	1, 53	239.55	10.53	22.76**
Quadratic: Trial x Condition	1, 53	176.38	5.60	31.52**

Note: N= 55. Seven participants did not complete trial 7 and so were not included in these analyses. Tests of sphericity were violated for all within-subjects analyses. However, given this is common for learning data, and that analyses aimed to provide manipulation checks only, 'sphericity assumed' measures are reported for simplicity. *p<.05, **p<.01.

Transfer phase.

Changes to weather conditions in the transfer phase were designed to assess the level of knowledge participants in each group had acquired relative to those in the other group. Participants' performance in the test phase was therefore compared against performance in the final trial of the acquisition phase for the opposite group. For example, the high rule space search group's transfer performance was compared to the end-training performance (Trial 7) of the low rule space group since both were performed under like-task conditions (i.e. poor weather). Trial 7 was selected as the point of comparison because it represented a maximal level of knowledge attainable for both groups and so served as a basis for evaluating transfer performance.

Table 5 displays the results of comparisons between transfer phase and Trial 7 performance measures for like weather sets. As shown, transfer phase performance for the high rule search group, was equivalent to that of the low rule search group during training for all three dependent measures. The high rule search group appeared to readily adapt to the changed transfer conditions, suggesting that the group had acquired a high level of rule knowledge.

Table 5: Results of independent samples t-tests comparing performance in the transfer trials with performance in the comparable weather condition immediately prior to transfer (trial 7).

Variable		Trial 7	Transfer	*df*	*t*
Landings					
High vs Low Rule Search	M (SD)	49.31 (6.53)	44.62 (7.91)	55	2.45*
Low vs High Rule Search	M (SD)	57.65 (11.36)	58.14 (6.43)	59	.21
Rule-4 Errors					
High vs Low Rule Search	M (SD)	1.69 (2.49)	7.85 (9.50)	55	-3.37**
Low vs High Rule Search	M (SD)	1.31 (3.50)	.65 (.85)	59	-1.07
OpShort					
High vs Low Rule Search	M (SD)	4.00 (3.61)	2.44 (2.02)	55	2.00*
Low vs High Rule Search	M (SD)	16.46 (5.05)	15.00 (4.22)	59	-1.23

*Note. *p<.05, **p<.01*

For the low rule search group, performance was significantly lower in all three dependent measures than that of the high rule search group during training. This suggests that the group did not adapt well to training, suggesting that the group did not acquire a high level of task knowledge during training. This is particularly evident in the pattern of Rule-4 errors exhibited by the low rule search group during transfer. As shown in Figure 2, the group committed a very high number of Rule-4 errors in the transfer phase suggesting that the group continued to try to land planes on the short runways despite the conditions largely preventing such landings and clear error feedback being provided in response to each error. This strongly suggests that the group lacked sufficient Rule-4 knowledge to adapt to the novel task conditions. Consistent with predictions, it would appear that training participants under low rule search conditions produced poorer learning than high rule search conditions.

Cognitive abilities.

Table 6: Correlations between the two cognitive ability measures and the three dependent measures for each trial of the study.

	Landings		Rule-4 Errors		OpShort	
	1	2	1	2	1	2
1 APM	(.68)					
2 Noughts & Crosses	.31*	(.75)				
Training						
Trial 1	.39**	.36** -	-.16	-.09	.22	.17
Trial 2	.45**	.43**	-.18	-.28*	.19	.12
Trial 3	.35**	.32*	-.19	-.13	.12	.09
Trial 4	.33**	.35**	-.15	-.19	.21	.17
Trial 5	.40**	.41**	-.20	-.20	.24	.17
Trial 6	.30*	.35**	-.04	-.30*	.19	.13
Trial 7	.36**	.29*	-.12	-.19	.15	.08
Transfer						
Trial 8	.29*	.27*	-.03	-.06	.26*	.14
Trial 9	.29*	.33**	-.06	-.09	.09	.10
Trial 10	0.22	.33*	-.06	-.28*	.08	.12

Note * p<.05, ** p<.01, reliabilities for the ability measures are displayed on the first diagonal.

Differences in learning outcomes were also anticipated to be independent of cognitive load (*Hypothesis 2)*. To test this prediction, correlations between cognitive ability measures, as (indirect) proxies for cognitive load, and performance measures were compared between the groups. Overall correlations between the two cognitive ability measures and the three dependent measures for each trial are shown in Table 6. Since only landings performance demonstrated consistently significant correlations with ability measures, only correlations between ability and landings were considered for further analyses. Interestingly, overall correlations between landings and cognitive ability were generally significant throughout the study suggesting consistently high cognitive load throughout performance (although somewhat contrary to Ackerman's (1988) resource allocation model). It may be that since the specific training conditions employed in the present research did not depend on automation to the same extent as those employed in Ackerman's research, reductions in the strength of correlations were not observed.

Table 7: Results of the three moderated regression examining whether correlations between ability measures and landings performance differed between the two study groups.

The three analyses correspond to landings performance in early and late training as well as in transfer

	Phase					
Predictor variable	Training – Trials 1 to 3		Training – Trials 4 to 7		Transfer – Trials 8 to 10	
	β	ΔR^2	β	ΔR^2	β	R^2
Step 1		.34		.39		.56
Group	.28*		.42**		-.64**	
Reasoning	.33**		.27*		.25**	
Working Memory	.28*		.30**		.28**	
Step 2		.35		.39		.56
Group x Reasoning	<.01		-.15		-.27	
Group x Working Memory	.28		.71		.33	
Overall R		.59		.66		.75
Overall R^2		.35		.44		.56
Adjusted R^2		.27		.39		.52
Overall F (5, 62)		5.99*		8.66**		14.40**

Note: * p<.05, ** p<.01,

Moderated regression analyses were conducted to determine whether the relationship between cognitive load and landings performance differed between the two groups. Since cognitive load was anticipated to be highest during the initial stage of training when participants were still acquiring the rules of the task, separate analyses were conduced for the first three training trials as well as for later training trials and transfer. As shown in Table 7, although ability measures were significant predictors of landings performance overall, the strength the relationships did not differ between the two groups at any stage of the study suggesting that cognitive load was equivalent for both groups throughout the study. Whilst this result could have been due to limited power, it does appear, as shown in Figure 3, that there was little difference in the strength of correlations between the two groups. As shown, the correlations between ability measures and landings performance were similarly high across all phases (with the possible exception of trial 7 for working memory), suggesting that load was equivalent between the groups. In support of predictions, cognitive load therefore appeared to be equivalent and high in the two groups at almost all stages of learning and transfer.

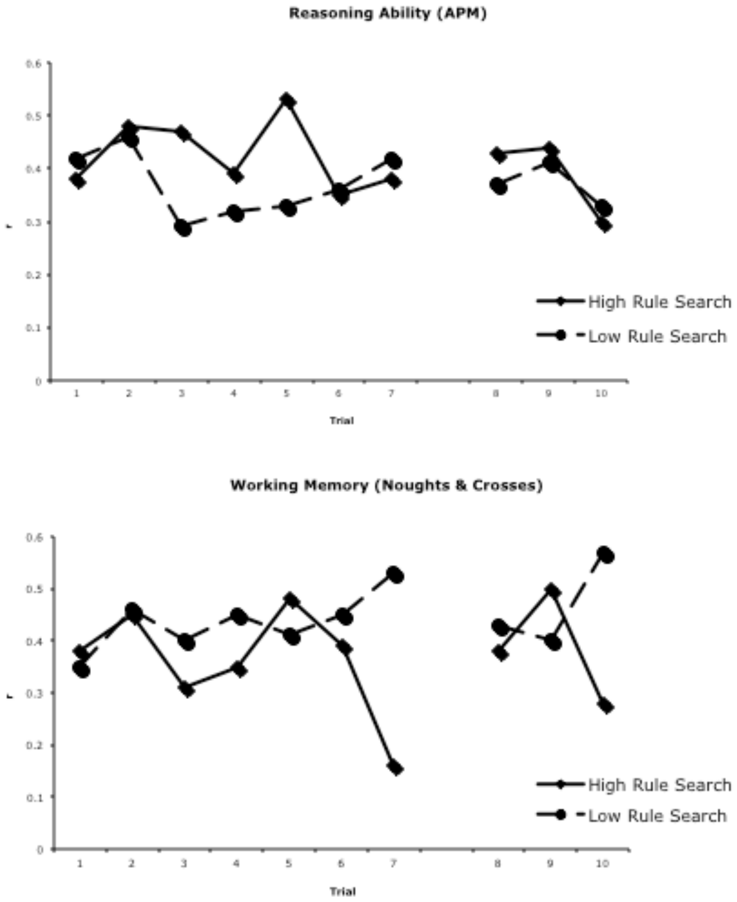


Figure 3: Correlations between landings and the two cognitive ability measures for each group.
Panels display reasoning ability (top panel), working memory (bottom panel), over the training (trials 1 to 7) and transfer (trials 8 to 10) phases. N = 50 for all correlations (n = 24 and 26 for groups 1 and 2 respectively)

Discussion

Study 1 sought to establish CLT and DST as independent theories of learning by manipulating rule space search independently of cognitive load. The study employed two versions of the KA-ATC task that encouraged or discouraged rule learning but were intended to elicit equivalent cognitive load. Participants were trained on one task version before performing the other version as a transfer task designed to assess their level of knowledge. Whilst results were generally supportive of the hypotheses, suggesting the independence of CLT and DST, the way in which rule space search was manipulated may cast come doubt on the accuracy of this claim.

Support for hypotheses.

In support of predictions, the high rule space search group performed better in transfer than the low rule space search group indicating that manipulations were successful in producing the anticipated differences in learning. More specifically, transfer performance of the high rule space search group was equivalent to that of the low rule search group at the end of training, suggesting that the group readily adapted to the changed transfer conditions. Since the group had never experienced most of the task conditions presented in the transfer task, the result suggests they had developed adequate rule knowledge to apply to the novel task situations. It seems likely then that greater encouragement to focus on task rules facilitated rule knowledge acquisition in the high rule space search group in turn permitting their high transfer performance. Consistent with DST therefore, those who were encouraged to conduct a greater search of rule space appeared to develop a higher level of task knowledge.

Contrary to the high rule space search group, the low rule space search group appeared to develop a relatively low level of task knowledge during training.

Compared to the high rule space search group's end training performance, the low rule space search participants performed more poorly in transfer suggesting the group had not acquired a sufficient level of task knowledge to adapt effectively to the changed transfer conditions. Most obviously, the group committed a very high number of Rule-4 (short runway landing) errors throughout transfer suggesting that the group lacked the necessary rule knowledge to recognise when OpShort landings were not possible. Notably, despite the repeated negative feedback, the group continued to use their previous response pattern of using the short runways, effectively demonstrating a set effect (Luchins & Luchins, 1959; Luchins & Luchins, 1991; Woltz, Gardner, & Bell, 2000). However, since negative feedback has been shown to prevent set effects (Sweller & Gee, 1978), it seems that the group simply lacked sufficient knowledge to change their behaviour. Directly consistent with DST, it would appear that the low rule focus group learnt only a specific way of performing the task that was not amenable to novel task situations. Focusing on task instances, and less on task rules, therefore appears to produce poorer learning, which appears to persist despite extensive task practice.

Also consistent with predictions, cognitive load was consistent, and consistently high, for both groups throughout task performance. This was particularly noteworthy for the initial stage of learning when cognitive demands were expected to be highest and differences in cognitive load were therefore most likely. It may be argued however that between group comparisons of cognitive load depended heavily on sample size (more so than the performance comparisons for which the rule space search hypothesis depended) and that the limited size of the present sample lacked sufficient power to detect differences. However, none of the differences in cognitive load approached significance indicating that even with a substantially larger sample, the groups would have been unlikely to differ. Although interpretation of the null hypothesis is problematic (and the measures of cognitive load indirect), it seems likely that the high levels of cognitive load in both groups did not influence rule learning. Rule space search appears, on the basis of this preliminary investigation, to have influenced rule learning independently of cognitive load.

Further consideration of rule space search manipulations.

Despite strong support for the hypotheses, these results may not be unequivocal. The way in which rule space search was manipulated may have produced unintended differences between the groups, potentially confounding results. Rule space search was manipulated by making rule knowledge, particularly that concerning OpShort landings (Rule-4), appear more or less relevant to the performance of each group during training. This was intended to produce differences in each groups' attention to task rules and consequently elicit differences in rule space search. Care was taken to ensure that rule knowledge was in fact equally irrelevant to both groups because if the actual relevance of rule knowledge differed, it may have created disparate incentives to learn. This could consequently have produced differences in rule knowledge independently of rule space search. Despite this, rule space search manipulations may have produced unequal rewards for acquiring rule knowledge.

The importance of Rule-4 knowledge to each of the groups was considered, a priori, to be equivalent because both groups could use this knowledge equally infrequently. For both groups, knowledge of Rule-4 could only be used in one of the four weather conditions presented and then only in relation to landing one of the task's four plane types on a short runway. However, although Rule-4 knowledge could be used equally infrequently, the benefit it provided to each of the groups may not have been equivalent. For example, since the low rule space search group could almost always perform OpShort landings during training, the only benefit of acquiring Rule-4 knowledge would have been to reduce Rule-4 errors. Since this group could only rarely commit Rule-4 errors (i.e. in only one of four weather conditions with one of the four plane types), the benefit of acquiring Rule-4 knowledge was therefore likely to have been minimal. Conversely, the benefit of acquiring Rule-4 knowledge for the high rule space search group was likely to have been substantial. This group was trained under conditions where OpShort, and consequently overall, landings were limited but by acquiring Rule-4 knowledge, participants could land (up to six) additional planes each trial[7]. This represented a substantial benefit given the group landed only 49 planes on average per trial at their peak. Although Rule-4 knowledge could be used equally infrequently by both groups, the benefit it could have provided to each may not therefore have been equivalent. Rule knowledge may have been more relevant to the high rule space search group, potentially creating a stronger incentive to learn and confounding interpretation of rule space search manipulations.

In addition to potential differences in the relevance of rule knowledge, the costs associated with learning the task rules may also have differed between the groups. For the low rule space search group, the high frequency of landings possible during training meant that response speed, rather than task knowledge was likely to have been the predominant constraint on their performance. If learning the task rules detracted from the speed with which these participants could land planes, it may therefore have been detrimental to performance. Not only may rule knowledge have been largely irrelevant then, investing the effort to learn it may have produced poorer training performance. Conversely, for the high rule search group, the low frequency

---

[7] Weather conditions for the high rule space search group during training would have been conducive to OpShort landings for 2.5min each trial. Since planes took 15sec to land, six additional landings could have been performed with maximal Rule-4 knowledge.

of landings permitted during training meant participants were more likely to be spending longer waiting for runways to become available rather than landing planes. Although this probably encouraged greater rule space search, it also meant that investing the effort required to learn task rules would have detracted minimally, if at all, from their performance. Not only may rule knowledge have been highly relevant to this group, acquiring it may not have produced any decrement in performance as it may have for the low rule space search group. If rule learning detracted from performance unevenly, the groups would again have differed in their incentive to acquire rule knowledge. Since the costs were likely less for the high rule space search group, the cost of learning rule knowledge may represent a further confound to results.

Differences in the relevance and cost of acquiring rule knowledge may confound interpretation of results, potentially diminishing the certainty with which CLT and DST can claimed as independent. Not only may the observed group differences not be attributable to manipulations of rule space search, but if groups differed in their incentive to acquire rule knowledge, CLT (along with most other theories of learning) would be consistent with results. Learning requires some incentive to occur so, for most theories including CLT, knowledge would be anticipated to be higher in the group who had stronger incentives to learn. This effect would also have been independent of both cognitive load and rule space search meaning observed differences may be of little use in determining whether CLT and DST are independent. The potential influence of task manipulations on each group's incentives to acquire rule knowledge therefore suggests that results, whist promising, are not unequivocal in establishing the independence of CLT and DST. Further research is therefore required to more conclusively demonstrate the independence of CLT and DST.

Summary.

The present study sought to establish CLT and DST as independent theories of learning by manipulating rule space search independently of cognitive load. Results suggested that manipulations had been successful, with higher learning demonstrated in the high rule space search group and equivalent cognitive load between the groups. However, manipulations may have also created different incentives for each group to acquire rule knowledge. This may have confounded results by evoking stronger incentives to acquire task rules in the high rule space search group, which may account for the observed results. Whilst results were promising, further research is therefore required to investigate the independence of CLT and DST.

CHAPTER 5: STUDY 2

AN INVESTIGATION OF THE INFLUENCE OF BOTH RULE SPACE SEARCH AND COGNITVE LOAD DURING LEARNING

Study 1 suggested that learners who had been encouraged to focus on rule search during learning developed better knowledge than those who had not. However, results may have been attributable to different incentives to acquire rule knowledge rather than differences in rule space search. To address this ambiguity, the present study substantially revised the experimental design of Study 1 to provide a more robust comparison of CLT and DST. The revisions retained the principle of manipulating rule space search independently of cognitive load (consistent with the first hypothesis of the thesis) but added a manipulation intended to influence rule space search and cognitive load in the same direction (consistent with the second hypothesis of the thesis). That is, manipulations were also intended to elicit high rule space search under conditions of high cognitive load and low rule space search under conditions of low cognitive load, contrary to the predictions of CLT. Study 2 therefore sought to more thoroughly investigate the independence of CLT and DST than had been achieved in Study 1.

Manipulation 1: Goal type.

CLT argues that that rule space search is a consequence of cognitive load such that rule space search increases (decreases) when load falls (rises) (Sweller et al., 2011). CLT therefore purports that it effectively subsumes DST since rule space search is not separable from, and can be explained by, cognitive load. Like Study 1, the aim of the first manipulation in the present investigation was therefore to demonstrate that rule space search and cognitive load could vary independently by eliciting either high or low rule space search under conditions of equivalent cognitive load. Goal type assigned during learning was selected for this purpose.

Goal Type. As outlined in Chapter 2, both CLT and DST agree that in complex settings specific goals impair learning relative to non-specific goals because they elicit means-end performance strategies (Burns & Vollmeyer, 2002; Sweller, 1988; Sweller et al., 2011; Sweller et al., 1998; Vollmeyer & Burns, 1996). The theories differ in that CLT argues the strategies elicit extraneous cognitive load whilst DST argues that they discourage rule space search. However, since both theories predict that specific goals will impair learning relative to non-specific goals and that they do so via the same means-end mechanism, separating the accounts of CLT and DST using specific goals is difficult. Goal specificity was therefore not considered a useful manipulation to investigate the differences between the theories. Instead, the present study employed a manipulation of goal type, rather than goal specificity, to investigate the independence of cognitive load and rule space search. The two goals types employed were learning and performance goals.

Goal specificity is not the only goal manipulation that has been shown to produce differences in learning outcomes in complex settings. Assigning a learning goal, where learners are encouraged to focus on improving their skill, versus a performance goal, where focus is placed on achieving some level of task performance, has also been shown to elicit marked differences in learning outcomes across a broad range of complex and applied settings (Chen & Mathieu, 2008; Cianci, Klein, & Seijts, 2010; Elliott & Dweck, 1988; Gist & Stevens, 1998; Seijts, Latham, Tasa, & Latham, 2004). Learning and performance goals were first investigated by Dweck and colleagues who established that individuals can be characterised as pursuing mastery (i.e. learning) or performance goals during task performance (Diener & Dweck, 1978, 1980; Dweck & Leggett, 1988). Those who pursued learning goals focused on improving their skills and abilities whilst those with performance goals focused on proving the adequacy of their abilities by achieving high levels of performance. Although Dweck and colleagues found no differences between each goal orientation in performance of straightforward tasks, when tasks were challenging or forced individuals to fail, learning goal individuals reacted and performed markedly better than those with performance goals. They argued that the disparity in performance under challenging conditions was due to differences in how the goals influenced perceptions of challenge. For learning oriented individuals, challenging situations were considered an opportunity to achieve their goal of improving their skills and abilities, whereas for those pursuing performance goals, challenge was viewed as a negative evaluation of their ability. When presented with a difficult scenario, learning goals therefore tended to encourage increased effort and the development of novel and more effective strategies to improve task performance whilst performance goals tended to promote reduced effort to avoid any threat of negative evaluation, thus preventing improvement.

Whilst pursuit of either performance or learning goals has historically been considered a trait disposition (Dweck & Leggett, 1988), goal orientations can be effectively manipulated through goal assignment to produce similar outcomes. (Cianci et al., 2010; Elliott & Dweck, 1988; Seijts et al., 2004; Winters & Latham, 1996). Elliott & Dweck (1988) for example, demonstrated that when a performance goal was emphasised, by filming and giving evaluative feedback, children gave up more readily when performing the task and attributed failures to the inadequacy of their abilities. When learning goals were emphasised, by increasing the perceived value of competence, children were more likely to persist despite failures and develop more effective strategies for problem solving. Cianci et al (2010) also found that by instructing adults to "learn how to approach the task as well as possible" (p622), they produced less negative affect during a verbal learning task than those instructed to "perform as well as possible", consistent with the influence of respective trait orientations. In analyses conducted on more applied, simulated, tasks, both Winters & Latham (1996) and Seijts et al (2004) found that by instructing participants to learn a specific number of strategies, as opposed to achieving a certain level of performance, effectively mimicking the outcomes of trait goal orientations, superior learning and self-efficacy outcomes were observed. Assigning either performance or learning goals may therefore be an alternative to goal specificity in influencing learning outcomes.

The influence of goal manipulations has traditionally been understood in terms of affective and conative outcomes, influencing on-task effort, persistence, and affect (Dweck & Leggett, 1988), but their influence may also be understood in terms of DST. Learning goals, for example, encourage greater focus on task understanding, a process that requires focusing on developing knowledge about task rules and the interrelationships between task variables (Dweck & Leggett, 1988). This is clearly similar to rule space search. Learning goals have also been found to increase strategy systematicity and hypothesis testing (Diener & Dweck, 1978), which is directly akin to the processes elicited by rule space search manipulations (Burns & Vollmeyer, 2002; Vollmeyer et al., 1996). Learning goals have also been repeatedly associated with deeper (i.e. evaluative and elaborative) processing and learning strategies (Elliot, McGregor, & Gable, 1999), which are similar to the evaluative nature of hypothesis testing in rule space search. Whilst performance goals are argued to produce poorer performance through negative self evaluation (Dweck & Leggett, 1988; Kanfer &

Ackerman, 1989), their negative influence may also be due, in DST terms, to encouraging focus on task instances, rather than rules, since they necessarily specify some ‘instance’ of performance. Manipulating goal type by assigning learning or performance goals may therefore be an effective method of manipulating rule space search in place of goal specificity.

However, manipulating rule space search using goal type is unlikely to be effective in separating CLT and DST if both theories make similar predictions about their effects. Differences in rule space search must be unrelated to differences in cognitive load to establish the theories as distinct. Manipulations of goal type appear to satisfy this requirement. For CLT, the poor learning outcomes caused by specific goals are due solely to their specificity. It is the provision of a specific performance criterion for which learners aim that elicits the deleterious means-end strategies. Non-specific goals, regardless of their content, cannot cause such strategies because they do not provide any reference point for individuals to continually compare their performance[8]. Therefore, if both learning and performance goals are non-specific, they cannot elicit means-end strategies and are thus likely to elicit equivalently low levels of cognitive load. Manipulating rule space search using non-specific performance and learning goals is therefore unlikely to affect cognitive load. Goal type therefore appears to be a suitable manipulation for influencing rule space search independently of cognitive load and therefore for investigating the independence of CLT and DST.

In the present study, participants were assigned either a performance or learning goal by instructing them to either perform as best, or learn as much, as possible. Since learning goals have been associated with greater focus on rule

---

[8] Moreover, if both goals are non-specific, the task motivational elicited by both should be equivalently low since it is goal specificity, not goal type, that affects task motivation (Locke & Latham, 1990, 2002)

learning, superior strategy development, and deeper processing, rule space search was anticipated to be higher for those assigned a learning goal. Conversely, since performance goals focus on task outcomes (i.e. instances), they are likely to encourage instance space over rule space search. As a result of these differences in rule space search, learning goal groups were anticipated to develop higher levels of rule knowledge than those assigned a performance goal. Also, since both performance and learning goals were not specific, instructing participants only to 'do their best', neither goal condition was anticipated to elicit differences in cognitive load. However, consistent with Study 1, cognitive load was anticipated to be equivalently high for both goal conditions. These predictions are summarised in *Hypotheses 1a*, *1b*, and *1c*, below.

> *Hypothesis 1a*: Participants assigned a learning goal will demonstrate greater rule space search than those assigned a performance goal
>
> *Hypothesis 1b*: Participants assigned a learning goal will, as a result of greater rule space search, develop superior rule knowledge than those assigned a performance goal
>
> *Hypothesis 1c*: Cognitive load will be equivalent in both learning and performance goal groups.

Manipulation 2: Level of information.

As demonstrated in Study 1, attempting to distinguish between CLT and DST by only manipulating rule space search can limit interpretations. The present study therefore included a further manipulation intended to influence both cognitive load and rule space search to more conclusively test the independence of the theories.

CLT and DST offer different predictions of learning under conditions of uncertainty. Uncertainty occurs when existing knowledge, or the information provided in instructions and examples, does not provide sufficient information to know, in advance, the outcome of an interaction with the task. As elucidated by Sweller and colleagues (Sweller et al., 2011), the only logical possibility for a learner in this situation is to randomly generate a response and then test whether it is effective. For CLT, this process produces high cognitive load because the number of possible responses is usually high. Unsuccessful attempts also need to be remembered so as not to be repeated. Because working memory is particularly limited in its capacity to process the (by definition) unorganised information generated by this process, learning is slowed. According to the theory, learning outcomes can therefore be improved for novice learners by providing sufficient information to perform a task through worked examples (Cooper & Sweller, 1987; Sweller & Cooper, 1985), more detailed instructions (Kalyuga, Chandler, Tuovinen, & Sweller, 2001), or better guidance (van Merrienboer, 1990), which serve to decrease the load associated with randomly generating and testing response options.

The process of generating and testing hypotheses about task responses is however central to rule space search in DST. By devising and continually testing hypotheses about a task, DST argues that learners engage in an active and constructive search to improve understanding of rules governing a task's operation (Klahr & Dunbar, 1988; Simon & Lea, 1974). For DST, greater hypothesis testing represents greater search of rule space, which produces better learning outcomes[9] (Burns & Vollmeyer, 2002; Vollmeyer et al., 1996). Increasing uncertainty in a task by not providing sufficient information should therefore increase hypothesis testing, rule space search, and, consequently, learning. CLT and DST thus make opposing predictions in situations of high uncertainty: CLT argues that increasing uncertainty will increase load and reduce learning, whilst DST argues that it will increase rule space search and improve learning. Manipulating uncertainty therefore represents an effective means of separating CLT and DST.

The present study sought to manipulate uncertainty by varying the amount of rule information provided to participants during learning. Participants were provided either full or partial information about the task to create conditions of low or moderate uncertainty. Under full information conditions, all rules were provided to participants during training. Uncertainty was therefore anticipated to be minimal since participants could continually refer to the rule information provided to guide their response selections. Under these conditions, both load and rule space search were anticipated to be relatively low because almost no search of the task was required to learn the rules governing performance. Under partial information conditions, no Rule-4 information was provided to participants during training. Uncertainty was therefore anticipated to be higher since participants could not refer to the provided information to guide all responses. Instead, participants would have had to generate and test responses to develop an understanding of task rules based on their outcomes. Under these conditions, both load and rule space search were anticipated to be higher because more random search was needed to learn the rules of the task. Whilst CLT would

---

[9] As Burns and Vollmeyer (2002) note however, a completely unconstrained search of rule space may not result in superior task learning outcomes. If learners continually test inappropriate or misleading hypotheses, learning is unlikely to proceed regardless of the extent of hypothesis testing. Assuming at least some guidance of search is provided, this is considered unlikely.

predict better learning under the full learning conditions where fewer random-generate-and-test strategies were required, DST predicts, somewhat counter intuitively, that learners provided less information will learn more because they will engage in greater search of rule space. Manipulating uncertainty by varying the amount of task information provided to participants is therefore an effective means of separating CLT and DST because the theories make divergent predictions under conditions of uncertainty.

Specific predictions concerning manipulations of information level were as follows. Since both CLT and DST predict that greater uncertainty would increase search of the task, it was anticipated that partial information groups would exhibit higher rule space search than full information groups. This was anticipated to produce higher cognitive load in the partial information groups, consistent with CLT. However, contrary to CLT (and consistent with DST), partial information groups were anticipated to develop better rule knowledge than full information groups due to their greater search of rule space. These predictions are summarised in *Hypotheses 2a*, *2b*, *and 2c* below.

> *Hypothesis 2a:* Rule space search will be higher under partial than full information conditions.

> *Hypothesis 2b:* As a result of greater rule space search, partial information participants will develop greater rule knowledge than those provided full information.

Hypothesis 2c: Cognitive load will be higher under partial information conditions but this will not impair development of rule knowledge.

Combined influence of goal type and information level manipulations.

In the present study, participants received either a performance or learning goal under either full or partial information conditions in a fully crossed design, as illustrated in Table 8. The combined effects of manipulations were intended to more clearly demonstrate the independence of cognitive load and rule space search.

Table 8: Experimental design and anticipated influence of manipulations on cognitive load and rule space search

		Information Level	
		Full	Partial
Goal Type	Performance	- Low Load - Low Rule Space Search	- High Load - Moderate Rule Space Search
	Learning	- Low Load - Moderate Rule Space Search	- High Load - High Rule Space Search

In particular, two groups were anticipated to demonstrate levels of cognitive load and rule space search that were directly contrary to the predictions of CLT. One group received both performance goal and full information manipulations. Since both manipulations discouraged rule space search, and neither was anticipated to elevate cognitive load, both cognitive load and rule space search were anticipated to be low. The other group received both learning goal and partial information manipulations. Since both manipulations encouraged rule space search, and one also likely elevated cognitive load, this group was expected to demonstrate both high rule space search and high cognitive load. Since CLT argues that rule space search is possible only under conditions of low cognitive load and that rule space search should be high under conditions of low cognitive load, these groups represented direct tests of CLT's predictions. Consistent with DST however, rule learning was anticipated to be consistent with the level of rule space search shown. These predictions are summarised in *Hypotheses 3a* and *3b* below.

> *Hypothesis 3a*: Contrary to the predictions of CLT, the Performance Goal/Full Information group was anticipated to demonstrate both low cognitive load and low rule space search whilst the Learning Goal/Partial Information group was anticipated to demonstrate both high cognitive load and high rule space search in the Acquisition Phase.

> *Hypothesis 3b*: For groups receiving consistent goal and information level manipulations, learning outcomes were anticipated to reflect the level of rule space search rather than cognitive load. Learning outcomes were therefore anticipated to be higher for the Learning Goal/Partial Information group than the Performance Goal/Full Information group.

The influence of combined manipulations for the other two groups (shown in the bottom left and upper right of Table 8) was less straightforward. For these groups, the effects of goal type and information level manipulations were anticipated to exert opposing influences on the level of rule space search. Vollmeyer et al (1996) also investigated the combined influence of conflicting rule space search manipulations: instruction in a hypothesis testing strategy (intended to increase rule space search) in combination with a specific goal (intended to discourage rule space search). Results indicated that the combined effect of the manipulations produced moderate rule space search and learning: above those of participants who were provided only a specific goal with no instruction, but below those of participants provided a non-specific goal with strategy instruction. Consistent with Vollmeyer et al (1996), the combined manipulations were therefore anticipated to elicit moderate levels of rule space search in both groups.

The influence of a moderate search of rule space was not however anticipated to be equivalent between the groups. For the group assigned a performance goal with partial information, extensive rule space search was required to discover the task rules because they had not been provided. However, for the group provided with a learning goal with full information, a far less extensive search of rule space was required because all rules were provided. Whilst a moderate search of rule space may have been sufficient for the Learning Goal/Full Information group, it was likely to have been inadequate for the Performance Goal/Partial Information group. Despite showing similar levels of rule space search, learning outcomes for the groups may have differed markedly.

If similar levels of rule space search elicit differences in rule knowledge depending on the level of information provided, results will reveal an interaction between goal type and information level manipulations. More specifically, differences in learning outcomes between performance and learning goal groups will be more pronounced under partial compared to full information conditions. For full information conditions, learners may have been able to acquire a high level of rule knowledge, because rules were provided, without a high level of rule space search. For partial information conditions, where rule space search was the only means of acquiring rule knowledge, differences in rule space search are more likely to have produced large differences in rule knowledge. Whilst learning goals were anticipated to produce better learning outcomes than performance goals (*Hypothesis 1b*), the difference was therefore anticipated to be more pronounced under partial compared to full information conditions. This prediction is summarised in *Hypothesis 3c* below.

> *Hypothesis 3c*: The influence of goal type on learning outcomes is anticipated to be moderated by level of information such that differences in leaning outcomes will be more pronounced under partial compared to full information conditions.

Research overview.

The two goal and information manipulations formed a 2 x 2 between-subjects factorial design. Goal type was manipulated by instructing participants to aim to either score as many points as possible (performance goal) or learn as much about how to play the game as possible (learning goal). Both goals were non-specific to avoid eliciting means-end strategies. Level of information was manipulated by providing participants either complete (full) or incomplete (partial) Rule-4 information about short runway landings.

To ensure differences in learning were attributable only to differences in early learning conditions, and not to conditions under which practice occurred (unlike Study 1), manipulations were administered during an initial ‘acquisition’ phase of the study only. Differences in cognitive load and rule space search were therefore anticipated only during this phase when participants were performing the task under their respective goal and information level conditions. By manipulating goal type and information level only in the early stages of learning, the study aimed to more effectively test whether participants could overcome the influences of initial learning conditions with practice or whether differences would persist throughout performance.

After the acquisition phase, when task conditions were equivalent, learning outcomes could be compared directly between the groups. The phase following acquisition was the ‘performance’ phase where all participants were instructed to maximise their score (performance goal) under partial (no Rule-4) conditions. These conditions were selected to test the rule knowledge participants had gained during the acquisition phase by requiring them to use their knowledge to achieve a task outcome under conditions where knowledge deficiencies could not be remedied simply by reading provided information.

Following the performance phase, a test phase was administered in which task conditions were altered to increase the importance of rule knowledge to performance. Since participants were expected to differ predominately in their knowledge of Rule-4 (particularly the knowledge concerning landing mid-sized planes on short runways-OpShort), test phase conditions forced participants to make greater use of the short runways than they had in preceding trials[10]. Those who had acquired a high level of knowledge in the preceding phases were expected to cope more effectively with the changed conditions by making greater use of the short runways. The test phase therefore provided a more stringent test of knowledge differences than previous trials.

In accordance with predictions, rule space search and cognitive differences were anticipated only in the initial acquisition phase of the study when the groups were performing the task under different conditions. Differences in learning outcomes were not assessed until the performance and test phases when participants were performing the task under like conditions. Differences between the groups were

---

[10] This manipulation based on Schunn & Reder (2001)

anticipated to be particularly pronounced in the test phase when task difficulty was increased to more explicitly test knowledge.

Method

Participants

82 first year Psychology students (55.8% female) aged between 18 and 37 from the University of Sydney participated in the study for course credit. Seven participants were excluded from analyses: five for not completing the study and two for obtaining task performance scores lower than 2.5 standard deviations below their group mean in every task trial, suggesting they were not taking performance seriously. A final sample of 75 was considered for analyses.

Apparatus

The apparatus used to present the task were identical to Study 1

Measures.

Dependent Measures. The primary dependent measure was the number of planes landed by participants during each task trial. Higher numbers of Landings indicated that participants had acquired higher levels of task knowledge and proficiency. OpShort landings, consistent with Study 1, were considered indicative of Rule-4 knowledge. Unlike Study 1 however, Rule-4 errors were considered indicative of task exploration (i.e. rule space search), though in the acquisition phase only[11].

---

[11] Although indirect, this method of measuring rule space search was considered preferable to direct measures such as think aloud protocols (e.g. Burns & Vollmeyer, 2002). Direct measures are typically employed in untimed problem-solving tasks where they provide insight into participants' decision-making processes. Since the present task relied less on decision-making than timely information processing, it is unlikely that thinking aloud would have been as informative. Indeed given the non-verbal and time-pressured nature of the task, direct measures may have impaired performance. Direct measures would also have required an almost impossible volume of analysis given present sample sizes.

This was because they represented a valid way of developing Rule-4 knowledge by attempting short runway landings and observing the consequences. This was particularly the case for partial information groups who were not provided with Rule-4 meaning this was the only way they could develop Rule-4 knowledge. Rule-4 errors were also not penalised during the acquisition phase to encourage task exploration in all groups. Following the acquisition phase however, Rule-4 errors were penalised and Rule-4 errors were then taken to indicate a lack of Rule-4 knowledge

Cognitive Ability. Like Study 1, the Raven's Advanced Progressive Matrices (APM) (Raven et al., 1994) test and the Noughts and Crosses working memory task (Mackintosh & Bennett, 2003), were administered to assess general cognitive ability. Given the failure to observe the anticipated correlation between these measures in Study 1 however, a third general cognitive ability measure was also introduced. Complex span working memory tasks (Unsworth & Engle, 2007) are widely considered to be highly representative of working memory capacity. Working memory is highly correlated with general cognitive ability (Ackerman et al., 2005; Colom et al., 2004) predictive of performance in cognitively complex domains, similarly predictive of performance in complex, attentionally demanding tasks (Conway, 1996). Also, complex span tasks are generally regarded as good measures of working memory capacity (Conway, Kane, Bunting, Hambrick, Wilhelm, & Engle, 2005). Each item in the measure presented participants with a series of simple equations followed by a single letter, e.g. "(4 x 2) -1 = 5 B". Participants were given 5 seconds to indicate whether the equation was true or false and remember the letter shown. Once the participant had selected true or false, or time ran out, the next equation and letter were presented. After viewing the series of equations, participants were asked to recall, in order, the letters only. Each level of the measure consisted of three sets of equations and letters and the length of each set increased incrementally from three to seven. One point was given for each complete series of letters correctly recalled. A full description of the task can be found in (Unsworth & Engle, 2007).

Design

As shown in Table 2, the research utilised a 2(Goal Type) x 2(Information Level) between subjects factorial design. Goal Type was manipulated by instructing participants either to maximise their performance score during (Performance Goal) or learn as much about the task as possible (Learning Goal), during the Acquisition Phase. Level of information was manipulated by providing, or not providing, participants with Rule-4 during instructions and the Acquisition Phase.

Table 2. Experiment design including number of participants in each group.

		Rules	
		Full (all rules shown)	Partial (Rule-4 not shown)
Goal Type	Performance	Group 1 (19)	Group 2 (18)
	Learning	Group 3 (17)	Group 4 (21)

Procedure

General procedure (all participants). Participants were randomly allocated to groups according to experiment session. Participation was completed over two days, a maximum of two days apart, in groups of between three and seven. Participants began the first session with task instructions presented via a Microsoft PowerPoint presentation projected onto an in-room screen. Accompanying verbal instructions, specific to each group, were read from a script specific. Instructions informed participants about the parts of the display, how to play the game (specific to each Goal Type group), the rules of the game (specific to each Level of Information group), and how to use the keyboard. Instruction duration was approximately 7min.

Following instructions, participants were shown a demonstration of the game in which each sub-task was performed and explained. Specifically, a plane (747) was added to the queue, moved down through the holding pattern, and then landed on one of the long runways. Participants then logged into the KA-ATC task and completed 6 x 10 minute trials of the task in succession. Participants were permitted to take short breaks between trials if required.

At the beginning of each trial an information screen was shown that displayed the response keys needed to play the game and the set of rules, appropriate to each group manipulation and trial. Before beginning trial 6, participants were told that the upcoming trial was "slightly different" to preceding trials but were not told of any specific changes to the task. At the completion of the six trials, participants left and were reminded to return at their allocated time.

In the second session, participants completed the individual difference measures in this order: demographics questionnaire, APM, Operations Span, and Noughts and Crosses.

As shown in Table 3, the six task trials were divided into three phases: Acquisition, Performance, and Test. Manipulations were administered during the instructions and initial Acquisition Phase only. After the Acquisition Phase, all participants completed the trials under identical task conditions.

Table 3. Experimental procedure. Manipulations were administered during instructions and during the first two trials (acquisition phase). Following the acquisition phase, all participants experienced the same conditions.

	Acquisition phase		Performance phase			Test phase
	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6
Group 1:	Performance Goal/ Full Information		Maximise performance score. No Rule-4 given			Task difficulty increased
Group 2:	Performance Goal/ Partial Information					
Group 3:	Learning Goal/ Full Information					
Group 4:	Learning Goal / Partial Information					

Acquisition Phase. Goal Type was manipulated during instructions and during the Acquisition Phase. Performance Goal groups (1 and 2) were told at the beginning and end of instructions that they should “aim to score as many points as possible by landing as many planes, and making as few errors, as possible”. Cumulative point scores were also visible to these groups both during and at the end of each trial of the training phase, in same way as Study 1. Learning goal groups (3 and 4) were instructed to “aim to learn as much about the game as possible and, in particular, [to] focus on learning the weather conditions governing landings on the short runways”. Point scores were not made visible to Learning Goal groups throughout the training phase, either during or at the end of each trial, to encourage participants to focus on learning rather than achieving high performance scores. Score labels were shown but the areas where actual scores were shown were left blank.

Information level was manipulated during instructions and the initial acquisition phase only. Participants in full information groups (1 and 3) were shown complete versions of Rule-4 (as shown in Table 1) during instructions and during the training phase where it was presented on both the information page that was presented before each trial, and as an error message that would appear during play if Rule-4 was violated. Full Information participants could also view Rule-4 at any time during play by pressing key ‘4’ which would display Rule-4 in the error message box. Participants in the partial information groups (2 and 4) were not given detailed information about Rule-4 during instructions and were instead told that “weather conditions determine when each plane type can land on the short runways” and that part of their task was to “work this out” during play. Partial Information groups were also not shown Rule-4 on the information page or as an error message if Rule-4 was violated; instead of the rule appearing, an alternative message was shown that read: “plane cannot use short runway under current weather conditions”.

In order to avoid the potential confound of penalising errors in the groups who could see their scores, Rule-4 errors were not penalised during the training phase for any of the groups. This was more relevant for performance goal groups who could see their score at all times and could therefore have sought to avoid Rule-4 errors by virtue point deductions rather than because of their focus on performance.

Performance phase. Following the acquisition phase, all participants performed the task under identical conditions. These conditions were identical to those experienced by the training phase conditions experienced by the performance goal/partial information group. Participants were instructed to maximise their point score, which was displayed both during and after each trial. Rule-4 was not viewable to participants at any time, and the alternative message “plane cannot use short runways under current weather conditions” was shown if Rule-4 was violated or if key ‘4’ was pressed.

Test phase. To evaluate participants’ Rule-4 knowledge acquisition more directly, task difficulty was increased in the test phase. By the test phase, all participants were likely to have reached, or were very near to reaching, a plateau in their performance, indicating that learning had reached a maximum. By increasing task difficulty, test phase performance aimed to identify those who had acquired higher levels of knowledge when learning was complete, and therefore independently of any differences in learning rates between groups.

Task difficulty was increased by altering the proportion of plane types presented, in order to increase the importance of Rule-4 knowledge in performance. In the acquisition and performance phases, each plane type was presented with equal frequency so that each plane type consisted of approximately 25% of all planes presented. In the test phase however, the proportion of large and medium-sized planes was increased such that 747s comprised 35% of planes, 727s 30%, DC10s 30%, and Props 5%. The higher proportion of 747s, which could only use long runways, and lower proportion of Props, which could always use short runways, was intended to force participants to land more 727s and DC10s on the short runways. Participants who had learnt the more complex aspects of Rule-4 would therefore be more able to use the short runways under these conditions. This manipulation has been used effectively for a similar purpose by Schunn & Reder (2000). Differences in Test Phase performance were therefore considered to be more sensitive to Rule-4 knowledge differences than previous phases' performance.

Results

Analyses

To test the effects of between group manipulations 2x2 (goal type x information level) between-subject ANOVAs were conducted for each phase of task performance. For each phase, one ANOVA was carried out for each of the task measures: Landings, Rule-4 errors, and OpShort. To clarify any significant interactions between goal type and information level, further pairwise tests were also conducted. ANOVAs were the same for all three experimental phases.

To determine whether manipulation influenced cognitive load during learning, three moderated regression analyses were conducted for each dependent measure in the acquisition phase. This was achieved using the procedure outlined in (Baron & Kenny, 1986). Analyses compared the strength of the relationship between cognitive ability and dependent measures (Landings, Rule-4 errors, and OpShort) between goal type and information level groups. Only acquisition phase data was analysed since it was the only stage in which cognitive load was anticipated to differ. Task conditions differed in this phase only and load was anticipated to be high only for the initial few trials of learning, consistent with previous research (Ackerman, 1987, 1988, 1992; Kanfer & Ackerman, 1989).

Manipulation check.

Fluid Ability. Means, standard deviations, reliabilities, and correlations between the three ability tests are provided in *Table 9*. As shown, the measures: APM, Noughts and Crosses, and Operations Span, were all moderately to highly intercorrelated (r >.34) suggesting that they could be combined to form a single fluid ability composite. Despite the small sample size, an exploratory factor analysis was conducted on the three measures to examine this possibility. Principal axis extraction with no factor rotation was used since data was normally distributed[12] and only one factor was anticipated. In support of a composite fluid ability indicator, the analysis revealed a single factor, which accounted for 46.25% of the variance in the measures. A composite measure was then created using unit-weighted *z* scores of the three fluid ability measures. This measure was used for all further analyses involving fluid ability.

Table 9: Means, standard deviations, reliabilities and correlations between fluid ability measures.

Measure	M	SD	1	2	3
1. APM	23.39	6.266	.86		
2. Operation Span	8.19	3.38	.36**	.74	
3. Noughts & Crosses	15.03	4.26	.57**	.39**	.73

Note: N= 75. ** Correlation is significant at .01. Reliabilities for each measure are displayed on the diagonal.

---

[12] Data satisfied assumptions of normality: skewness<|.37|; kurtosis<|.55|, and appeared amenable to |.37| factoring: Kaiser criterion= .66; Bartlett's test of sphericity: $\chi^2(3)= 42.01$, $p<.01$.

Dependent measures. To ensure that Rule-4 errors could be used as indicators of task exploration, the relationship between each of the three dependent measures was first analysed. Rule-4 errors were anticipated to represent task exploration during the acquisition phase (when they were not penalised) when they were likely to reflect attempts to develop Rule-4 knowledge, particularly for partial information groups. During the later performance test phases when these errors were penalised, Rule-4 errors instead anticipated to reflect poor Rule-4 knowledge. As shown in *Table 10* however, correlations did not appear to support these interpretations.

Table 10: Means, standard deviations, and intercorrelations for fluid ability and dependent measures for each phase of the study.

	M	SD	1	2	3	4	5	6	7	8	9	10
1. Fluid ability Composite (*z*)	0.00	1.00	-									
Landings												
2.Acquisition	36.87	7.67	.47**	-								
3.Performance	48.07	6.99	.55**	.70**	-							
4.Test	43.20	6.08	.42**	.59**	.67**	-						
Rule-4 Errors												
5.Acquisition	6.52	3.70	-.23*	-.09	-.03	-.07	-					
6.Performance	5.09	3.55	-.29*	-.19	-.11	-.09	.55**	-				
7.Test	7.09	7.16	-.19	-.27*	-.03	-.05	.29*	.66**	-			
OpShort												
8.Acquisition	5.18	2.72	.26*	.26*	.35**	.47**	.24*	.16	.01	-		
9.Performance	6.70	3.10	.19	.19	.44**	.38**	.11	.33**	.37**	.36**	-	
10.Test	9.28	5.25	.26*	.26*	.46**	.80**	-.03	.09	.25*	.44**	.55**	-

*Note: N= 75. *Correlation is significant at .05, **Correlation is significant at .01*

If Rule-4 errors were indicative of attempts to gain Rule-4 knowledge during the acquisition phase, errors made during the acquisition phase should be positively correlated with later knowledge and performance measures. As shown in *Table 10*, however, correlations between acquisition phase Rule-4 errors and later phase Landings and OpShort were not significant. It may have been the case however that a relationship between acquisition phase errors and later performance was evident only for participants provided partial rule information since committing Rule-4 errors was the only means they could acquire the rule knowledge they had not been provided. A series of moderated regression analyses were conducted to examine this possibility.

As shown in Table 11, the relationship between acquisition phase Rule-4 errors and test phase Landings differed according to information level with partial information groups indeed demonstrating a significantly more positive (though non significant: r= .18, ns) relationship than full information groups (r= -.32, p=.06). The same effect, though more pronounced, was also observed between acquisition phase Rule-4 errors and test phase OpShort; partial information: r= .27, ns; full information: r=-.33, p<.05). The pattern of findings therefore suggests that making Rule-4 errors during the initial stage of learning was more beneficial for partial information groups than full information groups and suggests that Rule-4 errors may be considered, at least for partial information groups, as attempts to explore the task to increase knowledge. Rule-4 errors in the acquisition phase were therefore considered a valid indicator of rule space search for partial information groups[13].

---

[13] The negative relationships observed for the full information groups were surprising. They suggest that rule-4 errors for these groups did not represent learning opportunities but a lack of knowledge. It seems that for participants provided full rule information, errors were an indication of failing to learn the provided information rather than attempts to improve their knowledge.

Table 11: Results of the four moderated regression analyses predicting Landings and OpShort for the Performance and Test Phases.

	Performance Phase				Test Phase			
	Landings		OpShort		Landings		OpShort	
Predictor variable	β	ΔR^2	β	ΔR^2	β	ΔR^2	β	ΔR^2
Step 1		>.01		.01		>.01		>.01
Rule-4 Errors	-.03		.11		-.07		-.03	
Step 2		>.01		.14**		>.01		.03
Goal Type	.06		.07		-.04		.02	
Information Level	-.03		-.38**		-.04		-.18	
Step 3		.01		.03		.08		.10
Goal Type x Rule-4 errors	-.14		.22		.22		.22	
Information Level x Rule-4 errors	.19		.39		.51*		.59*	
Overall R		.12		.43		.30		.37
Overall R^2		.01		.19		.09		.14
Adjusted R^2		-.06		.13		.02		.07
Overall F (5, 69)		.19		3.18*		1.32		2.15

*Note: N= 75. *p<.05, **p<.01*

In the performance and test phases, Rule-4 errors were penalised and were thus anticipated to correlate negatively with knowledge and performance scores. As shown in Table 10 however, correlations between Rule-4 and Landings were not significant but were positive for OpShort. The absence of correlations for landings suggests that Rule-4 errors and overall proficiency were unrelated; some low performing participants may have avoided the short runways, thus minimising their errors whilst some high performing participants may have made many errors in trying to maximise their use of the short runways. The positive correlation between Rule-4 errors and OpShort however suggests that rather than Rule-4 errors indicating a lack of rule knowledge, the combined Rule-4 and OpShort data appear to indicate a general propensity to use the short runways for mid-sized aircraft because those who had more rule knowledge, as evidenced by high levels of OpShort, also made a high number of Rule-4 errors. Rule-4 errors were therefore not interpreted as indicators of poor rule knowledge in performance and test phases, as had been expected.

Changes made to the test phase conditions were intended to increase the proportion of mid-sized aircraft landed on the short runways (OpShort) to more strongly test participants' Rule-4 knowledge. To assess whether these changes were effective, correlations between Landings and OpShort for the test phase were compared to previous phases. In support of the manipulations, Fisher's z comparisons revealed that the correlations between Landings and OpShort were significantly higher in the test phase than in either of the preceding phases (acquisition phase: z= 3.45, p<.01; performance phase: z= 3.76, p<.01). This suggests that Rule-4 knowledge was more relevant to performance under the test phase than previous phases and therefore supports consideration of the test phase as a stronger indicator of Rule-4 knowledge.

Acquisition phase.

Rule space search. Differences in Rule-4 errors (as a measure of rule space search) and cognitive load were anticipated in the acquisition phase when manipulations were administered and groups were performing the task under different conditions. As shown in *Table 12*, a 2x2 between-subjects ANOVA on Rule-4 errors revealed a significant difference according to goal type, with learning goal groups showing significantly higher Rule-4 errors than performance goal groups. As illustrated in *Figure 4*, the result was consistent with *Hypothesis 1a* with learning goal groups showing greater rule space search than those assigned performance goals. Contrary to *Hypothesis 2a* however, no difference in Rule-4 errors was observed between information level groups. This was somewhat surprising since committing Rule-4 errors was the only means for these groups to acquire Rule-4 knowledge and such errors were not penalised during the acquisition phase. As shown in *Figure 4* however, Rule-4 errors were higher for both partial compared to full information groups, consistent with predictions, but the difference did not achieve significance (*p*=.14). The information level manipulation therefore appeared to be a somewhat weaker influence of rule space search than goal type.

Table 12: ANOVA F tests for the three 2x2 (Goal Type x Information Level) Acquisition Phase analyses on Landings, Rule-4 errors and OpShort.

Factor	Landings	Rule-4 Errors	OpShort
Between-Subjects			
Goal Type	<1	6.76*	<1
Information Level	<1	2.25	2.16
Goal Type x Information Level	<1	<1	5.69*
MSE	60.42	12.47	6.95

*Note: N= 75. All df = (1, 71). *p<.05, **p<.01.*

Hypothesis 3c predicted that the difference in rule space search between performance and learning goal groups would be larger under partial than full information conditions. Contrary to this prediction however, analyses revealed no interaction between goal and information level in Rule-4 errors, as shown in *Table 12*. However, as illustrated in *Figure 4*, rule-space search appeared to be highest in the Learning/Partial group and lowest in the Performance/Partial group, consistent with *Hypothesis 3a*. The combined influence of consistent manipulations (i.e. those that were expected to have a similar effect on rule space search) therefore appeared to provoke particularly high or low levels of rule space search, as had been anticipated. Further pairwise analyses[14], revealed that the Learning/Partial group committed significantly more Rule-4 errors than either of the performance goal groups (Performance/Full: t(71)= 2.81, *p*<.01 ; Performance/Partial: t(71)= 2.30, *p*<.05), tentatively supporting the suggestion that combination of consistent manipulations

---

[14] Note however that type-1 error rates were not controlled for these analyses. If type-1 error rates had been controlled using the Bonferroni method, only the comparison between the Learning/Partial and Performance/Full groups would have achieved significance. Interpretations have been qualified as a result.

(i.e. learning goal with partial information or performance goal with full information) did produce complementary effects on rule space search.

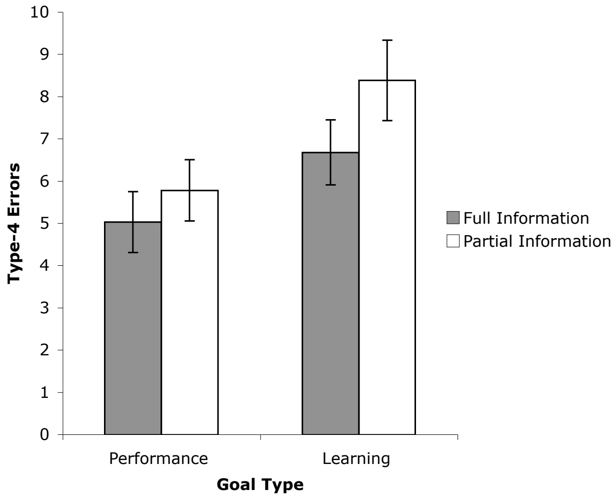


Figure 4: Mean Rule-4 errors (+/-1 SE) made by each group during the acquisition phase.

Cognitive load: Cognitive load was indirectly shown by the correlation between the fluid ability composite and the dependent measures of task performance with stronger correlations considered indicative of greater cognitive load. To examine whether cognitive load differed according to goal or information level manipulations during the acquisition phase, three moderated regression analyses (Baron & Kenny, 1986) were performed[15], one for each dependent measure. Fluid ability was entered first into each regression to determine whether fluid ability was predictive of each dependent measure. In the second step, goal type and information level were entered to control for any between group differences in the dependent measures before the two moderation terms were entered in the third and final step. Partial information groups were anticipated to show higher cognitive load than full information groups due to

---

[15] Analyses were performed only on acquisition phase results since this was when cognitive load was anticipated to be highest (Ackerman, 1992; Kanfer & Ackerman, 1989)

their greater need to conduct an unsystematic search of the task to develop adequate Rule-4 knowledge (*Hypothesis 2c)*.

Table 13: Results of moderated regression analyses predicting Landings, Rule-4 Errors, and OpShort, by Fluid Ability and group manipulations for the acquisition phase.

Predictor variable	Landings		Rule-4 Errors		OpShort	
	β	Δ R^2	*β*	Δ R^2	*β*	Δ R^2
Step 1		.23**		.05*		.07*
Fluid ability	.47**		-.23*		.26*	
Step 2		.05		.11*		.08
Goal Type	.28		.54*		.18	
Information Level	.28		.28		-.27	
Goal Type x Information Level	-.33		.37		.95*	
Step 3		.04		>.01		.06
Goal Type x Fluid ability	-.28		.04		.06	
Information Level x Fluid ability	-.31		-.04		-.51*	
Overall R		.56**		$.40^{\dagger}$		.46*
Overall R^2		.31**		$.16^{\dagger}$		.21*
Adjusted R^2		.25**		$.09^{\dagger}$		.14*
rall F (5, 69)		5.12**		$2.17^{\dagger}$		2.95*

Note: † p<.06, **p<.05, **p<.01.*

Table 13 displays the three moderated regression analyses predicting acquisition phase performance (landings, Rule-4 errors, and OpShort) by fluid ability and the group manipulations[16]. As shown, fluid ability significantly predicted acquisition phase landings, Rule-4 errors, and OpShort, indicating that those higher in fluid ability landed more planes, made fewer Rule-4 errors, and made more effective use of the short runways. This supports predictions that information-processing demands, and therefore cognitive load, were high during the initial phase of learning.

---

[16] Fluid ability scores were standardised prior to computing interaction terms, and standardised dependent variables were also used in all analyses (Friedrich, 1982). Group variables were effect coded so coefficients referred to between group differences and were therefore comparable to ANOVA analyses. 3-way interactions were not included as none related to hypotheses, however none were significant.

Interestingly, cognitive load was negatively related to rule space search overall, consistent with the predictions of CLT. Results were also consistent with ANOVA results despite controlling for fluid ability. This suggests that there was little need to control for fluid ability in ANOVA analyses.

In support of *Hypothesis 1c*, cognitive load was not found to differ according to goal type for any of the three dependent measures. The difference in rule space search that was observed between learning and performance goal groups therefore appeared, as predicted, to be independent of any variation in cognitive load. The relationship between cognitive load and OpShort was however found to differ according to information level. Whilst the significance of the result was consistent with *Hypothesis 2b*, the direction was opposite to predictions with higher load evident in full, rather than partial, information groups, as shown in *Figure 5*. Correlations were also higher in the Performance/Full compared to Learning/Partial group, contrary to *Hypothesis 3a.* A further test of simple main effects (Holmbeck, 2002) revealed that the relationship between fluid ability and OpShort was positive for the Full Information group (β = .99, t(68)= 2.96, p<.01) but effectively zero for the Partial Information group (β < .01, t(68)<.01, ns) suggesting that the provision of complete information produced higher, rather than lower, cognitive load during the acquisition phase.

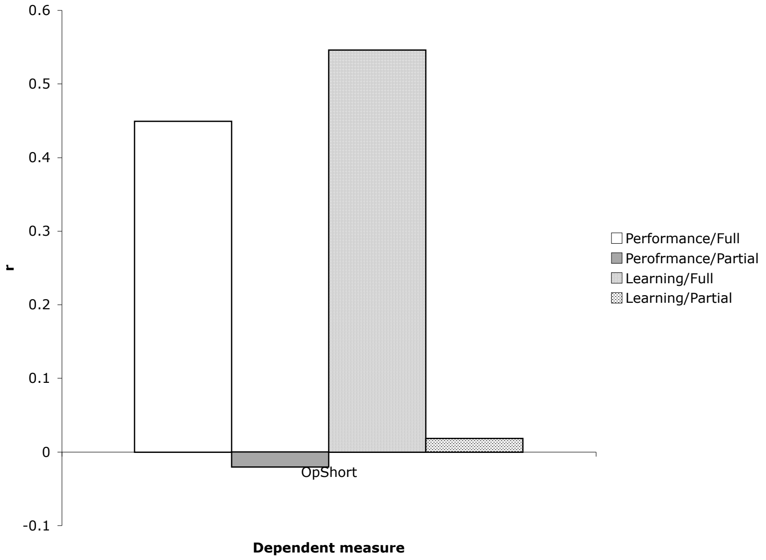


Figure 5: Correlations between the fluid ability composite and OpShort for all four groups.

Performance measures. Although no specific predictions were made about Landings and OpShort in the acquisition phase, both measures were analysed to determine whether groups differed in their level of task proficiency (landings) or knowledge (OpShort) under the diverse learning conditions. As shown in Table 3, ANOVA analyses revealed no differences in Landings according to goal type, information level, or their interaction, suggesting that all groups performed similarly over the phase. Differences were observed for OpShort, as shown in *Table 12*, with the interaction between goal type and information level achieving significance. As illustrated in *Figure 6*, this was likely due to the Performance/Partial group who appeared to perform few OpShort landings relative to other groups. Further pairwise tests supported this observation indicating that the Performance/Partial group performed fewer OpShort landings than either the Performance/Full (t(71)= 2.72, *p*<.01) or Learning/Partial (t(71)= 2.17, *p*<.05) groups, but a similar number to the

Learning/Full group (t(71)= 1.44, *p*=.16), with all other groups performing similarly (t(71)< 1.22, *p*>.23).

The poor performance of the Performance/Partial group, in particular compared to the Learning/Partial group, suggests that goal conditions have a more pronounced influence on learning when partial information is given, consistent with *Hypothesis 3c*. It would appear that the extensive rule space search undertaken by the Learning/Partial group facilitated knowledge acquisition, to a similar level of those provided full information, but the lower rule space search evidenced by the Performance/Partial group impeded knowledge acquisition, even as early as the acquisition phase. Interestingly, although the Performance/Full group showed relatively little rule space search, their OpShort performance was high suggesting that the provision of full rule information may mitigate the influence of rule space search on learning.

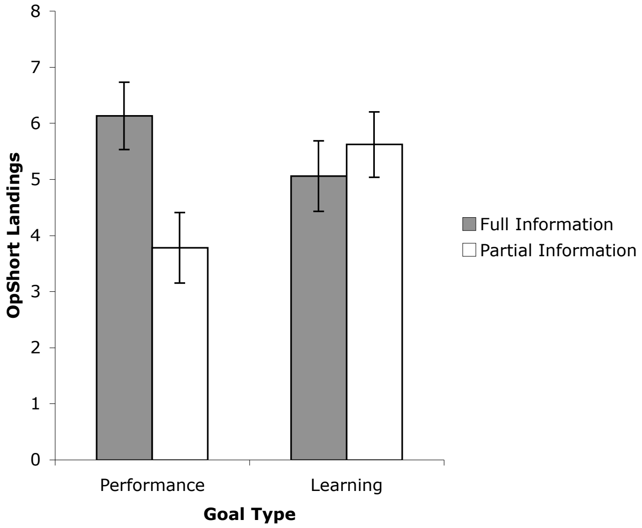


Figure 6: Mean OpShort landings (+/- 1 SE) by each group during the acquisition phase

Performance phase.

In the performance phase, all participants were instructed to maximise their point scores under conditions where scores were visible, no Rule-4 information was provided, and Rule-4 errors were penalised. It was anticipated that learning goal and partial information groups would show higher performance (Landings) and knowledge (OpShort) scores than performance goal or full information conditions (*Hypotheses 1b* and *2b*), reflecting the differences in rule space search in the acquisition phase. Differences in rule knowledge were also expected to be more pronounced for partial compared to full information conditions (*Hypothesis 3c*).

As shown in *Table 14*, no differences were observed between the groups for Landings, contrary to predictions. The result suggests that all groups were equally proficient at landing planes during the phase, despite anticipated knowledge differences. Knowledge differences may not have emerged in landings data however because participants were anticipated to differ only in their Rule-4 knowledge, and such knowledge was relevant only for approximately 14% of total landings that were OpShort. Whilst the results suggest that all groups were similarly proficient at landing planes during the phase, the measure may not have been sufficiently sensitive to detect differences in Rule-4 knowledge over the phase.

Table 14: ANOVA F tests for the three 2x2 (Goal Type x Information Level) Performance Phase analyses on Landings, Rule-4 errors and OpShort.

Factor	Landings	Rule-4 Errors	OpShort
Between-Subjects			
Goal Type	<1	<1	1.09
Information Level	<1	1.61	10.92**
Goal Type x Information Level	<1	1.42	5.95*
MSE	50.74	12.59	8.01

*Note: N= 75. All df = (1, 71). *p<.05, **p<.01.*

Significant differences were observed for OpShort, as shown in *Table 14*, suggesting that manipulations were successful at inducing knowledge differences between the groups. However, as shown in *Figure 7*, the direction of results was somewhat contrary to predictions. Contrary to *Hypothesis 1a,* OpShort was found to be similar between learning and performance goal conditions suggesting that the greater rule space search undertaken by learning goal groups did not produce superior learning in the performance phase. Moreover, full information groups showed higher OpShort than partial information conditions, directly contrary to *Hypothesis 2b*.

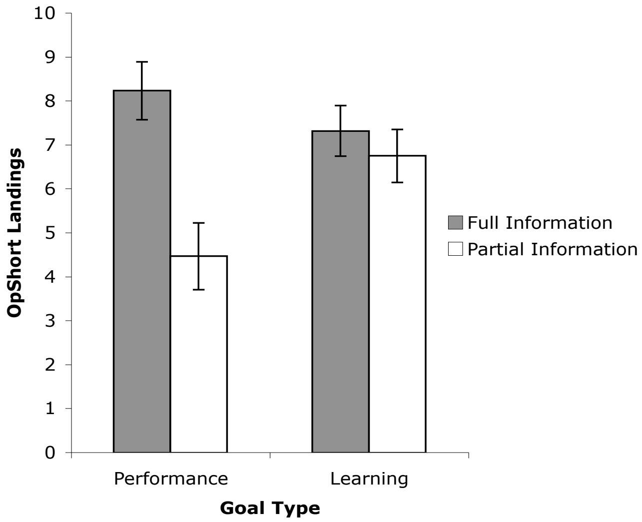


Figure 7: Mean number of OpShort landings (+/- 1 SE) for each group over the performance phase.

As shown in *Table 14*, a significant interaction was observed for OpShort suggesting that the influence of information level was not consistent for each goal type. The difference in OpShort between full and partial information groups appeared (Figure 7), to be significantly larger for performance goals as opposed to learning goals. This suggests that the aforementioned main effect for information level is attributable to this difference. To examine this possibility further, pairwise analyses were performed. Tests revealed that the Performance/Partial group performed significantly fewer OpShort landings than any other group across the phase (Performance/Full: (t(71)= 4.05, *p*<.01.; Learning/Full: (t(71)= 2.98, *p*<.01; Learning/Partial (t(71)=2.51, *p*=.01), and that all other groups performed similarly (all t(71)<1.66, *p*>.10). It would appear then that only the Performance/Partial, rather than both partial information groups, demonstrated a low level of knowledge during the performance phase. In sum, the negative influence of partial information conditions therefore appeared relevant only when combined with a performance goal.

The poor knowledge demonstrated by the Performance/Partial group was also consistent with *Hypothesis 3c* since the difference between the two partial information groups was larger than between the two full information groups. This result suggests, consistent with predictions, that the observed differences in rule space search in the acquisition phase, where the performance goal groups showed lower rule space search than those assigned learning goals, produced more pronounced differences in knowledge when only partial information was provided. Rule space search seemed particularly pertinent therefore, when participants had to discover information about the task rather than simply use information that had been provided.

The high knowledge demonstrated by the Performance/Full group was however contrary to *Hypothesis 3b*. The Performance/Full group was anticipated to demonstrate lower knowledge than both learning goal groups, but in particular the Learning/Partial group, due to their lower anticipated level of rule space in the acquisition phase. The high level of knowledge shown by the group despite their low rule space search during the acquisition phase suggests that rule space search may be of little benefit when complete task information is provided.

Although no specific predictions were made regarding Rule-4 errors, the universally low scores observed for all groups, as shown in Table 14, throughout the phase suggests that the introduction of error penalties effectively reduced errors to a minimum during the performance phase. Performance phase manipulations therefore appeared to be effective at discouraging rule space search by penalising errors.

Test phase

Task conditions were changed in the test phase to increase the importance of Rule-4 knowledge to performance. Conditions were consistent with the performance phase except proportionally more 747s were presented to force participants to perform more OpShort landings than previous phases. By forcing participants to perform landings that relied on Rule-4 knowledge, the conditions aimed to provide a stronger test of Rule-4 knowledge than had been achieved in the performance phase.

Table 15: ANOVA F tests for the three 2x2 (Goal Type x Information Level) Test Phase analyses on Landings, Rule-4 errors and OpShort.

Factor	Landings	Rule-4 Errors	OpShort
Between-Subjects			
Goal Type	<1	<1	<1
Information Level	<1	<1	2.72
Goal Type x Information Level	2.99	<1	4.77*
MSE	46.09	52.55	26.03

*Note: N= 75. All df = (1, 71). *p<.05, **p<.01.*

Like the performance phase, knowledge and performance scores were anticipated to be higher in learning goal and partial information groups as a result of greater rule space search undertaken during acquisition (*Hypotheses 1b* and *2b*). As shown in *Table 15* however, no significant differences between groups were observed for Landings although the interaction between goal type and information level did approach significance (p=.09). This suggests (*Figure 8*) that participants given manipulations that exerted a similar influence on goal and information level manipulations (i.e. the Performance/Full and Learning/Partial groups) developed somewhat better task knowledge and proficiency than those given contrary manipulations (i.e. Performance/Partial and Learning/Full groups). Consistent with the performance phase, the result again suggests that the Performance/Full group was able to acquire a high level of knowledge despite a limited search of rule space. Interestingly however, it also suggests that the high level of rule space shown by the Learning/Partial group may have been slightly beneficial in terms of Landings. Since results were not significant, a clear interpretation cannot be made.

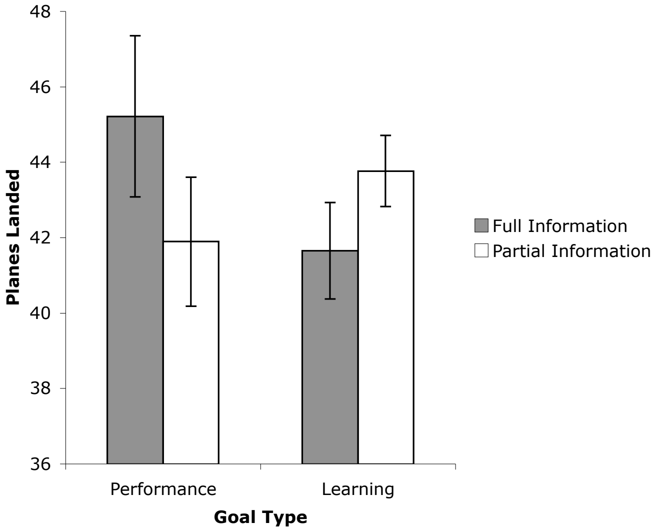


Figure 8: Mean number of planes landed (+/- 1 SE) for each group over the test phase.

As shown in *Table 15*, only the interaction achieved significance for OpShort. The pattern of OpShort results were in fact very similar to the performance phase, as shown in Figure 9, with differences appearing significantly larger between the two performance compared to the two learning goal groups. Whilst the pattern of results were similar, pairwise tests revealed only that the two performance goal groups differed with the Performance/Partial group showing lower OpShort than the Performance/Full group (t(71)= 2.70, *p*<.01), and no other comparisons achieving significance (t(71)<1.61, *p*>.11). The difference in OpShort was therefore not larger between two partial information groups as had been anticipated (*Hypothesis 3c*).

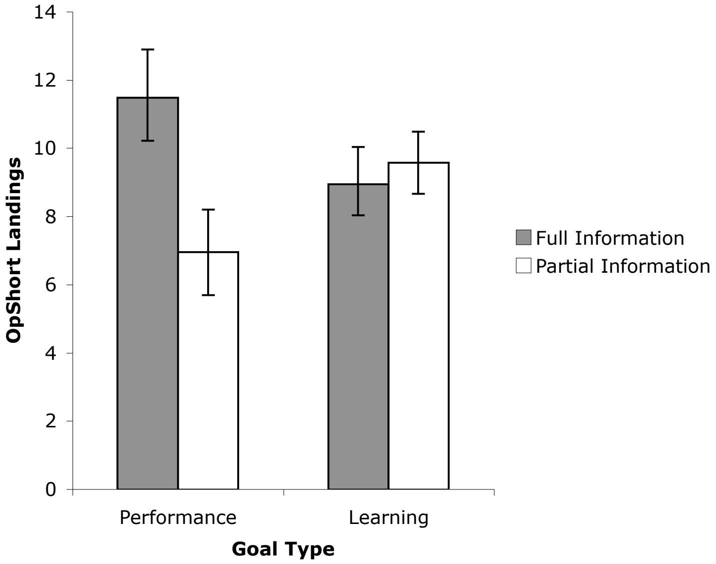


Figure 9: Mean number of OpShort landings (+/- 1 SE) for each group over the test phase.

As shown in Figure 9, the likely reason that the observed interaction for OpShort was not consistent with predictions was that the Performance/Full demonstrated higher, and the Learning/Partial group lower, OpShort than expected, contrary to *Hypothesis 3b*. It had been anticipated that the Learning/Partial group would develop a superior level of knowledge to the Performance/Full group due to their greater search of rule space during the acquisition phase. However, it would appear instead that the Performance/Full group developed a comparatively high level of knowledge in spite of a very limited rule space search. The result suggests that providing full rule information may effectively negate the benefits of rule space search.

Discussion

Overall, Study 2 aimed to provide a stronger comparison between CLT and DST than had been achieved in Study 1. The main prediction was that learning goal and partial information manipulations would elicit greater search of rule space search during initial learning relative to the performance goal and full information manipulations. This was in turn anticipated to promote better knowledge development and improve performance in subsequent phases, consistent with DST. Differences in cognitive load, whilst anticipated, were not expected to be related to differences in rule space search or the levels of knowledge participants developed, contrary to CLT.

Goal and information level manipulations.

Rule space search and learning outcomes. The influences of goal and information level manipulations on rule space search during the acquisition phase were consistent with predictions. Learning goals, intended to encourage participants to focus on learning the task rules, were found to produce greater rule space search than performance goals, which were instead expected to focus participants on performance outcomes. This difference was also particularly pronounced between Learning/Partial and Performance/Full groups, the two groups anticipated to show the highest and lowest rule space search respectively. This suggested that the combined influence of these manipulations produced particularly high and low levels of rule space search as anticipated. The Learning/Partial group also exhibited higher rule space search than the Performance/Partial group suggesting, as expected, that the performance goal manipulation suppressed rule space search, even when it was necessary to develop a full understanding of task rules. Although no difference in rule space search was observed between full and partial information conditions, contrary to predictions, rule space search was somewhat (though not significantly) higher for partial information groups, tentatively suggesting that providing incomplete task information may contribute, albeit slightly, to greater rule space search. However, it would appear that information level affected the outcome, rather than the extent of rule space search. The influences of manipulations on rule space search were therefore consistent with predictions and of DST.

In contrast to the rule space search findings however, results for learning outcomes were mixed. Consistent with predictions, and with their low level of rule space search, the Performance/Partial group developed a relatively poor level of task knowledge. This appeared due to the groups' low level of rule space search since the Learning/Partial group (who showed a high level of rule space search) developed a high level of knowledge despite also not being provided Rule-4. Cognitive load for the Learning/Partial group was also low, indicating that cognitive load was unlikely to have impeded knowledge acquisition. Interestingly, the Performance/Partial groups' relatively poor knowledge did not ameliorate with practice, with results suggesting that the group chose to avoid situations in which their knowledge was inadequate (i.e. OpShort landings), rather than search rule space to improve their knowledge. It seems that a performance goal, even if non-specific, may therefore dissuade learners from improving their knowledge by promoting avoidance of negative performance outcomes, consistent with Dweck (1988). In support of predictions then, assigning a performance goal when some relevant information had to be discovered appeared to have a deleterious effect on knowledge acquisition because of its limiting influence on rule space search. In addition, performance goals appeared to have successfully induced a focus on performance outcomes, i.e. task instances, despite use of a non-specific, rather than the more commonly used specific, performance goal (Burns &

Vollmeyer, 2002; Sweller, 1988). Consistent with DST, this suggests that it is likely an instance space focus, not just the focus on a specific goal, can impair learning.

Although learning outcomes for the Performance/Partial group were consistent with initial search of rule space, they appeared unrelated for other groups. For example, despite significant differences in rule space search, learning and performance goal groups demonstrated similar levels of task knowledge throughout the study. Even the Performance/Full and Learning/Partial groups, who differed most markedly in rule space search, demonstrated similar rule knowledge in the test phase. It would seem therefore that differences in rule space search did not, with the exception of the Performance/Partial group, produce anticipated differences in learning, possibly suggesting a dissociation between learning and rule space search.

The failure to observe a clear relationship between rule space search and knowledge appeared to be due largely to the high levels of knowledge achieved by the Performance/Full group, despite the group's low level of rule space search shown in the acquisition phase. Both learning goal groups for example developed high levels of knowledge, commensurate with their rule space search, but they were superior only to the Performance/Partial group. The high knowledge levels evidenced by the Performance/Full group therefore suggest that proving complete rule information may have negated the need for rule space search. Rather than suggesting that rule space search is unrelated to learning, the results may suggest that it is irrelevant when complete task information is provided. Rule space search may therefore be of benefit only when it can provide information that is not otherwise available.

In support of this interpretation, differences between learning and performance goal groups were found to be less pronounced under full compared to partial information conditions (at least in the acquisition and performance phases). For full information groups, there appeared to be negligible difference in task knowledge regardless of the pronounced differences in rule space search shown. Conversely, when partial information was provided, large differences were observed in knowledge that appeared directly related to rule space search. Participants provided partial information appeared to be more sensitive to the influence of goal type on rule space search than those provided full information. These results also suggest that rule space search may be beneficial only to the extent that it compensates for a lack of relevant task information, rather than providing any additional benefit.

The suggestion that rule space search is of benefit only when task performance is somewhat uncertain may also be attributable to the methodology of the present study rather than the inefficacy of rule space search per se. In particular, rule space search may not have produced an advantage due to ceiling effects or an inadequate training duration. The similarity of knowledge scores shown by the Learning/Partial, Learning/Full and Performance/Partial groups, despite different levels of rule space search, may, for example, have been due to all groups reaching a response ceiling. This is a distinct possibility given the knowledge measure (OpShort) was an aspect of task performance and was therefore constrained by the opportunities the task presented. Alternatively, the relatively short two-trial duration of the acquisition phase, the only phase in which rule space search manipulations were administered, may have limited the extent to which participants were able to search rule space, consequently curtailing their knowledge acquisition. This seems plausible given that rule space search likely takes a greater amount of time to generate knowledge compared to acquiring the same knowledge from information provided. Allowing a longer acquisition phase for participants to conduct a more thorough search of rule space may then elicit the anticipated superior levels of knowledge in high rule space search groups. Present methodology may therefore account for the failure to observe superior knowledge in groups who conducted greater rule space search. Both ceiling and training duration limits will therefore be addressed in the following chapter.

Cognitive load. Cognitive load was anticipated to be higher in partial information groups during the acquisition phase because greater exploration was required for these groups. This exploration was anticipated to place higher demands on working memory than simply learning rules from provided information (Sweller et al., 2011). Contrary to this prediction however, cognitive load was found to be substantially higher in the full information groups with load in the partial information groups effectively zero. Since it cannot be argued that the partial information groups experienced low cognitive load because they failed to learn task rules or conduct sufficient exploration (the Learning/Partial group clearly did both), the difference in load between the full and partial information groups was likely due to differences in the element interactivity of each manipulation.

Providing full rule information may represent additional task elements that a learner must process in conjunction with learning how to perform a novel task. This is consistent with CLT's suggestion that rule learning imposes a high cognitive load, rather than simplifying the task by guiding responses in situations of uncertainty. In other words, full information may make a task more complex by giving a learner more things to consider at a time when load is already high. This possibility seems plausible given the complex and contingent nature of the Rule-4 information that differentiated the full and partial information groups. Learning how to utilise such a rule was likely challenging and required additional cognitive resources to do effectively. That learning the same rule for the Learning/Partial group did not produce elevated load may be due to partial information conditions permitting learners to develop rule knowledge more slowly, at their own pace. Alternatively, consistent with CLT's explanation of split attention effects (Sweller et al., 1998; Tarmizi & Sweller, 1988), it may be due to the group learning rules from the same source in which it was utilised (i.e. the task), rather than having to translate rule information from text into action (as for full information groups). The elevated load evidenced by the full information groups is therefore consistent with CLT since full information conditions produced higher cognitive load than partial information manipulations.

The higher cognitive load observed for full compared to partial information groups does however suggest some dissociation between cognitive load and learning. Despite the high load, both full information groups developed high levels of task knowledge suggesting that their learning was unimpeded by the high load, contrary to CLT. Whilst it may be argued that cognitive load ameliorated with practice (Ackerman, 1988; Kanfer & Ackerman, 1989) allowing full information participants to acquire high levels of knowledge despite initially high cognitive load, that both groups demonstrated high rule knowledge during the initial stage of learning when cognitive load was elevated suggests otherwise. Whilst it may be that cognitive load, although elevated, was not sufficient to have overloaded working memory, the failure to observe any, even non-significant, decrement in knowledge, especially given the large size of the load difference between the groups, suggests that load did not influence knowledge acquisition. It would appear that in the current task, cognitive load did not influence knowledge acquisition, contrary to CLT[17].

The higher cognitive load observed for full compared to partial information groups also suggests, albeit less strongly, some dissociation between cognitive load and rule space search. CLT predicts rule space search to be low under conditions of

---

[17] It should be noted however that the high knowledge of the full information groups was also contrary to predictions of DST so neither theory was consistent with this result.

high cognitive load but rule space search appeared to be higher for the Learning/Full compared to Performance/Full groups. Whilst this difference in rule space search was not significant, that the difference approached significance with almost no difference observed in cognitive load experienced suggests, albeit tentatively, that load and rule space search may vary independently. It should be noted however that there was a significant negative correlation observed between rule space search and cognitive load overall suggesting that, on average, that load was inversely related to cognitive load, consistent with CLT.

Theoretical implications.

Regarding the success with which the study achieved its aim of separating CLT and DST, results were mixed. On the one hand, rule space search appeared to vary independently of cognitive load since both learning goal groups showed high rule space search under conditions of different cognitive load. In addition, full information groups showed similarly high cognitive load but different levels of rule space search. Cognitive load and rule space search therefore appeared to vary independently. On the other hand, cognitive load was negatively related to rule space search overall suggesting, consistent with CLT, that rule space search and cognitive load were related and inversely proportional. Results were therefore equivocal.

Although direct evidence was inconsistent, results were generally more supportive of DST than CLT suggesting some separability of the theories. In the acquisition phase, patterns of rule space were directly consistent with DST-based predictions with higher learning shown in learning goal groups and, to a lesser (non-significant) extent, partial information groups. As well, in partial information groups, the extent of rule space search was directly related to learning outcomes, in direct support of the theory. Whilst learning outcomes for the full information groups were not consistent with predictions, in that differences in rule space search did not produce differences in knowledge, this may be due to limitations in design, rather than a disconfirmation of the theory. If rule space search is indeed effective only to the extent that it provides information that is not otherwise available, then results are almost entirely consistent with DST.

Conversely, whilst CLT can account for the observed higher cognitive load in full information groups, these differences appeared largely unrelated to learning outcomes. High cognitive load was evidenced in groups that showed both high and low learning and low cognitive load in groups that, similarly, showed high and low learning. This suggests that cognitive load had little influence on the level of knowledge participants acquired. On balance therefore, it seems DST is more consistent with the present results.

A second implication of the study consistent with DST is that initial differences in learning did not appear to ameliorate with practice, contrary to CLT. The present study clearly indicates that initial learning conditions may have lasting effects on knowledge acquisition despite repeated and prolonged task practice. Poorer knowledge levels were observed for the Performance/Partial group throughout the study. If their poor performance was due to cognitive load, they should have reduced their disadvantage with practice, as cognitive load declined (although load was low for the group initially). Their persistently poor performance suggests instead that the group focused on maximising their performance in instance space and avoided making errors that would have detracted from this goal. Whilst such an approach certainly minimised errors, it also suggests that the participants did not focus on acquiring rule knowledge since errors were the only way of acquiring rule knowledge. It seems likely that the group's focus on performance outcomes in instance space meant that they avoided rule space search and consequently did not acquire a high level of rule knowledge despite their prolonged practice. Consistent with DST, continued encouragement to focus on task instances appeared to prevent rule learning, potentially indefinitely.

Summary

Overall Study 2 presented a stronger test of the independence of CLT and DST than had been achieved in Study 1. Whilst direct comparisons of cognitive load and rule space search were mixed, making separation of the theories difficult, learning outcomes were more consistent with DST suggesting, albeit weakly, that the theories are separable. The following study will attempt to address the limitations of the present study and provide a more conclusive comparison of CLT and DST.

CHAPTER 6: STUDY 3 -

A FURTHER INVESTIGATION OF THE INFLUENCE OF BOTH RULE SPACE SEARCH AND COGNITVE LOAD DURING LEARNING

Study 2 provided preliminary evidence that DST and CLT are separable explanations of learning. In particular, performance and learning goal groups demonstrated different levels of rule space search despite showing no differences in cognitive load overall. Further, whilst those who received full rule information during training experienced higher cognitive load, this appeared unrelated to their level of rule space search since groups were found to demonstrate both high and low rule space search under these conditions. Finally, the persistent effects of task manipulations on knowledge differences were more supportive of DST than CLT. However, contrary to these findings, cognitive load was, overall, negatively correlated with rule space search and rule learning appeared to be unrelated to both cognitive load and rule space search when full information was provided. Further research was therefore required to address these somewhat inconsistent findings.

The present study sought to resolve the inconsistent findings of Study 2, as well as replicate its consistent findings, by conducting a more detailed and more strictly controlled investigation of the influences of rule space search and cognitive load on learning. Of specific focus was whether greater rule space search during the initial stages of learning would necessarily lead to better knowledge and performance, and whether high cognitive load during training could be shown to impede learning. To achieve these aims, three modifications were made to the design of Study 2. First, the duration of the acquisition phase was increased to allow greater time for rule space search; second, the accuracy of knowledge assessment was improved to provide a better indication of knowledge differences between groups; third, performance variability not attributable to manipulations was reduced to improve the salience of between group differences in learning and cognitive load. These modifications are explained in detail below.

Modification 1: Increased the duration of the acquisition phase

One of the primary aims of Study 3 was to establish whether higher levels of rule space search would, regardless of task conditions, provide a clear advantage to both knowledge and performance outcomes. According to DST, the more time a learner engages in rule space search during learning, the greater their acquisition of the task's underlying rule structure and the better their knowledge and performance. In Study 2 however, the only groups to demonstrate differences in knowledge commensurate with rule space search were the two partial information groups. For the two full information groups, knowledge scores were similarly high despite the performance goal group showing low, and the learning group high, levels of rule space search. By differentiating only partial information groups, the result suggested that rule space search may be advantageous only to the extent that it provides knowledge not otherwise obtainable from the task or instructions, and may therefore be redundant when complete task information is readily available.

The lack of a consistent relationship between rule space search and learning, especially for the full information groups, may however not be due simply to its utility; the length of time available to conduct rule space search would also appear likely to influence its effectiveness. Rule space search requires repeated testing and retesting of task-related hypotheses and is therefore likely to take longer to generate relevant and accurate task knowledge than using a list of provided information. Using provided information as a basis for rule space search, as would have been the case for the Learning/Full group in Study 2, would also likely have taken longer than using provided information to perform the task, as was the case for the Performance/Full group. Permitting too short a duration for rule space search may then limit its effectiveness because it may end rule space search prematurely. Since Study 2 provided only two trials where rule space search was encouraged, this may not have allowed sufficient time for rule space search to produce knowledge and performance advantages. Rather than suggesting that rule space search is redundant when full information is available, the previous results may instead indicate that training duration was too short to allow a search of rule space sufficient to elicit advantages in knowledge and performance.

To establish whether rule space could indeed confer an advantage on learners, even when full information was provided, the present study extended the duration of the acquisition phase. This was anticipated to allow a greater opportunity for participants to search rule space, thus providing a more thorough test of the influence of rule space search on subsequent knowledge acquisition and performance.

Modification 2: Improving the accuracy of knowledge assessment

Failure to observe differences between the two full information groups, or a clear knowledge advantage for those who engaged in greater rule space search, may also have been due to the way in which task knowledge was assessed in Study 2. In the previous study, knowledge was operationalised in terms of OpShort landings: the complex, weather-contingent landings of mid-sized aircraft on short runways, particularly under the more difficult conditions of the single test phase trial. Assessing knowledge in this way raises two potential concerns as to the accuracy of the measure.

First, as a component of overall performance, OpShort potentially confounds knowledge with proficiency. The rate participants landed planes in the test phase may well have influenced their rate of OpShort landings independent of their level of task knowledge. It is possible, for example, that the high knowledge exhibited by the Performance/Full group was not wholly indicative of their task knowledge, but attributable, at least in part, to their faster overall rate of landing planes. The high knowledge achieved by this group, despite its low rule space search and high cognitive load, may therefore have been spurious.

To overcome this potential confound, and provide a more accurate assessment of task knowledge, a separate knowledge measure was introduced in Study 3. This measure provided a performance-independent assessment of task knowledge that was insensitive to any systematic between-group differences in task proficiency.

The second potential flaw in the previous study's measurement of knowledge was the use of only a single trial for knowledge assessment. Using only one trial, where task conditions had been changed, may have biased results towards the participants who were most able to adapt their performance to task changes. As demonstrated by Schunn & Reder (2000), individuals differ in the speed and extent to which they can adjust to task changes, with such individual variation likely to be independent of differences in task knowledge. Presenting only a single trial to test knowledge may therefore have advantaged participants with better adaptive abilities rather than those with higher levels of task knowledge.

Advantaging higher adaptive ability participants would not have been problematic if the ability to adapt was evenly distributed across the groups, as would be expected with random allocation of participants. However, participant's adaptive abilities may have been disproportionately affected by the different experimental manipulations. Many researchers have noted that increasing the variability of task training can improve the effectiveness with which individuals respond to later task changes (Catrambone & Holyoak, 1990; Gick & Holyoak, 1983; Kimball & Holyoak, 2000; Novick & Holyoak, 1991; Reder, Charney, & Morgan, 1986; Singley & Anderson, 1989). In Study 2, the variability of training differed between the groups because the transition between acquisition and performance phases was not uniform. For the Performance/Partial group, for example, acquisition and performance phases were identical because the performance goal and partial information conditions they experienced during acquisition were the same conditions all participants were provided during the performance phase. Since all other groups experienced at least some change in task conditions between acquisition and performance phases, this group's poor knowledge development could therefore attributable to the lack of variability between acquisition and performance phases, rather than the group manipulations themselves. Differences in the changes experienced the groups could therefore have advantaged (or disadvantaged) some groups more than others.

To ensure that the test phase represented a test of task knowledge, and not adaptive abilites, the duration of the test phase was extended. The longer duration was anticipated to provide all participants sufficient time to adapt to task changes and achieve a level of performance that was commensurate with their knowledge. In addition to the separate knowledge assessment, the longer test phase sought to improve the accuracy of the knowledge measures in the present study to better distinguish between groups.

Modification 3: Minimising unsystematic variation in task performance

Failure to observe anticipated differences in both task and load measures may also have been attributable to a high level of individual variability in task performance caused by way task variables were generated. In Study 2, the plane and weather condition sequences shown in each trial were generated randomly by task software. For any given trial or participant, the sequence of planes and weather conditions could vary markedly, potentially providing harder or easier task conditions independent of experimental manipulations. Some participants could, for example, have been presented with milder weather conditions during the test phase thus improving their results independently of their task knowledge.

Whilst the random nature of the sequence variation likely ensured task difficulty was equivalent across groups, it also increased unsystematic variation to knowledge and performance scores. This may potentially have obscured between-group differences. The absence of anticipated differences in knowledge, performance, and cognitive load, in Study 2 may therefore have been attributable to this extraneous task variability.

In order to better control task conditions and minimise unwanted variability in knowledge and performance scores, the present study fixed the generation of planes and weather conditions so that all participants experienced identical sequences on equivalent trials. In other words, whilst the sequences differed between task trials, the specific sequence for trial one, two, three, etc was the same for all participants. This change was intended to ensure that any differences in knowledge and performance scores attributable to experimental manipulations, rule space search, and cognitive load, were accurately detected.

Hypotheses

The modifications to the present study did not alter the fundamental design of Study 2 but instead aimed to improve it with the view to achieving stronger support for the hypotheses. Consistent with Study 2 then, the present study manipulated goal type and level of information during training in a 2 x 2 factorial design across three phases: acquisition, performance, and test. Manipulations were again administered during the initial acquisition phase and were intended to influence participants' search of rule space and cognitive load to produce differences in the level of knowledge participants developed.

Goal manipulations were anticipated to influence the level of rule space search during the initial acquisition phase. Learning goals were expected to encourage greater focus on learning the task rules, and less on maximising performance, and were therefore anticipated to elicit greater search of rule space than performance goals, consistent with previous results.

Since the acquisition phase was longer in the present study, analyses were also able to examine whether learning goals also elicited a different pattern of rule space search than performance goals over the course of the phase. Previous research has suggested that goal type may also influence the duration, not simply the level, of rule space search during learning. Burns and Vollmeyer (Burns & Vollmeyer, 2002; Vollmeyer & Burns, 2002; Vollmeyer et al., 1996) demonstrated that non-specific performance goals can encourage more persistent and extended rule space search during learning relative to specific goals, which instead encourage learners to prematurely switch from searching rule space to trying to maximise performance. Since learning goals were anticipated to elicit greater rule space search, like non-specific goals, it seems plausible that they would also prolong the duration of rule space search during the acquisition phase. Similarly, if performance goal manipulations reduce rule space search, it also seems reasonable to suggest that they would do so by prematurely encouraging participants to switch from an exploratory to performance focus. The lower performance focus of the learning goal groups was therefore anticipated to not simply encourage higher levels of rule space search but also a more persistent pattern of task exploration throughout the acquisition phase. Performance goal groups were instead anticipated to show a rapid decline in rule space search indicative of an earlier switch from an exploration to performance focus. These predictions are summarised in Hypotheses 1a and 1b below.

Hypothesis 1a: Learning goal groups will demonstrate higher levels of task exploration, i.e. rule space search, during the acquisition phase than performance goal groups.

Hypothesis 1b: Learning goal groups will demonstrate more persistent levels of task exploration during the acquisition phase than performance goal groups. Performance goal groups are instead anticipated to show a more rapid decline in rule space search across the acquisition phase.

The amount of rule information provided to participants was also anticipated to influence rule space search during the acquisition phase. In Study 2, rule space search was anticipated to be higher for both partial information groups due to their need to explore the task to discover information that had not been provided. However, results suggested that rather than universally encourage greater rule space search, the influence of partial information conditions depended on goal type, encouraging high levels of rule space search when learning goals were assigned but appearing to be ineffective when combined with performance goals. Performance goals appeared, in effect, to suppress any positive influence of partial information conditions on rule space search. Despite the prolonged duration of the acquisition phase in the present study, present predictions were consistent with the findings of Study 2. In combination with learning goals, the partial information conditions were anticipated to elevate rule space search relative to other groups. However, when combined with performance goals, partial information conditions were anticipated to have a limited influence on rule space search. This was because performance goals were likely to suppress rule space search during the acquisition phase. These predictions are summarised in Hypotheses 2a and 2b below.

Hypothesis 2a: Partial information conditions will have no overall effect on rule space search during the acquisition phase but will influence rule space search differently depending on the goal type assigned. Differences in rule space search will be greater for groups provided partial, compared to full, rule information.

Hypothesis 2b: When combined with a learning goal, provided partial information conditions will elicit a greater, and more persistent pattern of rule space search over the acquisition phase than other groups.

Manipulation of level of information was also intended to affect the level of cognitive load experienced by participants during the acquisition phase when the attentional demands of the task were at their highest (Anderson, 1982; Kanfer & Ackerman, 1989) and task conditions differed between the groups. In Study 2 it was anticipated that greater task exploration would increase cognitive load on participants since exploration involves processing a greater number of cognitive elements than using provided rules or examples. Results however suggested that load was higher in the full information groups during the acquisition phase, particularly when combined with a performance goal, suggesting the provision of complete rule information may increase cognitive load since participants are initially provided a larger amount of information to learn rather than learning at their own pace through task experience. Given the contrary nature of predictions and findings of the previous study, no direction is anticipated for the influence of information manipulations on cognitive load because either can be accounted for by CLT. Cognitive load between the full and partial information groups is therefore only anticipated to differ during the acquisition phase. However, contrary to CLT, whatever the influence of information-level manipulations on cognitive load, they are not anticipated to influence learning. Groups that experience high cognitive load during the acquisition phase, for example, are not anticipated to show lower knowledge or performance scores than those of other groups during the subsequent phases. These predictions are summarised in hypothesis 3 below.

Hypothesis 3: Cognitive load is anticipated to differ between full and partial information groups during the acquisition phase but the differences are not anticipated to influence learning, contrary to CLT. No differences in performance or knowledge in subsequent phases are anticipated as a result of load differences during acquisition.

In the performance and test phases, knowledge and performance scores were anticipated to largely reflect the amount of rule space search undertaken during the acquisition phase. Groups who had undertaken greater amounts of rule space search were therefore anticipated to achieve higher scores, particularly under the more challenging test phase conditions. Whilst this was not observed in Study 2, the extended duration acquisition in the present study was anticipated to provide a greater opportunity for rule space search, thus increasing its likely influence on later performance. The greater rule space search anticipated in learning goal, relative to performance goal, groups was expected to produce higher knowledge and performance scores in both the performance and test phases. Unlike Study 2, higher scores were expected in both full and partial learning goal groups since both were anticipated to show greater rule space search relative to their performance goal counterparts. Given the dual rule space search incentives provided the Learning/Partial group however, this group was expected to show the highest overall knowledge and performance scores in both performance and test phases. These predictions are summarised in Hypotheses 4a and 4b below.

Hypothesis 4a: Learning goal groups will achieve higher knowledge and performance scores than performance goal groups in both the performance and test phases. This will be observed for both learning goal groups relative to their performance goal counterparts.

Hypothesis 4b: The Learning/Partial group will achieve the highest performance and knowledge scores in the performance and test phases due to its relatively high level of rule space search undertaken during the acquisition phase.

Although partial information conditions were intended to elevate rule space search during the acquisition phase, no knowledge or performance differences were anticipated between the full and partial information groups in later phases. Instead, consistent with Study 2, information level was anticipated to influence participants differently depending on the type of goal they were assigned. Partial information conditions were anticipated to increase rule space search more when combined with learning, as opposed to performance, goals because they were likely to encourage, rather than suppress, the incentive to search rule space during the acquisition phase. Although some elevation in rule space search was anticipated for the Performance/Partial group, it was unlikely that the limited search of rule space would be sufficient to acquire the rule information that they had not been provided. Despite some limited rule space search then, the Performance/Partial group was anticipated to display much lower levels of knowledge and performance than the Learning/Partial group, who was anticipated to show relatively high levels of both. Full information groups were anticipated to differ less markedly than the two partial information groups since provision of full information was anticipated to ameliorate, to a large extent, knowledge differences caused by differences in rule space search. Although the Learning/Full group was anticipated to demonstrate higher knowledge and performance scores than the Performance/Full group, the difference was not anticipated to be as pronounced as the two partial information conditions. These predictions are summarised in Hypothesises 5a and 5b below.

Hypothesis 5a: No overall differences in knowledge or performance scores will be observed between full and partial information groups in the performance or test phases.

Hypothesis 5b: In the performance and test phases, full and partial information conditions will influence participants differently depending on their assigned goal type. Greater knowledge and performance differences are anticipated between the two partial information groups than the two full information groups, with the provision of partial information benefiting those assigned a learning goal but impairing those assigned a performance goal.

Summary

The present study sought to provide a more comprehensive examination of the influences of rule space search and cognitive load in the context of skill acquisition than had been achieved in Study 2. To this end, changes were made to the previous design that increased the opportunity for, and therefore potential influence of, rule space search during the acquisition phase. Changes were also made to improve the assessment of task knowledge and reduce unsystematic variation in task performance to improve the accuracy of knowledge and performance measures.

Consistent with DST, learning goals were anticipated to elicit greater rule space search than performance goals during the acquisition phase, particularly in combination with partial information conditions. This was anticipated to produce higher knowledge and performance scores in the later phases. Performance goal groups were anticipated to demonstrate lower rule space search during the acquisition phase and consequently poorer knowledge and performance in the later phases, particularly for the group provided partial information conditions. Differences in cognitive load were anticipated between full and partial information groups but were not anticipated to influence learning, counter to CLT. Overall, Study 3 sought to provide a more comprehensive examination of the influences of cognitive load and rule space search on learning in the context of skill acquisition.

Method

Participants

One hundred and twelve (59.8% female) first and second year undergraduate students aged between 19 and 41 years (*M* = 20.1) participated for course credit or payment. Thirty-two second year students from the Economics and Business faculty participated for payment of $40 each. The remaining eighty participants were first year psychology students who participated for course credit. 5 participants were excluded from analysis: two paid participants experienced computer failures that compromised their data and three unpaid participants were 2.5 standard deviations below their respective group means in performance for every trial suggesting they were not trying. A final sample of 107 was considered for analysis.

Dependent measures

Task measures: Consistent with previous studies, the primary data consisted of the number of planes landed (Landings), the number of incorrect short runway landing attempts (Rule-4 errors), and the number of successful short runway landings of mid-sized planes (OpShort), at the completion of each trial. Landings were considered an indicator of overall task performance and proficiency, Rule-4 errors an indicator of task exploration (in the acquisition phase) and knowledge (in the performance and test phases), and OpShort an indicator of Rule-4 knowledge throughout training.

Task knowledge: In addition to OpShort, a 25 item, computer-based, task knowledge questionnaire was developed to test participant's knowledge of Rule 4. The questionnaire presented a screenshot of the KA-ATC task where a single plane in level 1 had been selected (other planes were present in the holding pattern, but were not selected). Participants were presented with four response options, corresponding to the four runways present in the task, and were required to select the runway(s) on which the plane could land given the weather conditions shown in the screenshot. Participants were instructed to select multiple runways if the plane could use more than one runway given the conditions and 15 seconds was allowed for each question (a count-down timer was displayed on the screen to indicate time remaining). Screenshots were selected so as to provide a test as closely resembling actual performance as possible.

Since the aim of the knowledge questionnaire was to ascertain the level of rule 4 knowledge participants had acquired, the questionnaire was biased towards mid-sized planes. In 18 of the 25 items, 727s and DC10s were the selected plane types (9 each), with each of these items corresponding to one of the nine weather combinations possible in the task (i.e. each mid-sized plane was shown with every one of the nine possible task weather combinations). Of the remaining items, 747s were selected in three and Props in four. Items were scored correct if the participant selected the maximum number of correct runway(s) for the plane, i.e. items were not scored correct if a participant selected only 1 runway when a plane could have used one of two runways.

Cognitive ability. Participants completed two measures of cognitive ability in the present study. Consistent with the previous studies, participants completed the APM to assess general cognitive abilities. Unlike Study 2 (Chapter 5) where the full 36-item version was administered, participants in Study 3 performed a 20-item computer-based short form of the test (Raven, 1993). This was identical to the version used in Study 1 (Chapter 4). The short form was chosen in place of the full version to save time due to time constraints on testing and the extended duration of training and test phases. Short forms of the APM measure have been used extensively in research to save time whilst maintaining the psychometric properties of the test (e.g. Arthur, Tubre, Paul, & Sanchez-Ku, 1999; Jaeggi, Studer-Luethi, Buschkuehl, & Su, 2010). A 40-minute time limit was imposed in order to keep testing time to a minimum and limit the possibility of ceiling effects (Jaeggi et al, 2010).

The second cognitive ability measure was the same complex span working memory task employed in Study 2 (Unsworth & Engle, 2007). The Noughts and Crosses test employed in the preceding two studies (Mackintosh & Bennett, 2003) was dropped to save time and since it provided a similar assessment of working memory as the complex span task but with less desirable psychometric properties.

Design

Consistent with Study 2, and as shown in Table 16, the research utilised a 2(Goal Type) x 2(Level of Information) between subjects factorial design (N= 107). The two goal types were again performance or learning goals and level of information conditions were full and partial.

Like Study 2, performance was broken down into three distinct phases: training, performance, and test. Manipulations in each phase were identical to Study 2 with goal and information manipulations administered during training and performance goal, partial information conditions administered thereafter. In the present study however, participants completed three trials in each phase, as shown in Table 17. Participants thus completed 9 trials of practice on the task.

Table 16: Experiment design including number of participants in each group.

		Rules	
		Full	Partial
Goal Type	Performance	Group 1 (24)	Group 2 (29)
	Learning	Group 3 (27)	Group 4 (27)

Procedure

Participants completed participation over two days, not more than one day apart, in groups of between four and eight. Participants began the first session with task instructions. Instruction script, presentation, and timing were identical to those employed in Study 2.

Following instructions, participants logged into the KA-ATC task and completed six trials: three under training phase conditions and three under the performance goal, partial information conditions of the performance phase. At the commencement of each trial an information screen was again shown that displayed the set of rules (appropriate to the participant's group and phase of practice) as well as the response keys needed to play the game. Participants were allowed to take short breaks between trials if desired. At the completion of the sixth trial, participants were asked to take a short break whilst the experimenter loaded onto their computers the knowledge questionnaire and the test-phase version of the KA-ATC task. Participants were not permitted to speak with others about the task during this time. When participants returned, they completed the task knowledge questionnaire and 2 trials task under test phase conditions. Before commencing the test phase trials, participants were told that the task was a "slightly different version to the one [they] had previously played" but were not told of any specific changes to the task. At the completion of the two trials, participants left and were reminded to return at their allocated time.

Session two began with participants completing the final test phase trial of the task. Participants then completed the APM and Complex Span measures. The final trial of the transfer phase was conducted between 24 and 48 hours after completion of the eighth trial in order to assess performance differences after a time delay.

Table 17. Outline of experimental procedure.

Manipulations were administered during instructions and during the first 3 trials (acquisition phase). Subsequently, all participants experienced the same conditions for the remainder of the trials

	Acquisition Phase			Performance Phase			Test Phase		
	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Trial 7	Trial 8	Trial 9
Group 1:	Performance Goal/ Full Rules			Maximise performance score. No Rule-4 given			Task difficulty increased		Completed on Day 2.
Group 2:	Performance Goal/ Partial Rules								
Group 3:	Learning Goal/ Full Rules								
Group 4:	Learning Goal / Partial Rules								

Results

Analyses

To test the effects of between group manipulations mixed 2x2x(3) (goal type x information level x trial) ANOVAs were conducted for each phase of task performance. For each phase, one ANOVA was carried out for each of the task measures: Landings, Rule-4 errors, and OpShort. For significant within-subject effects, additional trend analyses were conducted to determine the nature of differences. For between-subject effects, further pairwise tests were carried out to clarify any ambiguous results. ANOVAs were the same for all three experimental phases.

To determine whether groups differed in terms of cognitive load, moderated regression analyses were carried out for each phase of the experiment. This was achieved using the procedure outlined in (Baron & Kenny, 1986). The analyses compared the relative strength of the relationship between cognitive ability and performance and knowledge measures (landings and OpShort) between goal type and information level groups.

Manipulation check

In order to ensure that the three dependent measures, Landings, Rule-4 errors, and OpShort, were in fact indicative of the intended constructs, correlations between the measures were first analysed. The number of Landings made during each trial was anticipated to represent overall performance, Rule-4 errors both task exploration (during the training phase) and task knowledge (during performance and test phases), and OpShort the level of participants' Rule-4 knowledge. The relationships between Landings, Rule-4 errors, and OpShort shown in Table 18 generally support for these conceptualisations.

Table 18: Means, standard deviations, and intercorrelations for performance measures by phase.

	M	*SD*	1	2	3	4	5	6	7	8	9
Landings											
1.Acquisition	41.24	7.74	-								
2.Performance	52.27	7.27	.76**	-							
3.Test	46.93	6.11	.64**	.85**	-						
Rule-4 Errors											
4. Acquisition	7.68	4.79	.06	.26**	.22*	-					
5.Performance	6.48	4.42	.01	.11	.15	.59**	-				
6.Test	8.02	6.56	-.09	.05	.06	.56**	.75**	-			
OpShort											
7.Acquisition	19.86	9.10	.66**	.66**	.64**	.39**	.28**	.16	-		
8.Performance	25.35	10.53	.53**	.74**	.67**	.42**	.32**	.31**	.80**	-	
9.Test	36.65	14.08	.49**	.70**	.85**	.27**	.30**	.30**	.67**	.76**	-

*Note: N=107. *Correlation is significant at .05, **Correlation is significant at .01*

If Rule-4 errors were indicative of attempts to gain Rule-4 knowledge during the acquisition phase, the number of errors during acquisition phase should be positively correlated with knowledge and performance measures from later phases. As shown in Table 18, the number of Rule-4 errors made in the acquisition phase was positively correlated with Landings in both the Performance and Test phases indicating that short-runway landing errors in early training were associated with better knowledge and performance later in task performance.

Making Rule-4 errors should have been more beneficial however, for participants in the partial information groups since this was the only means by which they could acquire the knowledge that they had not been provided. To determine whether the overall positive relationship between Rule-4 errors in the acquisition phase and later measures of knowledge and performance was stronger for partial information groups, four moderated regressions (Baron & Kenny, 1986) were conducted on Landings and OpShort measures from both the performance and test phases. As shown in Table 19 however, no differences between goal-type or information-level groups were observed in the relationship between Rule-4 errors and later knowledge or performance. Making Rule-4 errors during the initial phase of learning therefore appeared similarly beneficial to all groups, regardless of experimental manipulations.

Table 19: Results of the four moderated regression analyses predicting Landings and OpShort for the performance and test phases.

Analyses examined whether the relationship between Rule-4 errors in training and later performance differed according to group

	Performance Phase				Test Phase			
	Landings		OpShort		Landings		OpShort	
Predictor variable	β	ΔR^2	β	ΔR^2	β	ΔR^2	β	ΔR^2
Step 1		.07**		.18		.05*		.08**
Rule-4 Errors (Training)	.26		.42**		.22*		.27**	
Step 2		.06*		.01		.08*		.03
Goal Type	.20*		.08		.27**		.17	
Information Level	-.12		<-.01		-.06		-.04	
Step 3		.01		.01		.01		.03
Goal Type x Rule-4 errors	.13		.08		.10		.11	
Information Level x Rule-4 errors	.01		-.09		-.03		-.33#	
Overall R		.36		.44		.36		.37
Overall R^2		.13		.19		.13		.14
Adjusted R^2		.09		.15		.09		.01
Overall F (5, 101)		3.05*		4.73**		3.07*		3.25**

*Note: N= 75. *p<.05, **p<.01*

Unlike the acquisition phase, Rule-4 errors were penalised in the performance and test phases as a disincentive to attempt short runway landings when unsure of success. Rule-4 errors in the performance and test phases were therefore anticipated to be indicative of a lack of Rule-4 knowledge, and therefore negatively correlated with OpShort. Contrary to predictions however, Rule-4 errors were correlated positively with OpShort in the performance and test phases, as shown in Table 18. The strongly positive relationships likely suggests that both OpShort and Rule-4 errors were indicative of participants' overall propensity to attempt OpShort landings, with a greater number of errors associated with a greater number of attempts. Given that the positive correlation was counter to predictions however, Rule-4 errors in the performance and test phases were not interpreted as indicators of poor knowledge in subsequent analyses.

Changes made to the task during the test phase were designed to increase task difficulty by forcing participants to make greater use of the short runways, particularly for mid-sized plane landings (OpShort). This was intended to more strongly assess participants' rule-4 knowledge than previous phases. To determine whether this manipulation was successful, the proportion of total Landings comprising of OpShort was compared between the test and preceding phases. As shown in Table 18, correlations between OpShort and overall Landings were stronger in the test phase than either preceding phase, observations supported by Fisher's z comparisons (Acquisition Phase: $z = 3.34$, $p<.01$; Performance Phase: $z = 2.2$, $p<.05$). OpShort landings thus formed a greater proportion of overall landings in the test phase as expected. This supported consideration of test phase performance (both Landings and OpShort landings) as stronger indicators of Rule-4 knowledge than previous trials.

Acquisition phase:

The first hypothesis concerning the acquisition phase was that learning goal groups would exhibit greater (*Hypothesis 1a*) and more persistent (*Hypothesis 1b*) task exploration, evidenced by higher Rule-4 errors, than groups provided performance goals. As shown in Table 20, between-subject comparisons of Rule-4 error scores across the phase did not reveal any significant differences between learning and performance goal groups, contrary to *Hypothesis 1a*, but within-subject comparisons did reveal differences in the pattern of Rule-4 errors according to goal-type across the phase. Trend analyses of the within-subject result revealed that participants generally reduced Rule-4 errors over the training phase (linear: F(1,103) = 13.62, p<.01) but the pattern of reduction differed according to goal type (goal type x quadratic: F(1,103) = 6.19, p<.05). As shown in Figure 10, the performance goal groups showed a steady decline in Rule-4 errors over the training phase. Learning goal groups however showed a small increase in Rule-4 errors in the second trial before a reduction in the third, suggesting that Learning Goal groups, particularly the Learning/Partial group, maintained exploratory behaviour for longer during the training phase than performance goal groups. This is in direct support of *Hypothesis 1b* and suggests that learning goal manipulations encouraged a more persistent, if not greater, search of rule space than performance goal groups during the acquisition phase.

Table 20: F statistics for the three acquisition phase 2x2x(3) Goal Type x Information Level x (Trial) ANOVAs.

Factor	Landings	Rule-4 Errors	OpShort
Within-Subjects			
Trial	294.28**	11.11**	131.95*
Trial x Goal Type	7.33**	3.62*	0.10
Trial x Information Level	.11	0.17	2.01
Trial x Goal Type x Information Level	.54	1.23	0.57
MSE (Trial)	198	13.46	6.58
Between-Subjects			
Goal Type	6.65*	1.42	.76
Information Level	3.11	3.40	.82
Goal Type x Information Level	2.50	2.79	9.41**
MSE	163.96	66.10	25.98

*Note: N= 107. Within-subjects df= (2, 206). Between-subjects df = (1, 103). *p<.05, **p<.01.*

Hypothesis 2a predicted that although information level would have no direct influence on rule space search during the acquisition phase, it would interact with goal type such that a larger difference in rule space search was anticipated between the two partial, compared to two full, information groups. Partial information was anticipated to encourage rule space search but its influence was expected to be suppressed by the score-focus elicited by performance goals, thus promoting high and low rule space search in the Learning/Partial and Performance/Partial groups respectively. Although the Learning/Partial group showed relatively high and persistent levels of rule space search during the acquisition phase and the Performance/Partial group showed low and declining levels, as shown in Figure 10, the interaction between goal type and information level was not significant, as shown in Table 20. Given the apparent difference shown in Figure 10 however, further pairwise comparisons were conducted, suggesting, tentatively[18], that Learning/Partial group committed more Rule-4 errors than the Performance/Partial group over the acquisition phase (Performance/Partial: t(103) = 2.05, p<.05) in support of *Hypothesis 2a.* The learning/Partial group also showed higher rule space search than the other

---

[18] Comparisons were conducted post hoc without control of type-1 error rate and would not have been significant if they were. Interpretations have been qualified as a result.

two groups in support of *Hypothesis 2b* (Performance/Full: t(103) = 2.09, p<05; Learning/Full: t(103)= 2.52, p<.05), with all other groups performing similarly (all t(103) < .34, ns). Although weaker than anticipated, results suggest that partial information does encourage greater search of rule space, but it is of benefit only when combined with a complementary (i.e. learning) goal. When combined with a goal that discourages rule space search, the goal appears to suppress the influence of partial information, limiting necessary task exploration, and likely impairing learning.

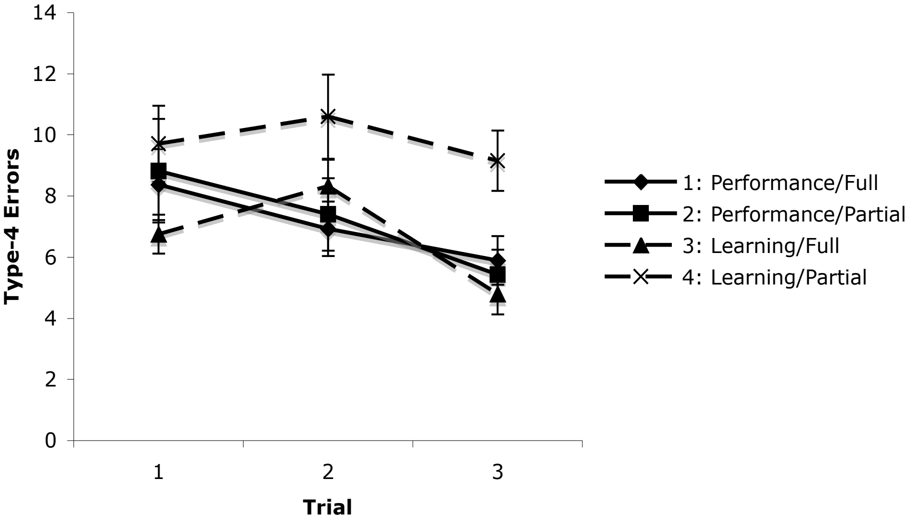


Figure 10: Mean number of Rule-4 errors (+/- 1 SE) committed by each group over the three acquisition phase trials.

Although not considered in hypotheses for the acquisition phase, Landings and OpShort data were also analysed to determine whether the groups differed in their task proficiency or task knowledge during the initial stage of learning. As shown in Table 20, between-subject comparisons for Landings data revealed that full information groups landed significantly more planes than partial information groups. Since no Landings differences were observed in Study 2, it would appear that the three modifications introduced in the present study, along with a greater sample size, were successful at increasing the salience of between group differences. As shown in *Figure 11*, and remarkably consistent with Study 2, this difference between full and partial information conditions appeared largely attributable to the relatively poor performance of the Performance/Partial group. Pairwise comparisons supported this observation revealing that the Performance/Partial group landed significantly fewer planes than any other group during the training phase (Performance/Full: t(103) = 3.10, p<.01; Learning/Full: t(103) = 3.09, p<.01; Learning/Partial: t(103) = 2.93, p<.01), with other groups landing a similar number of planes (all t(103) < .26, ns). The low task exploration shown by the Performance/Partial group during the acquisition phase therefore had a detrimental effect on performance from a very early stage of learning.

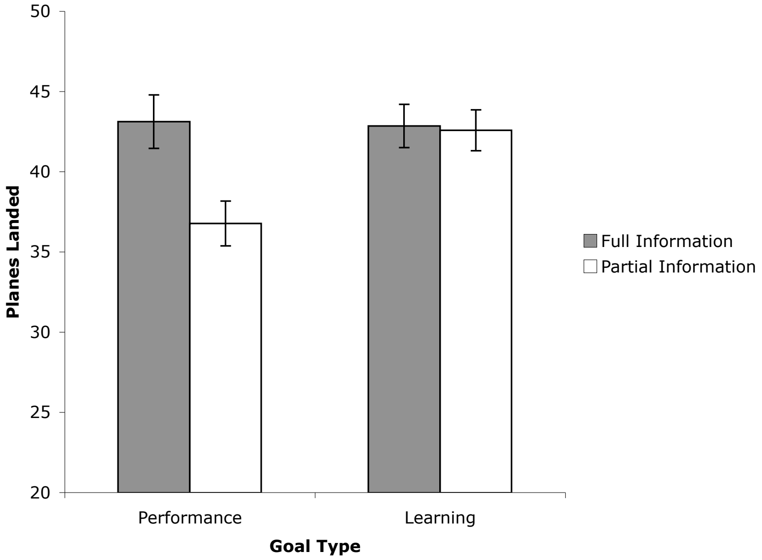


Figure 11: Planes landed by each group during the acquisition phase. Error bars represent +/- 1 SE from the mean.

Results of within-subject comparisons for Landings performance across the trials of the acquisition phase also revealed, as shown in Table 20, that participants improved their Landings performance over the course of training at a slightly decreasing rate (linear: F(1, 99) = 393.88, *p*<.01; quadratic: F(1, 99) = 23.51, *p*<.01).

An interaction between goal type and trial was also observed with trend analyses suggesting that performance goal groups increased their performance more over the training phase than learning goal groups (linear interaction: F(1, 99) = 9.65, *p*<.01), however, as shown in Figure 12, this was likely due to the lower trial 1 performance of the Performance/Partial group. Landings results therefore suggest that the Performance/Partial group was the least proficient of all groups throughout the acquisition phase, a result likely attributable to the group's limited search of rule space despite needing rule space search to discover the information they were not provided.

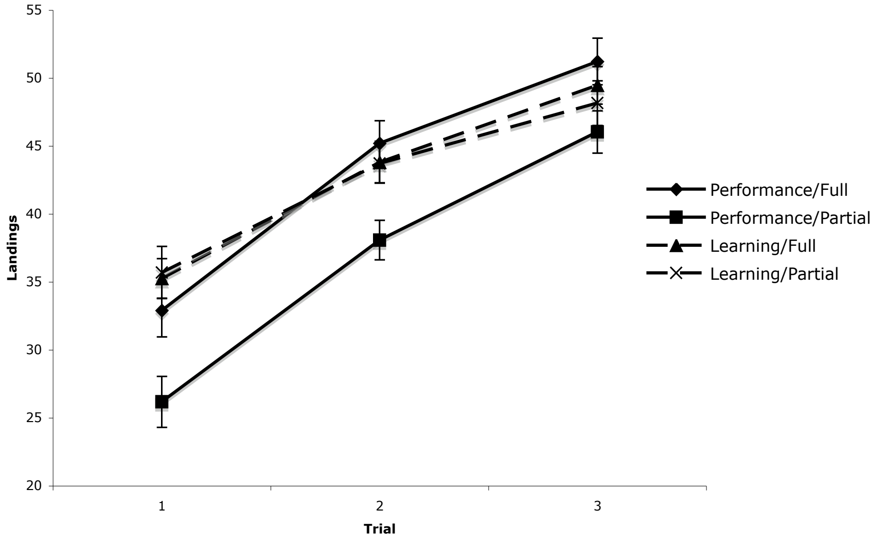


Figure 12: Mean Landings performance (+/- 1 SE) for each group over the three acquisition phase trials.

OpShort results for the acquisition phase revealed that participants increased OpShort similarly across the training phase (linear: F(1,103) = 201.95, *p*<.01) but the level of OpShort differed according to both goal-type and information level, as shown in Table 20. As shown in Figure 13, the difference in OpShort between full and partial information conditions was effectively reversed depending on the type of goal assigned: partial information conditions appeared to be detrimental to performance goal participants but beneficial to those assigned a learning goal, whilst the full information conditions appeared to be beneficial for those provided performance goals but detrimental to those provided learning goals. Pairwise simple effect tests supported this finding indicating that OpShort was higher in the Performance/Full and Learning/Partial conditions than Performance/Partial and Learning/Full conditions (respectively: t(103) = 2.30, p<.05; t(103) = 2.03, p<.05) with OpShort also higher in the Learning/Partial condition than the Performance/Partial condition (t(1,103) = 2.42, p<.05). The combination of complete information with a performance goal, or incomplete information with a learning goal, therefore appeared to produce a relatively high level of rule knowledge acquisition during the training phase.

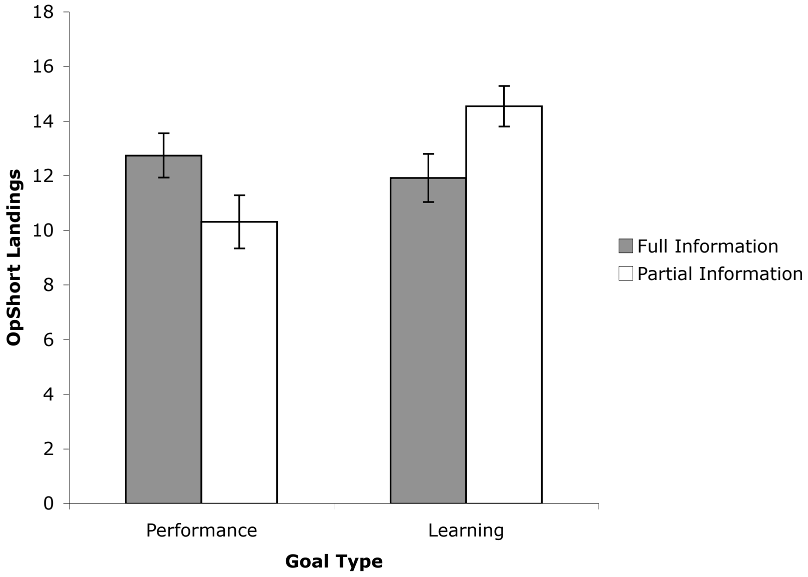


Figure 13: Mean number of OpShort landings (+/- 1 SE) for each group over the acquisition phase.

Whilst both Performance/Full and Learning/Partial groups were expected to develop high levels of knowledge (since the Performance/Full group was provided with complete rule information and an incentive to use it, and the Learning/Partial was strongly encouraged to explore the task), the somewhat lower OpShort landings demonstrated by the Learning/Full group was unexpected. The group did however demonstrate high landings performance during the phase suggesting that their relatively low OpShort may not necessarily be indicative of lower knowledge but also the differences in task conditions. For example, the Learning/Full group was not provided score information during the acquisition phase, reducing the incentive for the group to perform what were more difficult OpShort landings since no positive reinforcement was received. If Learning/Full participants had already acquired a high level of rule-4 knowledge, the conditions of the learning phase would therefore have provided little incentive to perform OpShort landings. This possibility will be explored further in the later phase results.

Cognitive Load: Correlations between fluid ability (a proxy for cognitive load) and performance were anticipated to differ between the groups during the acquisition phase (*Hypothesis 3*) when attentional demands were at their highest and task conditions were not consistent. Both APM and Operation Span performance provided measures of fluid ability and were to be combined to form a composite general ability measure, however, as shown in Table 21, despite demonstrating adequate reliability, the measures did not correlate strongly[19] (r = .29). Since APM performance was more highly correlated with the task performance measures over all phases of the study, this measure alone was used as the indicator of cognitive ability. Operation Span data was not considered for further analyses.

---

[19] Conversations with some participants following testing revealed that many from the paid, Economics Faculty, cohort had recently completed a memory course in mnemonics. This assisted them in achieving higher scores on the Operation Span measure and thus likely contributed to the measure failing to correlate as strongly as had been anticipated with knowledge and performance measures.

Table 21: Means, standard deviations, and correlations between APM and Operation Span measures and Landings, Rule-4 error, and OpShort, scores for each phase of the experiment.

Measure	M	SD	1	2
1. APM	12.74	3.49	.75	
2. Operation Span	9.41	3.19	.29**	.77
Landings				
3. Acquisition	41.24	7.74	.23*	.12
4. Performance	52.27	7.27	.25*	.15
5. Test	46.93	6.11	.29**	.17
Errors				
6. Acquisition	7.68	4.79	.08	-.10
7. Performance	6.48	4.42	.04	-.10
8. Test	8.02	6.56	.06	-.23*
OpShort				
9. Acquisition	19.86	9.10	.26**	.14
10. Performance	25.35	10.53	.27**	.09
11. Test	36.65	14.08	.25**	.10

Note: N= 107. * Correlation is significant at .05, ** Correlation is significant at .01. Reliabilities for both cognitive measures are displayed on the diagonal.

As shown in Table 21, correlations between cognitive ability (APM) and the three task performance measures were unexpectedly low during the acquisition phase (and remained so throughout the study). Correlations with the ability measure were anticipated to be approximately r = .4 during the acquisition phase, consistent with the previous two studies[20] and with previous research with the task (Ackerman, 1988; Kanfer & Ackerman, 1989). The cause of such low correlations is unclear. The longer duration of the acquisition phase in the present study may have allowed cognitive load to fall to low levels by the end of training, but correlations between APM, and performance measures were similarly small across each trial of the phase (Landings: $r_{T1} = .21$, $r_{T2} = .22$, $r_{T3} = .16$; OpShort: $r_{T1} = .14$, $r_{T2} = .30$, $r_{T3} = .20$). Cognitive load did not therefore appear to diminish over the acquisition phase but instead remained low and constant. It is also possible that because the APM task was administered at the end of training rather than the beginning, as had been done previously, participants may have completed the task with less enthusiasm than preceding studies. However,

---

[20] By contrast, the correlation between cognitive ability and landings during acquisition for study 2 was r = .48

means and standard deviations of the measure were almost identical to those of previous studies (Study 3: M= 12.74, SD= 3.49; Study 1: M= 12.36, SD= 3.12)[21] suggesting that performance was comparable. Participants may have simply found the task easier than previous studies, or the APM measure may not have been as effective in assessing cognitive load as it had been previously.

Whilst differentiating between groups on the basis of such small correlations is difficult, moderated regression analyses (Baron & Kenny, 1986) were conducted to determine whether the groups differed in cognitive load during the acquisition phase, consistent with *Hypothesis 3*. As shown in Table 22, cognitive ability predicted both Landings and OpShort performance during the acquisition phase, but the relationship did not differ according to goal type or information level, contrary to predictions. Failure to observe any differences in load between the groups may suggest that load did not differ between groups, or simply that the overall strength of the load-performance correlations were insufficient to allow detection of between group differences. Contrary to *Hypothesis 3* then, the group manipulations appeared to be ineffective in eliciting differences in cognitive load during the acquisition phase.

---

[21] Study 2 used the complete 36-item version of the APM measure rather than the short 20-item version used in Studies 1 and 3 so direct comparison is difficult. To approximate, the mean for Study 2 was 23.29 (SD= 6.27), or 64.7% correct, whilst the mean percent correct for Studies 1 and 3 were respectively 61.8% and 63.7%. APM scores were therefore considered similar across the three studies.

Table 22: Results of three moderated regression analyses of predicting Landings, Rule-4 errors, and OpShort, by fluid ability and group manipulations for the acquisition Phase.

Predictor variable	Landings		Rule-4 Errors		OpShort	
	β	ΔR^2	β	ΔR^2	β	ΔR^2
Step 1		.05*		.01		.07**
Fluid ability	.23*		.08		.26**	
Step 2		.10**		.07		.08*
Goal Type	.15		.10		-.01	
Information Level	-.21*		.18		-.02	
Goal Type x Information Level	.19*		.16		.28**	
Step 3		<.01		.01		.01
Goal Type x Fluid ability	-.05		.07		.05	
Information Level x Fluid ability	-.02		.07		.10	
Overall R		.40		.29		.39
Overall R^2		.16		.08		.16
Adjusted R^2		.11		.03		.10
Overall F (5, 69)		3.11**		1.50		3.05**

* p<.05, ** p<.01

Performance phase

All groups performed the task under identical conditions in the performance phase with groups instructed to maximise their point score under conditions where Rule-4 errors were penalised and no Rule-4 information or feedback was provided. Learning goal participants, and in particular Learning/Partial participants, were expected to show higher landings and OpShort performance than their performance goal counterparts due to their greater search of rule space during the acquisition phase (*Hypotheses 4a* and *4b*). As shown in Table 23 however, between-subject analyses revealed that only landings, not OpShort, differed according goal type. As displayed in *Figure 14*, higher Landings were observed in the learning goal groups, with particularly high Landings observed for the Learning/Partial group, consistent with predictions. Landings results therefore suggested that greater rule space search during the initial stages of learning facilitate development of task proficiency.

Table 23: F statistics for the three performance phase 2x2x(3) Goal Type x Information Level x (Trial) ANOVAs.

Factor	Landings	Rule-4 Errors	OpShort
Within-Subjects			
Trial	21.39**	16.81**	107.90**
Trial x Goal Type	.44	1.60	3.97*
Trial x Information Level	.61	1.06	.05
Trial x Goal Type x Information Level	.34	.64	.99
MSE (Trial)	23.49	9.26	5.96
Between-Subjects			
Goal Type	7.55**	.41	1.07
Information Level	.62	.59	.15
Goal Type x Information Level	5.42*	.85	8.89**
MSE	135.50	59.95	32.42

*Note: Within-subjects df = Landings (2, 206); OpShort (2, 206). Between-subjects df = (1, 103). *p<.05, **p<.01.*

As shown in Table 23, an interaction was also observed for landings with a greater difference in Landings performance found between the two partial, compared to two full, information conditions, consistent with *Hypotheses 5a and 5b*. Whilst further pairwise tests supported this observation with Landings differing significantly between the two partial information conditions (*t*(103) = 3.65, *p*<.01) and not at all between full information groups (*t*(103) = .29, *p*=.77), the result appeared to be largely attributable to the poor performance of the Performance/Partial group who again displayed significantly lower Landings than each of the other groups (Performance/Full: *t*(103) = 2.17, *p*=.03; Learning/Full: *t*(103) = 2.35, *p*=.02). It would appear again that the effect of goal manipulations is particularly pronounced when partial information is provided, likely due to the influence of goals on rule space search during the initial stage of learning.

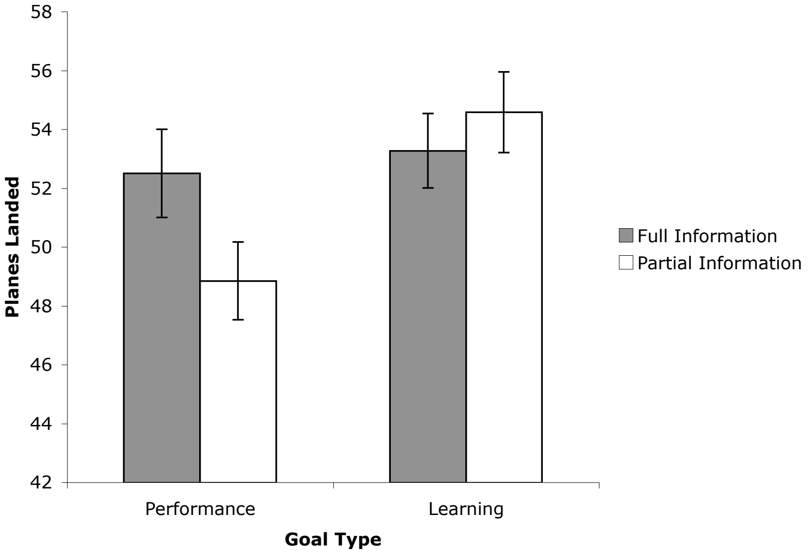


Figure 14: Mean number of Landings (+/- 1 SE) for each group during the performance phase.

Although OpShort was not found to differ between the two goal-type conditions, contrary to *Hypothesis 4a*, other OpShort data was consistent with predictions. First, as shown in Figure 15, OpShort was particularly high in the Learning/Partial group, consistent with *Hypothesis 4b*. Second, the interaction between goal-type and information-level manipulations was significant, consistent with *Hypothesis 5b*, suggesting that differences in OpShort were more pronounced under partial, compared to full, information conditions. Like Landings then, OpShort results generally indicated that the influence of goals is particularly pronounced when partial information is provided during learning.

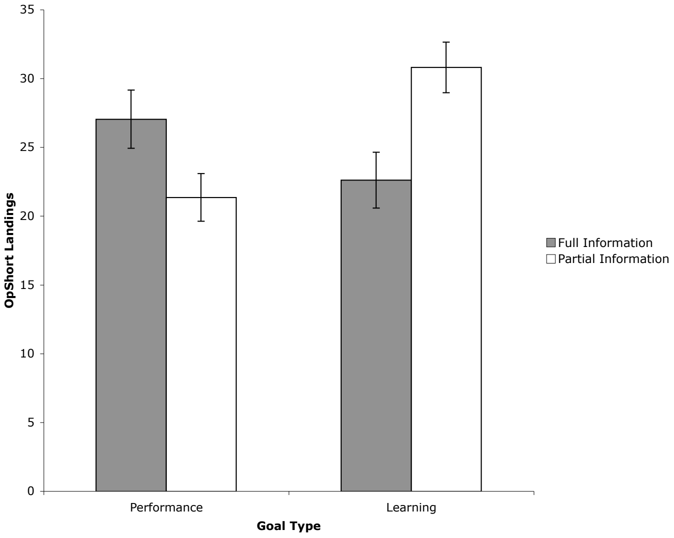


Figure 15: Mean number of OpShort landings (+/- 1 SE) for each group over the performance phase.

OpShort results were however unexpectedly low for the Learning/Full group, as shown in *Figure 15*. Given the group's learning focus and the more persistent pattern of rule space search during the acquisition phase, OpShort was anticipated to be relatively high for this group during the performance phase. Pairwise tests indicated that although the Learning/Full group performed a similar number of OpShort landings to the Performance/Full group ($t(103) = 1.35$, $p=.18$), it was also not different to the low performance of the Performance/Partial group ($t(103) = .47$, $p=.64$). Whilst this result may indicate that the group did not acquire a high level of Rule-4 knowledge during the acquisition phase, the group's high Landings performance (and their relatively high test phase OpShort, to be discussed in the following section), suggests that they may have simply elected not to use the short runways as often as they could during the performance phase. Since performance phase conditions did not force participants to make extensive use of the short runways for mid-sized plane landings, OpShort during the phase may not have been sufficiently sensitive to differences in rule knowledge. This may possibly account for the somewhat lower than anticipated OpShort scores for the Learning/Full group.

Whilst not specifically addressed in hypotheses, analyses also considered whether groups differed in their pattern of Landings, OpShort, and Rule-4 errors across the three performance phase trials. As shown in Table 23, within-subject analyses revealed that all three measures differed across trials in performance phase with only the goal-type manipulation interacting with OpShort. Trend analyses indicated that participants generally improved during the performance phase but at a declining rate with both landings and OpShort scores increasing at decreasing rates across the phase (Landings: linear: $F(1, 103) = 16.60$, $p<.01$; quadratic: $F(1, 103) = 31.230$, $p<.01$; OpShort: linear: $F(1, 103) = 99.23$, $p<.01$; quadratic: $F(1, 103) = 112.89$, $p<.01$) and Rule-4 errors decreasing to an effective minimum (linear: $F(1, 103) = 16.60$, $p<.01$; quadratic: $F(1, 103)= 17.31$, $p<.01$). Results were therefore consistent with participants nearing a performance plateau at the completion of the performance phase, as had been intended.

Trend analyses for the interaction between goal type and OpShort indicated that the linear rate of improvement in OpShort across the performance phase was higher in learning compared to performance goal groups (linear: $F(1, 103) = 8.88$, $p<.01$; quadratic: $F(1, 103) = 1.59$, $p=.28$). This indicated that learning goal participants improved more during the performance phase. This result is consistent with acquisition phase manipulations since learning goal groups were first provided a performance goal and access to their scores during the performance phase, likely providing an incentive to increase their performance. Conversely, performance goal participants had experienced these conditions throughout the acquisition phase and would therefore be less likely to demonstrate a marked improvement. Differences in the rate of improvement in OpShort between the learning and performance goal groups may therefore be explained by differences in the transition between acquisition and performance phases.

Test phase

***Knowledge test**:* At the commencement of the test phase, participants completed the knowledge test to provide an independent assessment of task knowledge prior to their experiencing the changed task conditions. Consistent with *Hypotheses 4a and 4b*, learning goal participants were anticipated to achieve higher scores than their performance goal counterparts, particularly the Learning/Partial group, and differences between the two goal-type groups were anticipated to be larger for those provided partial rule information (*Hypothesis 5b*). Although the measure demonstrated adequate reliability (α= .78), a 2x2 (Goal Type x Level of Information) between-subjects ANOVA failed to reveal any significant differences between the groups (Goal Type: (F(1,102) = .92, ns; Level of Information: F(1,102) = .12, ns; Interaction: F(1,102) = .76, ns). Further examination of results revealed that all groups performed similarly well on the test, scoring between 70% and 75%, suggesting that the measure did not effectively discriminate between participants. Such high and uniform scores may have been due to the inclusion of easier 747 and prop items, for which rules were straightforward, so a further between-subjects ANOVA was conducted only on the 18 items concerning the more complex 727 and DC10 plane landings. This analysis also failed to reveal any significant differences between groups (Goal Type: F(1,102) = .52, ns; Level of Information (F(1,102) =.12, ns; Interaction: F(1,102) = 1.90, ns), though as shown in *Figure 16*, the direction of mean differences was consistent with performance phase data. The Performance/Partial group showed the lowest, and the Learning/Partial group the highest, knowledge scores whilst the two full information groups performed the test similarly. Whilst it may be the case that knowledge did not differ between the groups, it would seem, based on the observed performance differences, that the test was not sufficiently sensitive to accurately assess task knowledge.

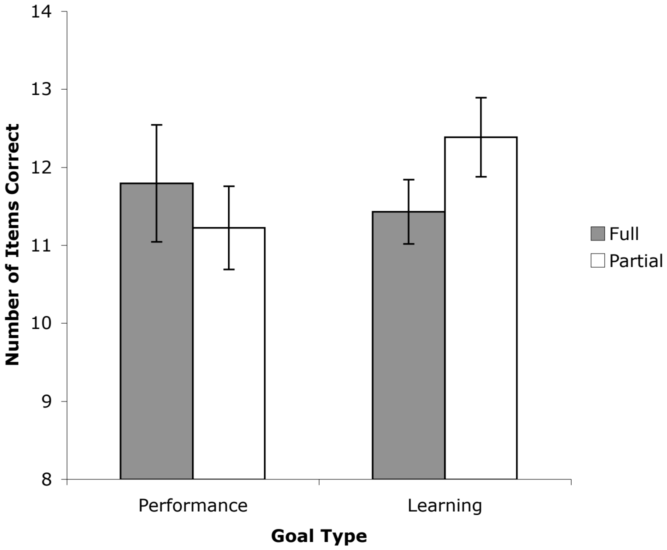


Figure 16: Mean number of 727 and DC10 knowledge test items correct (+/- 1 SE) by each group.

Task measures. Task conditions were changed in the test phase to ensure that performance depended more heavily on Rule-4 knowledge. Like the performance phase, all participants were instructed to maximise their overall scores under conditions where no Rule-4 information was provided, but changes were made to the frequency of plane-type presentations with the frequency of 747s, 727s and DC10s increased, and Props decreased, to force participants to perform more OpShort landings than they had in previous phases. By forcing participants to make greater use of the short runways, the test phase provided a stronger test of Rule-4 knowledge than had been achieved previously.

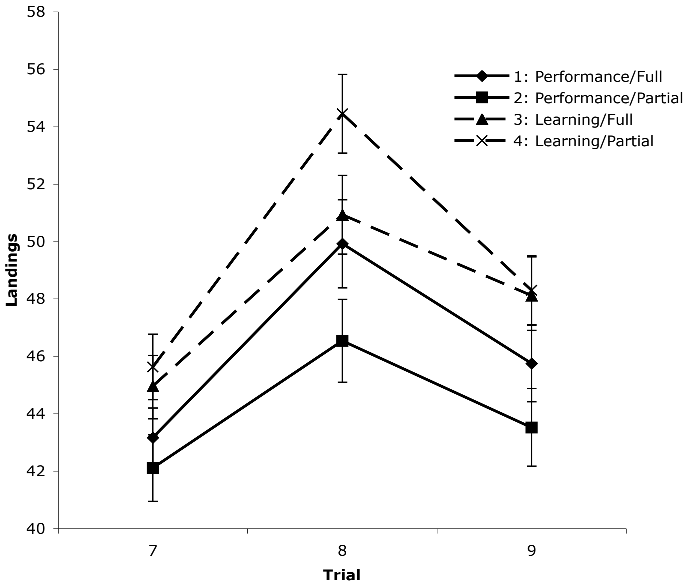


Figure 17: Mean Landings (+/- 1 S.E) for each group over the three test phase trials.

The test phase was also conducted over two days with participants performing the first two test phase trials on day one, and the last test phase trial on day two. Whilst it had been planned to combine all test phase trials into a single phase for analyses (consistent with preceding phases), the performance of all groups was found to have fallen markedly in the final (day-two) trial of the phase, as shown in Figure 17. This fall suggested a discontinuity in performance and that the final trial may have been assessing different knowledge structures (such as retention) than the previous two test phase trials. To ensure test phase analyses were consistent, analyses examined only the first two (i.e. day-1) trials of the test phase with results from the final trial not considered further.

Table 24: F statistics for the three test phase 2x2x(2) Goal Type x Information Level x (Trial) ANOVAs.

Factor	Landings	Rule-4 Errors	OpShort
Within-Subjects			
Trial	147.96**	33.55**	190.80**
Trial x Goal Type	2.85	.44	1.65
Trial x Information Level	.06	2.08	.01
Trial x Goal Type x Information Level	5.87*	5.77*	3.70
MSE (Trial)	15.17	12.82	13.65
Between-Subjects			
Goal Type	9.38**	.07	3.18
Information Level	<.01	<.01	.02
Goal Type x Information Level	3.44	5.57*	10.74**
MSE	71.98	86.35	41.57

*Note: Within-subjects df = (1, 103). Between-subjects df = (1, 103). *p<.05, **p<.01*

Consistent with the performance phase, and *Hypotheses 4a* and *4b*, learning goal participants, especially the Learning/Partial group, were anticipated to show higher landings and OpShort performance in the test phase than their performance goal counterparts due to their greater initial search of rule space. As shown in Table 24, between-subject analyses revealed that only Landings, not OpShort, differed according to goal type with learning goal participants showing higher Landings performance than participants given performance goals, in support of *Hypothesis 4a*. As shown in Figure 18, and consistent with *Hypothesis 4b*, landings were also particularly high for the Learning/Partial group with further pairwise tests indicating that the group performed significantly more Landings than the Performance/Full ($t(103) = 2.09$, p=.04) and Performance/Partial ($t(103) = 12.47$, p<.01) groups but not the Learning/Full group ($t(103) = 1.67$, p=.20). Landings results were therefore in direct support of the hypothesis: groups who were assigned a learning goal, and therefore showed greater rule space search during the acquisition phase, demonstrated a higher level of task proficiency under the more challenging test phase conditions than those assigned a performance goal.

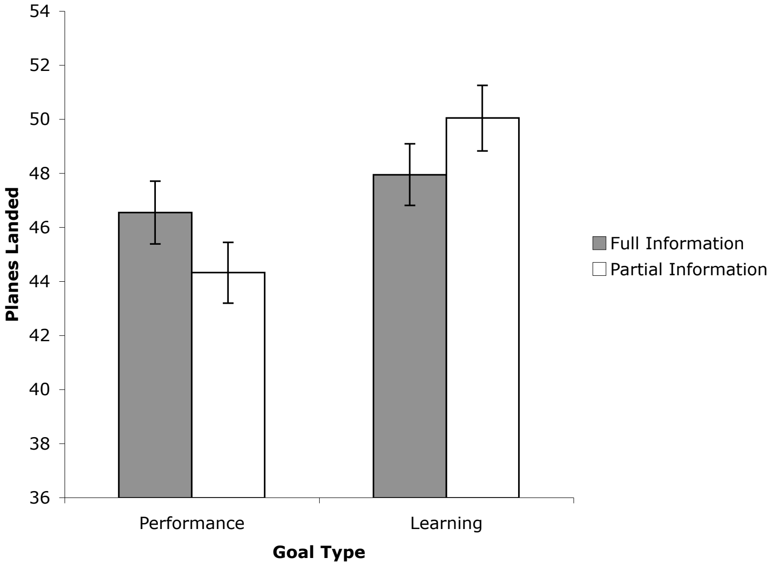


Figure 18: Mean Landings (+/- 1 SE) by each group during the test phase.

Although OpShort was not found to differ between the learning and performance goal groups, contrary to *Hypothesis 4a,* OpShort did appear to be particularly high for the Learning/Partial group, as shown in Figure 19. Consistent with *Hypothesis 4b*, pairwise tests supported this observation with OpShort significantly higher for the Learning/Partial group than the Performance/Partial ($t(103) = 13.08$, $p<.01$), Learning/Full ($t(103) = 6.13$, $p=.02$), but not Performance/Full ($t(103) = 1.54$, $p=.13$) groups. Whilst the high OpShort observed in the Performance/Full group was unexpected, the similarly high level of OpShort in the Learning/Partial group supports predictions that even when not provided complete rule information, participants can, through extensive rule space search, acquire similar, and possibly (given the previous landings results) superior, level of task knowledge and proficiency.

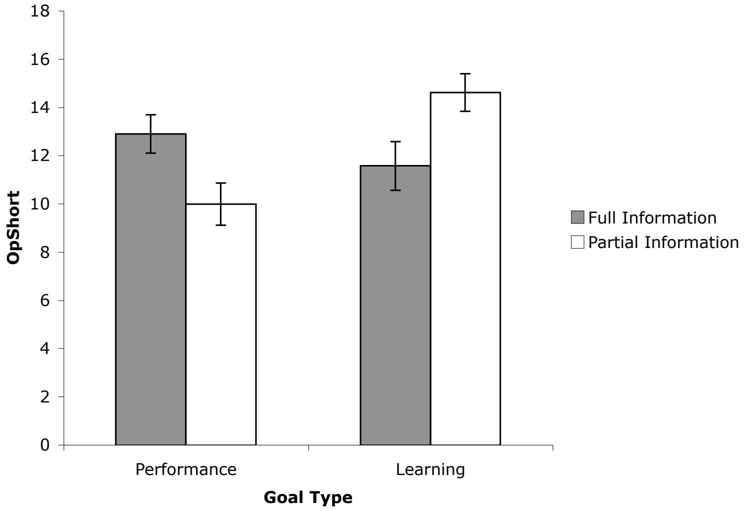


Figure 19: Mean number of OpShort landings (+/- 1 SE) for each group over the test phase.

Hypotheses 5a and *5b* predicted that although no differences in landings or OpShort would be observed between the full and partial information level conditions, an interaction would be observed with goal type such that differences between performance and learning goal groups would be greater under partial compared to full information conditions. In support of predictions, no differences were observed in either landings or OpShort for the level of information conditions, as shown in Table 24, but interactions were observed only for the OpShort measure. As shown in Figure 18, the pattern of Landings differences between learning and performance goal groups did appear to be larger for the groups provided partial than full information, but the interaction was only marginally significant (p=.07). The same, though more pronounced, pattern was however observed for OpShort data, as shown in Figure 19. Whilst the differences appeared consistent with *Hypothesis 5b*, with a larger difference observed between the two partial, compared to two full, information conditions, results for the Learning/Full group were again lower than anticipated. Although significantly lower than only the Learning/Partial group ($t(103) = 2.48$, *p*=.02)[22], the result may again suggest that the group acquired less task knowledge than would be anticipated from their search of rule space during the acquisition phase. Given the group's high overall Landings performance however, in what were more knowledge-dependent task conditions, it would appear that the group did acquire at least a reasonable level of task knowledge. The significantly larger difference in OpShort between the two partial information groups however, and the similar, though not significant pattern for Landings results, are in support of *Hypothesis 5b.* Rule space search again appeared to be of greater importance when participants needed to discover information than when it was provided.

Although not considered in hypotheses, Rule-4 error analyses revealed a significant interaction between goal-type and level of information conditions, as shown in Table 24. The interaction indicated that Rule-4 errors were highest in the Performance/Full and Learning/Partial groups and low in the Performance/Partial and Learning/Full groups. As shown in Figure 20, the interaction seemed largely attributable to performance in trial 7, the initial trial of the test phase, where both Performance/Full and Learning/Partial groups showed relatively high error scores before reducing errors to a level comparable with other groups in trial 8. As shown in Table 24, this pattern gave rise to a significant three-way interaction between trial, goal type, and level of information.

---

[22] Results of pairwise tests for comparisons between the Learning/Full and other groups were as follows: Performance/Full: *t*(103)= 1.05, *p*= .30; Performance/Partial: *t*(103)= 1.31, *p*= .19.

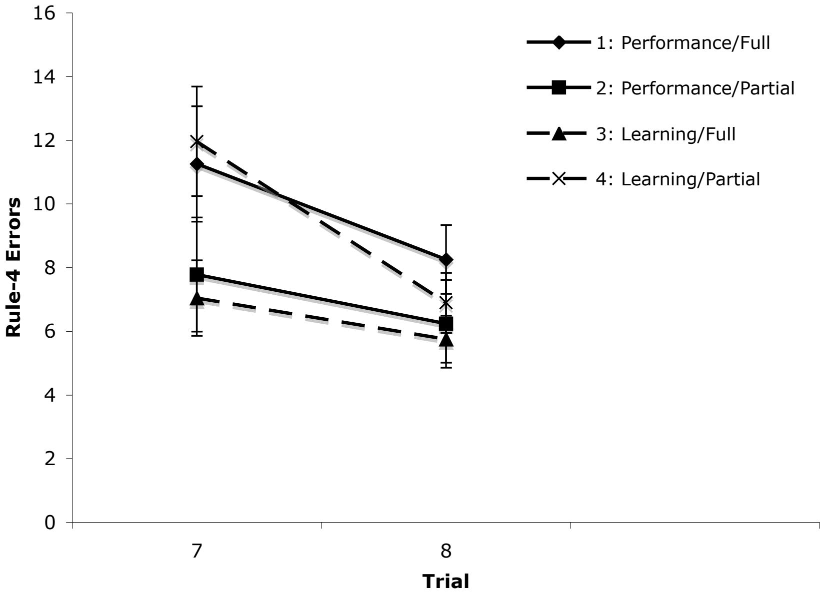


Figure 20: Mean Rule-4 errors (+/- 1 SE) for the first two trials of the test phase.

The reason that that Performance/Full and Learning/Partial groups showed such high initial Rule-4 error scores when introduced to the more difficult conditions of the test phase may be due to an *expertise reversal effect* (Kaluga, Ayres, Chandler, & Sweller, 2003). Both groups had demonstrated a high level of task knowledge in the preceding phases and so may have been more disrupted by the change in task conditions. Alternatively, the elevated Rule-4 error scores may however suggest that the two groups decided to undertake rule space search in response to the increased difficulty of the test phase. That is, participants in these groups may have realised that their state of rule knowledge was insufficient for the test phase conditions and rather than avoid short runway landings, they increased short runway landing attempts, effectively increasing their search of rule space. Conversely, the Performance/Partial group who had shown low knowledge prior to the test phase demonstrated little elevation in Rule-4 errors in the test phase, suggesting that the group was, likely as a result of their knowledge, avoiding OpShort landings.

Support for this explanation may be found in the pattern of landings performance over the first two test phase trials. As shown in Table 24, a three-way interaction for Landings was observed with the Performance/Full and Learning/Partial groups showing stronger improvements between trials 7 and 8 than the Performance/Partial or Learning/Full groups, as illustrated in Figure 17. The same pattern was also observed for OpShort results although the three-way interaction was only marginally significant (p= .06). Landings results therefore indicated that the two groups who showed the highest Rule-4 errors in trial 7, also showed the greatest improvements in Landings performance in trial 8. It seems plausible then that errors, similar to the acquisition phase, were instances of rule space search, and served to facilitate knowledge acquisition and performance. This may also account for the Learning/Full group's lower than expected OpShort performance since the group committed relatively few Rule-4 errors and thus did not improve their knowledge from the performance phase.

Rather than simply providing a stronger assessment of task knowledge, the test phase may have encouraged participants to acquire a higher level of task knowledge in order to maintain and improve their performance. Possibly because they were more motivated to achieve high performance scores, both the Performance/Full and Learning/Partial groups appeared to increase their error scores early in the test phase, which likely contributed to their greater knowledge and performance development across the test phase.

Given that improvements in Landings (and marginally OpShort) performance appeared to be related to rule space search during the initial test phase trial, analyses were also conducted to determine whether cognitive load could account for any of the differences observed differences. Moderated regression analyses were conducted on trial 7 Rule-4 errors, and the difference in Landings and OpShort scores between trials 7 and 8. As shown in Table 25, analyses revealed no differences in cognitive load that could account for the aforementioned differences in error, Landings, or OpShort scores during the test phase. Whilst this is again likely due to the unexpectedly low correlations between performance and load measures, the low load across all groups suggests that load could not account for the differences observed in Rule-4 errors or improvements in Landings and OpShort. It would appear that task exploration, in the form of increased Rule-4 errors, is a better explanation of why the Performance/Full and Learning/Partial groups improved more strongly than other groups in the test phase.

Table 25: Results of the three moderated regression analyses predicting trial 7 Rule-4 errors, and the difference in both Landings and OpShort between trials 7 and 8.

Predictor variable	T7 Rule-4 Errors		T7-T8 Landings		T7-T8 OpShort	
	β	ΔR^2	β	ΔR^2	β	ΔR^2
Step 1		.09		.05*		.03
Fluid ability	.10		.23*		.16	
Step 2		.07		.12*		.07
Goal Type	-.02		.13		.10	
Information Level	.05		.03		.02	
Goal Type x Information Level	.24		.22*		.18	
Step 3		.08		.13*		.07
Goal Type x Fluid ability	.08		-.11		-.07	
Information Level x Fluid ability	.03		.04		-.02	
Step 4		.08		.14*		.08
Goal Type x Information Level x Fluid ability	.02		.06		.11	
Overall R		.28		.37		.29
Overall R^2		.08		.14		.08
Adjusted R^2		.01		.07		.02
Overall F (7, 99)		1.17		2.20*		1.27

*Note: N = 75. *p<.05, **p<.01*

Summary.

Results across all phases were broadly consistent with predictions, and with DST. In the acquisition phase, learning goal groups showed more persistent rule space search, particularly the Learning/Partial group, suggesting that learning goal manipulations were effective at encouraging rule space search, particularly when it was needed to discover information not provided. When incomplete information was provided in combination with a performance goal, rule space search remained low suggesting that despite a distinct need to conduct rule space search in these conditions, performance goals effectively discouraged rule space search when learning. None of the differences in rule space search were explained by differences in cognitive load, although this was likely due to unexpectedly low load-performance correlations in all phases of the study.

Performance and test phase results were largely consistent with rule space search undertaken during the acquisition phase. Groups who showed a higher level of rule space search, i.e. the two learning goal groups, landed more planes during both phases, however, the pattern was more pronounced when only partial information was given. OpShort results were only consistent with predictions (and acquisition phase results) for the partial information groups, with both full information groups demonstrating similar levels of OpShort in both phases. When full information is made repeatedly available during learning, rule space search does not provide a clear advantage to learning outcomes. When participants have to discover information not provided, learning goals appear to foster greater rule space search, and greater knowledge acquisition as a result.

OpShort results were of particular interest in the test phase where both Learning/Partial and Performance/Full groups showed greater improvements in

OpShort. The improvements appeared to be related to the high number of Rule-4 errors these groups committed during the initial trial of the test phase suggesting that when confronted with the greater difficulty of the test phase, these groups increased their search of rule space, improving their knowledge as a result. Whilst consistent with DST, it remains unclear what prompted these specific groups to do this.

Taken together, results were generally supportive of predictions and DST, though no results were directly contrary to the predictions of CLT. A qualification to DST however, appeared to be the relative ineffectiveness of rule space search when full rule information was readily available suggesting that rule space search is effective to the extent that it provides knowledge not otherwise attainable.

Discussion

Study 3 sought to provide a more thorough comparison of CLT and DST than had been accomplished in the preceding studies. Changes were made to the design of Study 2 to increase the salience of experimental manipulations to better elucidate the relationships between rule space search, cognitive load, and knowledge acquisition. Design changes included increasing the length of the training phase to allow greater opportunity for rule space search, extending the duration of the test phase and adding a knowledge test improve assessment of task knowledge, and fixing the previously random generation of task variables to reduce unsystematic variation in task performance. The effectiveness of these modifications are first evaluated before the implications of results are discussed.

Evaluation of task modifications.

Modifications to the previous design were intended to achieve a greater separation of experimental groups and more detailed analysis of how manipulations influenced rule space search. The first change to this end involved increasing the duration of the acquisition phase to allow participants a greater opportunity to search rule space. This was intended to increase the likely influence of rule space search on learning and permit closer analysis of the pattern of rule space search undertaken during learning. Participants in the learning goal groups, and in particular the Learning/Partial group, were found to exhibit a prolonged search of rule space during the acquisition phase relative to performance goal groups, suggesting that the modification was effective for those encouraged to search rule space. Analyses also revealed that performance goal groups showed a steadily declining pattern of rule space search over the acquisition phase which contributed to interpretations by suggesting that such goals do not totally prevent rule space search, but seem instead to encourage a faster reduction in rule space search strategies compared to learning goals. Extending the duration of the acquisition phase was therefore successful in permitting greater search of rule space and more detailed analysis of the patterns of rule space search shown during learning.

The second set of modifications aimed to increase the accuracy of knowledge assessment by introducing a separate knowledge measure and extending the final test phase of task performance. The success of these changes was however mixed. The addition of the separate measure of knowledge was intended to provide a task-independent assessment of participant’s knowledge not influenced by their task proficiency. However, whilst results from the measure were consistent with task-based knowledge indicators, none of the observed differences achieved significance.

Given that large differences in knowledge were observed on task-based indicators, it seems likely that the separate measure was not sufficiently sensitive to detect differences in task knowledge. Further revision of this measure is therefore required before it may be used as a valid assessment of task knowledge.

Extending the duration of the test phase was anticipated to minimise the influence of participant's adaptive abilities (Schunn & Reder, 2001) on performance in the changed, and more difficult, test phase. Participants who had experienced a greater change between acquisition and performance phases were anticipated to be more adaptive when faced with the changed test phase conditions. Participant's adaptive abilities appeared to be relatively unimportant to test phase performance. Differences between the groups were found to be consistent in the first two trials of the test phase suggesting that prior experience of task changes did not influence participant's adaptive abilities. Had adaptive abilities strongly influenced performance, groups who had experienced greater change from earlier phases would have shown the highest initial, but diminishing, advantage in the test phase. Rather than preventing participant's adaptive abilities from influencing performance, the extended test phase allowed adaptive abilities to be dismissed a potential confound of knowledge assessment.

The extended duration of the phase also revealed that some groups actively engaged in rule space search in response to increased task difficulty, likely to improve their task knowledge and increase their subsequent performance. Although the extended duration of the test phase did not necessarily provide a more accurate assessment of task knowledge, it also provided a more detailed analysis of patterns of behaviour under changed, and more complex, task conditions.

The third modification to the previous study's design was the fixing of randomly generated plane and weather sequences to reduce variation in task performance not attributable to experimental manipulations. Participants were therefore expected to perform the task under highly similar plane and weather conditions thereby increasing the salience of between-group differences. In the previous study, no group differences were observed for Landings performance in any of the three experimental phases and differences in Rule-4 errors were observed in only the initial acquisition phase. In the present study, group-differences in Landings performance were observed in all phases of the study, as were differences in Rule-4 errors. Observation of differences not detected in the previous study suggests that the reduction of random task variation (in conjunction with an increased sample size) was successful at reducing unsystematic variation in task performance and, as a consequence, increasing the salience of between group differences.

Influence of goal and information manipulations

Goal and information manipulations were anticipated to influence rule space search during the acquisition phase, thereby affecting knowledge acquisition and performance in the later task phases. Consistent with hypotheses, overall results indicated that participants assigned learning goals showed a more persistent, if not higher, search of rule space during the acquisition phase than performance goal groups, as well as higher levels of knowledge and performance in subsequent task phases. However, whilst overall results were broadly consistent with DST, with high rule space search groups developing higher levels of task knowledge and proficiency, more detailed analyses revealed that results were consistent with DST only for participants provided partial information. Regardless of goal type, full information groups performed similarly throughout the study. Results of full and partial information groups are therefore discussed separately.

Full information groups. Although provided with complete rule information, Learning/Full participants were anticipated to show a greater search of rule space than the Performance/Full group during the acquisition phase. Consequently, they were anticipated to develop superior levels of knowledge and performance in later phases, consistent with DST. Greater rule space search was, in other words, anticipated to confer an advantage even when all relevant task information was provided. Consistent with predictions, the Learning/Full group did show a more persistent search of rule space than the Performance/Full group during the acquisition phase but their overall level of rule space search during the phase was similar as knowledge and performance scores were similarly high in the later phases. This suggests that the greater persistence in rule space search had little influence on knowledge acquisition. Results therefore suggest that under full information conditions, goal manipulations exerted only a weak influence on rule space search during the initial stage of learning, an influence that was inconsequential to knowledge and performance in later phases. It would appear that when all relevant task information is provided, individuals can acquire a good level of task knowledge and proficiency regardless of rule space search. The benefit of rule space search may therefore exist only when it can provide information that is not otherwise available.

The similarity of both full information groups was however tempered by the consistently, though not significantly, higher knowledge and performance scores shown by the Performance/Full group. It had been predicted, consistent with DST, that the Learning/Partial group would show higher knowledge and performance scores due to their greater search of rule space during the acquisition phase but the somewhat higher scores shown by the Performance/Full group, despite their low rule space search, suggested that performance goals may be preferable under such conditions. Although not significant, the results may indicate that rule space search is redundant when complete rule information is provided, and serve only to distract participants by encouraging a search for information that is readily available. Alternatively, providing complete rule information may have made the task easier by making the behaviours required for maximal performance abundantly clear. Under such conditions, the motivational influence of performance goals are likely to be more pronounced (Locke, 2000; Wood et al., 1987), potentially advantaging learners assigned performance goals. Including a motivation measure in further research could assist in establishing whether this is the case. Although rule space search may have been distracting or performance goals more appropriate when full information was provided, the slightly lower performance and knowledge scores exhibited by the Learning/Full do suggest a qualification to DST. Rule space search appears to be irrelevant, perhaps even mildly detrimental, to learning when full task information is provided.

Partial information groups. For participants given partial rule information during the acquisition phase, results were directly consistent with predictions: assigning a learning goal produced consistently higher rule space search, knowledge, and performance, than assigning a performance goal. In the acquisition phase, the Leaning/Partial group showed higher rule space search than any other group, suggesting, as anticipated, that learning goal and partial information manipulations combined to encourage a particularly high level of rule space search. As a result of this higher initial search of rule space, the Learning/Partial group also consistently achieved the highest levels of knowledge and performance of any group in later phases, even compared to those who were provided complete information, in strong support of the theory. This result is particularly noteworthy since it suggests that extensive rule space search can not only overcome the inherent disadvantage of being provided partial information, but may in fact be more advantageous to learning than providing complete information. Consistent with DST then, assigning a learning goal in combination with partial information appeared to encourage very high levels of rule space search, which in turn produced similarly high knowledge and performance in later phases. It would appear that in the absence of complete task information, giving learners sufficient time to explore and test various hypotheses about a task may produce optimal learning outcomes.

Contrary to the Learning/Partial group, the Performance/Partial group showed persistently low knowledge and performance scores throughout the study, consistent with predictions. In all phases, the group's scores were significantly below those of other groups (with the occasional exception of the Learning/Full group) suggesting that despite repeated and prolonged task practice, the group did not overcome the initial disadvantages from being provided partial rule information and conducting a limited search of rule space. The Performance/Partial group in fact appeared to avoid using the short runways in any situation where they lacked the requisite knowledge, instead of using such opportunities to improve their knowledge. Despite needing to conduct rule space search to improve their performance, the Performance/Partial group appeared then to settle on an error-minimisation approach that produced persistently suboptimal performance. This is consistent with DST since it suggests that knowledge could not be developed simply through task practice but instead required a direct search of rule space.

The persistence of knowledge and performance deficits of the Performance/Partial groups is also consistent with the accounts of both Anderson (Anderson, 1982; Anderson, Fincham, & Douglass, 1997; Anderson & Schunn, 2000) and Ericcson (Ericsson, 1996, 2006; Ericsson et al., 1993) since both suggest that improvements in knowledge are not achieved through simple practice (at least not when learners need to discover relevant task information) but instead require a direct and concerted effort. The result is also contrary to predictions of CLT since the lower cognitive load or resource demands associated with highly practiced performance did not result in the Performance/Partial group reducing their knowledge and performance deficits to the other groups with practice.

Whilst the generally poor knowledge and performance scores shown by the Performance/Partial group were consistent with predictions, the pattern of rule space search shown by the group was somewhat contrary to expectations. The low rule space search and poor knowledge and performance scores shown by the Performance/Partial groups were not entirely consistent however with predictions and DST. Performance goal manipulations were intended to discourage participants from searching rule space by instead focusing attention on achieving a high task score rather than on learning task rules. The generally poor performance of the Performance/Partial relative to Learning/Partial group was therefore consistent with predictions suggesting that when partial information is provided, goal type has a marked influence on rule space search and subsequent learning. However, performance goals were anticipated to produce some conflict in the Performance/Partial participants because of the competing need to search rule space to learn the rules that had not been provided. Although rule space search for the Performance/Partial group was anticipated to be lower than that of the

Learning/Partial group, it was not anticipated to be as low as that of the Performance/Full group who had little incentive to search rule space. However, this prediction was not supported. Both performance goal groups demonstrated almost identical, declining, patterns of rule space search over the acquisition phase suggesting that the partial information manipulation was ineffective at increasing rule space search when combined with a performance goal. Since partial information appeared to elevate rule space search when combined with a learning goal, it would appear that performance goals effectively suppressed the influence of partial information conditions on rule space search. Performance goals therefore appeared to reduce rule space search, even under conditions where it was critical to knowledge acquisition.

Despite showing a strong pattern of decline in rule space search during the acquisition phase, both performance goal groups initially showed a level of rule space search that was similar to that of learning goal participants. Performance goals therefore did not appear to completely discourage search of rule space, but rather promote an earlier shift away from rule space search to a more performance focused task approach. This result is directly consistent with previous DST research that has suggested that specific goals encourage faster switching from rule to instance focus during learning (Burns & Vollmeyer, 2002; Vollmeyer et al., 1996). The result extends previous research however since it suggests that a goal's content or focus, in addition to it's specificity, may elicit earlier switching away from rule space search strategies. Rather than completely discouraging rule space search, performance goal manipulations seemed to produce a faster reduction of rule space search with practice, likely limiting knowledge acquisition for the Performance/Partial group who required greater rule space search to develop a reasonable level of task knowledge.

Cognitive load. Goal and information manipulations were anticipated to elicit differences in cognitive load during the acquisition phase to determine whether such differences were related to rule space search. However, results showed uniformly low cognitive load for all groups during the acquisition phase, and throughout the study. The weak relationships between load and task measures were likely attributable to the use of only a single shortened measure of ability that may not have adequately assessed the load of the task. However, given that an identical measure had been used successfully in Study 1, it may imply that participants simply found the present task more straightforward. Despite the weak relationships observed, differences in rule space search were independent of any differences in cognitive load. This implies, albeit weakly, that rule space search and cognitive load are potentially independent, contrary to the argument of CLT that rule space search is a by-product of cognitive load (Sweller et al., 2011). Whilst further research is necessary to fully delineate cognitive load and rule space search, the present research suggests some separation is at least possible.

Theoretical implications

The principle aim of Study 3 was to provide a more thorough comparison of CLT and DST than had been accomplished in the preceding studies. Whilst the study was more effective in testing the differences between the theories, it did not establish a clear separation of the theories. Results were more consistent with DST than CLT, but none were directly contrary to CLT.

Results that were most supportive of DST were those of the partial information groups. Goal manipulations for these groups influenced rule space search in the predicted directions and rule space search appeared strongly linked to knowledge development and performance. Even when undertaken later in practice, (in response to the increased difficulty of the test phase), rule space search appeared strongly related to subsequent improvements in knowledge, in clear support of the theory. Differences between the groups that arose during the acquisition phase also persisted throughout the study suggesting, consistent with DST, that differences in rule space search can produce lasting differences in knowledge. For partial information groups, rule space search therefore appeared to have been influenced in the predicted directions and appeared to contribute substantially to knowledge acquisition in broad support of DST.

Results from full information groups were not generally consistent with DST. Rather than being contrary to the theory, the results were more suggestive of its limitations. Under full information conditions, goal manipulations had only a slight influence on rule space search, and this had almost no effect on subsequent knowledge development or performance. In conjunction with partial information findings, the results suggest that rule space search may only be of benefit when it generates knowledge that could not otherwise have been acquired. Rather than rule space search being a universally beneficial approach to learning, the present research suggests that is more likely to be of benefit when a task is opaque or relevant information is not readily accessible. When a task is transparent, or where complete task information is provided, rule space search may be superfluous, and may potentially be inferior to encouraging learners to focus on performance. Whilst this limitation of the theory is understandable given that the theory was developed to explain problem solving where tasks are often opaque, complex, and require problem solvers to generate relevant task information, the finding is an important contribution in the context of learning complex tasks.

Whilst results were generally supportive of DST, no clear evidence was found contrary to CLT. CLT claims that rule space search is inexorably linked to cognitive load, falling when load increases and rising when load falls (Sweller et al., 2011). Contrary to this perspective, it was anticipated that Learning/Partial participants would demonstrate high rule space search and high cognitive load, thereby uncoupling load and rule space search. However, no differences in cognitive load were observed at any stage of the study, contrary to predictions. All groups instead showed a uniformly low cognitive load, regardless of experimental manipulations or stage of practice. Whilst not contrary to CLT, results were inconsistent with the theory to the extent that observed differences in rule space search and learning were independent of cognitive load, but the result may easily be explained by a potentially flawed assessment of load. Also inconsistent with the theory was that group differences persisted throughout practice suggesting that even though load should fall with practice, knowledge deficits persisted. However, such persistence of group differences could also be explained by other non-cognitive factors, such as potentially unequal task motivation between groups, rather than a clear refutation of the theory. Whilst results were therefore more consistent with DST than CLT, it is difficult to clearly distinguish the theories on the basis of the present data.

Summary

The present study sought to improve on the results of Study 2 and more definitively separate the explanations of DST and CLT on learning. The study was successful in improving upon previous research demonstrating the effectiveness of encouraging rule space search during the initial stages of learning, particularly when individuals are not provided complete task information. In support of previous results, this study also demonstrated that rule space search is unlikely to benefit learners when complete information is provided, suggesting that it may be beneficial only to the extent that it can provide information not otherwise accessible. Whilst results were broadly consistent with DST, and less so with CLT, no clear distinction between the theories was achieved. This study elucidated some of the mechanisms by which rule space search may influences learning and so provides at least preliminary evidence to suggest that CLT and DST are separable.

CHAPTER SEVEN:

GENERAL DISCUSSION

This thesis aimed to investigate the claim made by CLT researchers (Sweller et al., 2011) that CLT and DST are complementary theories of learning. These researchers have proposed that cognitive load determines the level of rule space search such that rule space search is possible only under conditions of low cognitive load. This proposed relationship between CLT and DST has not been tested directly and indirect evidence has been mixed. The present thesis therefore sought to directly investigate whether CLT and DST are independent or complementary explanations of learning and, specifically, whether cognitive load determines the extent to which rule space is searched during learning. This chapter summarises the results of the three empirical investigations conducted as part of this thesis and discusses their direct and broader implications.

Summary of research findings

Study 1 (Chapter 4) sought to investigate the independence of CLT and DST by manipulating rule space search whilst holding cognitive load constant. This was achieved by training participants under one of two KA-ATC task conditions that either encouraged or discouraged rule learning. Cognitive load was held constant between the conditions by carefully matching the element interactivity of each condition. Participants were trained to approximately asymptotic levels of performance before being presented with the opposite task variant to test the transferability of their knowledge. In clear support of hypotheses, results demonstrated higher rule knowledge in the group encouraged to search rule space despite equivalently high cognitive load across for both conditions. Contrary to the proposed interdependence of CLT and DST then, high levels of rule space search were observed in conjunction with high cognitive load. Results therefore suggested that CLT and DST were independent accounts of learning.

However, results of Study 1 may also have been attributable to motivational differences between the groups. The different task manipulations had inadvertently created conditions where rule knowledge was likely to have been more beneficial to those encouraged to search rule space. The observed superiority of the high rule space search group could therefore have been due to differences in rule space search and/or motivation. Since the cause of knowledge differences could not be determined, interpretation of results was uncertain. A revised research method was therefore proposed for Study 2 (Chapter 5), which included, in particular, a more direct measure of rule space search.

Study 2 sought to conduct a more comprehensive investigation of the relationship between rule space search and cognitive load than had been achieved in Study 1. The study employed a 2x2 between-subjects design where participants were trained under either non-specific performance or non-specific learning goal manipulations with either full or partial rule information provided. Goal manipulations were intended to influence rule space search but, because they were both non-specific, not cognitive load. Information level manipulations were intended to influence both rule space search and cognitive load to produce conditions high in both rule space search and cognitive load (and vice versa). The manipulations were therefore intended to demonstrate that rule space search could vary independently of cognitive load and that high (low) cognitive load did not impair (facilitate) rule space search. Rule space search was assessed by the number of attempts participants made to learn the rules, as evidenced by unpenalised error scores.

Results of Study 2 were mixed though generally supportive of the independence of CLT and DST. Learning goal groups showed similarly high levels of rule space search but differing levels of cognitive load suggesting that the two constructs varied independently. Full information groups showed higher cognitive load than partial information groups but the difference was unrelated to differences in rule space search with groups in each information condition exhibiting different levels of rule space search despite the similar levels of cognitive load. Even though cognitive load differed for groups with similar levels of rule space search and was equivalent for groups with different levels of rule space search, there was an overall negative correlation between cognitive load and rule space search, contrary to the proposed independence of the theories. Moreover, neither cognitive load nor rule space search appeared overall to be related to learning outcomes. Thus results were generally supportive of the independence of cognitive load and rule space search, but they were also equivocal.

The mixed findings of Study 2 appeared largely due to the two performance goal conditions. Whilst both demonstrated low rule space search, the Performance/Full group developed markedly higher knowledge than the Performance/Partial group, who developed very poor knowledge overall. This difference in knowledge appeared not to be attributable to rule space search per se but instead appeared to be due to the benefit generated by searching rule space. When full information is provided, a high level of rule knowledge can be acquired simply by reading the provided information. Rule space search, in this case, is unnecessary. However, when only partial information is provided, rule space search is needed to discover the missing information. The benefit of rule space search is therefore greater when information has to be discovered. This explains why low rule space search was more detrimental to the Performance/Partial than Performance/Full groups.

In clear support of this interpretation, the relationship between rule space search and learning was stronger for partial compared to full information groups. The Performance/Partial group conducted very little rule space search and acquired a very low level of knowledge whilst the Learning/Partial group conducted a high level of rule space search and acquired a high level of knowledge. Conversely, despite disparate levels of rule space search, both full information groups developed similarly high levels of knowledge. Results of Study 2 therefore suggested a qualification to DST: rule space search may be beneficial only to the extent that it provides knowledge not otherwise available.

Since the results of Study 2 were promising but somewhat ambiguous, Study 3 (Chapter 6) modified the experimental design to improve the salience of between group differences. Modifications provided participants a longer duration in which to search rule space, improved the accuracy of knowledge measures, and reduced random task variability, whilst retaining the same 2x2 design as Study 2.

Despite improvements, results of Study 3 were again equivocal, but overall consistent with Study 2 and with the independence of CLT and DST. Learning goal groups demonstrated more persistent levels of rule space search than performance goal groups who instead showed a more rapid decline in rule space search with practice. This was consistent with previous findings that specific performance goals encourage a premature switch from rule to instance space search (Burns & Vollmeyer, 2002; Vollmeyer et al., 1996). Learning goal groups derived a greater benefit from rule space search, and achieved a higher level of performance, than performance goal groups. Rule space search therefore appeared to directly benefit learning outcomes, particularly for those encouraged to search rule space. Relative to Study 2, allowing a longer time to search rule space appeared to increase its effectiveness.

Cognitive load results were less clear. Load was comparatively low for all groups in Study 3 (relative to the other studies) and partial information groups exhibited slightly higher cognitive load than full information groups (the inverse of Study 2). The unexpected results may be attributable to the use of only a single, shortened, and indirect measure of cognitive load, raising some doubts about measurement accuracy. It is therefore unclear whether the observed null correlation between rule space search and cognitive load indicates independence, or simply failure of the measure. However, that only cognitive load differed between the full and partial conditions suggests, albeit weakly, some independent variation of cognitive load and rule space search/learning outcomes. Results therefore provide some suggestion of independence between CLT and DST.

Consistent with Study 2, knowledge and performance measures indicated a greater difference between the two partial compared to two full information groups. Since the Learning/Partial group demonstrated greater rule space search than the Performance/Partial group, results again suggest that the benefit of rule space search is greater when complete information is not provided. Notably, the Performance/Partial group showed an almost identical decline in rule space search to the Performance/Full group, despite being provided far less information, suggesting that assigning a performance goal, even if non-specific, can suppress rule space search. In support of previous findings then, rule space search appeared to benefit learning only when it generated information that was not otherwise available. That differences in knowledge appeared related to rule space search, rather than cognitive load, also suggests some independence of each.

In summary, results were generally, if not definitively, consistent with an independent interpretation of rule space search and cognitive load. Consistent with hypotheses presented in Chapter 3, rule space search appeared to vary, and predict knowledge outcomes, independently from cognitive load. Whilst the strongest support found was in Study 1, results could have been confounded with motivation. Results from Studies 2 and 3 were more mixed, though generally supportive. Taken together, the results provide the first direct, albeit qualified, evidence of the independence of CLT and DST.

Implications for the independence of CLT and DST

The general premise of this thesis has been that by demonstrating independent variation in cognitive load and rule space search, CLT and DST could be shown to be separable theories of learning. However, on reflection, this interpretation may be overly simplistic. In describing each theory, Chapter 2 noted the substantial overlap in the mechanisms by which both CLT and DST explain learning outcomes, particularly those concerning the goal free effect. In essence, both theories agree that specific goals elicit means-end strategies that impair learning whilst non-specific goals encourage task exploration that facilitates learning. It is therefore unlikely that the theories are entirely independent. Instead it may be, as Sweller et al (2011) suggest, that CLT does in fact subsume DST, but perhaps not in every situation.

CLT has been suggested to subsume DST because cognitive load can explain the learning mechanisms proposed by both theories. For example, it may be the higher cognitive load of means-end strategies that prevent attention to learning task rules. As Sweller et al (2011) note: “.. an emphasis on instance space prevents attention to the rule space… [because] limited working memory… prevents us from attending to both instance and rule space simultaneously” (p96). In effect, this means CLT proposes an additional, superordinate construct (i.e. cognitive load) to explain learning differences compared to DST. For DST, task outcomes themselves define the strategies and attentional focus of learners. If a task is goal directed, it will prompt goal-directed strategies and attention. If a task is rule or knowledge directed, it will prompt rule-focused strategies and attention. Under a DST conception then, cognitive load may be superfluous since task outcomes alone may be sufficient to explain the observed differences in learning. In some cases DST may represent a more parsimonious explanation of the goal free effect than CLT.

The question of cognitive load’s relevance in explaining the goal free effect may be more clearly illustrated by comparing two studies frequently cited as evidence for CLT. In the first, Sweller & Levine (1982) trained blindfolded university students on simple maze tracing problems under specific or non-specific goal conditions. The authors explained the observed goal-free effect by reference to DST (Simon & Lea, 1974) not CLT, because the latter had yet to be developed. This interpretation, without reference to cognitive load, seems sufficient. The task was straightforward and the participants likely high in cognitive abilities so it is conceivable that the capacity constraints of working memory (i.e. cognitive load) did not influence learning outcomes, even for those assigned specific goals. In this case, cognitive load may be unnecessary to explain the goal free effect.

In the second study, Owen & Sweller (1985) trained high school students to solve a number of trigonometry problems under specific and non-specific goal conditions, again observing the goal free effect. In this case, the problems were novel and sufficiently complex that the students had to be first taught the general principles required to solve each. A CLT interpretation, as invoked by the authors, seems apt in this case because the working memory demands of the task were almost certainly high. These results therefore are more conducive to a CLT interpretation than the previous study (Sweller & Levine, 1982). In this case, cognitive load may be useful to account for the observed differences.

The above examples suggest that invoking cognitive load to account for the goal free effect may not always be necessary. Sometimes the use of goal-directed or exploratory strategies and the amount of attention devoted to learning task rules may be sufficient, to account for learning differences. A pure DST interpretation may be more parsimonious in these situations.

The suggestion that cognitive load is not always required to account for observed differences in learning accords with the present findings. Cognitive load was found to be unrelated to either strategy use or learning outcomes in task performance. Strategy use and rule focus were generally found to predict learning outcomes alone. Whilst very tentative, the results of the present thesis therefore imply that some goal free effects may not be influenced by the limited nature of working memory capacity. In these situations, the way in which a task encourages learners to acquire rule knowledge or simply achieve a goal may be all that is necessary to explain differences in learning. This interpretation is not entirely inconsistent with CLT since the theory acknowledges these mechanisms contribute to learning outcomes, however, it does suggest a qualification to the theory.

This interpretation in no way implies that cognitive load is an invalid or inappropriate construct to explain learning differences in a number of situations. It is difficult to imagine for example how DST alone could explain why redundant information impairs learning (Chandler & Sweller, 1991, 1996) or how integrating task instructions within a problem can facilitate learning (Sweller, Chandler, Tierney, & Cooper, 1990; Tarmizi & Sweller, 1988). Moreover, findings such as the expertise reversal effect, where information that facilitates novice performance but inhibits expert performance, are highly conducive to a CLT interpretation (Kalyuga, Ayres, Chandler, & Sweller, 2003; Kalyuga et al., 2001). Given the substantial evidence supporting CLT, cognitive load is almost certainly a valid explanation of learning.

Rather than suggesting the independence of CLT and DST overall, the present findings suggest that the mechanisms proposed by both theories to account for the goal free effect may, in some cases, be independent of cognitive load. The current research therefore tentatively suggests a limitation to CLT. Future research may seek to examine further conditions under which load is, and is not, necessary to account for learning differences associated with the goal free effect.

Further implications

Beyond informing the distinction between CLT and DST the present research identified three additional implications that may contribute to further investigations of learning in complex settings.

First, the findings indicate a qualification to DST. Results of Studies 2 and 3 consistently found that the benefits of rule space search applied only to those groups provided partial rule information. When full information was provided, differences in rule space search appeared to have no influence on the acquisition of knowledge. The results therefore suggest that rule space search is of benefit only to the extent that can provide information not otherwise available. This is consistent with Simon & Lea's (1974) original description of the purpose of rule space search to 'discover' rule information. Encouraging rule space search is therefore of particular importance when tasks are opaque or ambiguous.

Second, results suggest that differences in learning conditions can have a persistent influence on performance, despite repeated task practice. Previous research in the domain of skill acquisition had found that initial learning differences ameliorate with practice (Ackerman, 1987, 1988; Kanfer & Ackerman, 1989). This has been posited to be a result of a reduction in the cognitive demands as task knowledge becomes increasingly proceduralised (Anderson, 1982; Anderson & Schunn, 2000; Lee & Anderson, 2001). The present findings suggest however that if task information is not provided, and task exploration is discouraged, that learners may never acquire the missing knowledge. They may instead simply adopt a suboptimal response strategy, perpetuating their disadvantage. Results therefore emphasise the importance of initial learning conditions for future performance. Further research may seek to investigate what other conditions lead to such persistent learning impairments and how these can be ameliorated.

Third, Study 3, which demonstrated a goal free effect using non-specific goals, suggests that goal specificity may not be the only goal characteristic that influences learning. This finding accords with previous research where goal content has, over and above the effect of goal specificity, influenced learning (Winters & Latham, 1996; Wirth et al., 2009). Future research on the effects of goals may wish to consider both the content and specificity of goals when examining their effects on learning.

Limitations.

Perhaps the most obvious limitations of the present research were its measurement of both cognitive load and rule space search. Cognitive load was assessed using multiple psychometric measures of working memory and cognitive ability. These measures were correlated with aspects of task performance to provide an indicator of the cognitive load of the task. Whilst this method is almost certainly valid given cognitive load's definition, it was not an ideal method to compare cognitive load between groups.

Correlations are inherently less stable statistics than means. Comparing differences between groups on correlations therefore requires large sample or effect sizes to achieve significance. The relatively small samples presented here may not have been sufficient to reliably assess knowledge differences between groups. Standard self-report measures of cognitive load (Paas & Van Merrienboer, 1993, 1994a) would have been preferable in this case since they compare means and are therefore inherently more stable indices than correlations. However, since psychometric measures represent a measure of cognitive load that is independent from both task performance and the subjective reflections of participants, further research may benefit from the use of such measures. They may, for example, be used to better validate existing self-report measures of cognitive load[23].

Rule space search was operationalised in the present research as the number of exploratory attempts to use the short runways to land mid-sized planes (Rule-4 errors). Whilst this measure may have been a reasonable test of exploration behaviour, this does not necessarily indicate that participants were searching rule space (Burns & Vollmeyer, 2002). Like Vollmeyer et al.'s (1996) findings, it is unclear whether present participants were testing hypotheses about rules they had generated through rule space search or were simply testing different responses without predicting a specific outcome (i.e. nonpredictive testing). Whilst it is likely based on previous research (Burns & Vollmeyer, 2002), that participants' higher exploration activity did constitute hypothesis testing, this cannot be verified from the

---

[23] This could also establish whether traditional self-report measures of cognitive load are consistent with psychometric measures such as working memory

present data. Further research may wish to use verbal protocols to better elucidate the nature of exploration behaviours to more unequivocally demonstrate rule space search[24].

The present research was also limited in that it did not assess task motivation. In time pressured, outcome oriented tasks like the KA-ATC task, motivational factors almost certainly contribute to performance (Kanfer & Ackerman, 1989; Kanfer et al., 1994). The absence of a motivational measure may therefore qualify interpretations. This was particularly evident for Study 1 where motivation could have completely accounted for the observed learning differences and in Studies 2 and 3 where the different conditions could have exerted differing effects on task motivation. Inclusion of a simple motivational measure in future research could help to distinguish between the cognitive and motivational effects of experimental manipulations.

Summary and concluding remarks.

CLT and DST propose different explanations for the goal free effect. On the one hand CLT argues that specific goals impair knowledge acquisition by eliciting greater cognitive load than non-specific goals. On the other, DST argues that specific goals impair knowledge acquisition by directing focus away from task rules towards specific task instances. To reconcile these approaches, CLT has proposed that the search for task rules is determined by cognitive load. Under this approach, higher cognitive load effectively prohibits rule space search whilst low load encourages it. This proposition, whilst certainly plausible, had never been directly tested.

---

[24] This could also assist in determining whether observed task exploration was an explicit attempt to learn task rules or an implicit response to task manipulations. Geddes and Stevenson (1997) have suggested that rule learning is likely to be explicit under rule space search conditions but implicit under instance space search conditions, though more direct measures of rule space search would be needed to support this interpretation.

This thesis aimed to directly test the relationship between cognitive load and rule space search. To achieve this, three empirical studies were conducted in which rule space search was manipulated in opposition to the level of cognitive load. Whilst tentative, results generally supported their independence. Rule space search, at least in some cases, may vary independently of cognitive load.

In conducting this research, the present thesis has made a number of contributions to general understanding of the processes underpinning learning in complex settings. First, it provides the first direct evidence that cognitive load and rule space search may be independent explanations of learning in complex settings. Consequently, the research tentatively suggests that even under conditions of high cognitive load, encouraging rule space search may facilitate, rather than impede, learning. Second, results suggest that rule space search may only be of benefit when tasks are opaque or ambiguous and relevant information has to be discovered. When complete and clear task information is available, rule space search is likely to be redundant. Third, findings indicate that initial learning conditions can have lasting effects on learning, which may not necessarily diminish with practice. Encouraging a learning rather than a performance focus when a task is opaque or ambiguous may facilitate acquisition of a more comprehensive task knowledge. Finally, the results suggest that the goal free effect may not apply only to specific and non-specific goals but also to performance and learning goals. This is an important extension to a well-established effect.

The present results are however tentative, particularly in relation to the independence of cognitive load and rule space search. They therefore require replication. The challenge of future research will be to do so more convincingly by using tasks and designs that are more similar to those historically used by CLT and DST researchers.

The investigation and specification of the cognitive processes underpinning learning, particularly in relation to complex tasks, is important for the continued evolution of educational methods. By doing so, the field progresses not only the knowledge of how learning occurs, but of knowledge development generally. Although preliminary, the present research has sought to contribute to the progression of knowledge in this field.

REFERENCES

Ackerman, P. L. (1986). Individual differences in information processing: An investigation of intellectual abilities and task performance during practice. *Intelligence, 10*(2), 101-139.

Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin, 102*(1), 3-27.

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General, 117*(3), 288-318.

Ackerman, P. L. (1990). A correlational analysis of skill specificity: Learning, abilities, and individual differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(5), 883-901.

Ackerman, P. L. (1992). Predicting individual differences in complex skill acquisition: Dynamics of ability determinants. *Journal of Applied Psychology, 77*(5), 598-614.

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working Memory and Intelligence: The Same or Different Constructs? *Psychological Bulletin, 131*(1), 30-60.

Ackerman, P. L., & Cianciolo, A. T. (2002). Ability and task constraint determinants of complex task performance. *Journal of Experimental Psychology: Applied, 8*(3), 194-208.

Anderson, J. R. (1982). Acquisition of Cognitive Skill. *Psychological Review, 89*(4), 369-406.

Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem situations. *Psychological Review, 94*(2), 192-210.

Anderson, J. R., Fincham, J. M., & Douglass, S. (1997). The role of examples and rules in the acquisition of a cognitive skill. *Journal of Experimental Psychology: Learning Memory and Cognition, 23*(4), 932-945.

Anderson, J. R., & Schunn, C. D. (2000). Implications of the ACT-R learning theory: No magic bullets. In R. Glaser (Ed.), *Advances in instructional psychology: Educational design and cognitive science* (Vol. 5, pp. 1-33). Mahwah, NJ: Lawrence Erlbaum Associates Publishers; US.

Anzai, Y., & Simon, H. A. (1979). The theory of learning by doing. *Psychological Review, 86*(2), 124-140.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*: Upper Saddle River NJ: Prentice-Hall.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. New York, NY: Cambridge University Press.

Berry, D. C., & Broadbent, D. E. (1984). On the relationship between task performance and associated verbalizable knowledge. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 36A*(2), 209-231.

Bors, D. A., & Stokes, T. L. (1998). Raven's Advanced Progressive Matrices: Norms for first-year university students and the development of a short form. *Educational and Psychological Measurement, 58*(3), 382-398.

Burns, B. D., & Vollmeyer, R. (2002). Goal specificity effects on hypothesis testing in problem solving. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 55*(1), 241-261.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York, NY: Cambridge University Press.

Catrambone, R. (1998). The subgoal learning model: Creating better examples so that students can solve novel problems. *Journal of Experimental Psychology: General, 127*(4), 355-376.

Catrambone, R., & Holyoak, K. J. (1990). Learning subgoals and methods for solving probability problems. *Memory & Cognition, 18*(6), 593-603.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293-332.

Chandler, P., & Sweller, J. (1996). Cognitive load while learning to use a computer program. *Applied Cognitive Psychology, 10*(2), 241-261.

Chase, W. G., & Simon, H. A. (1973a). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215-281). New York: Academic Press. .

Chase, W. G., & Simon, H. A. (1973b). Perception in chess. *Cognitive Psychology, 4*(1), 55-81.

Chen, G., & Mathieu, J. E. (2008). Goal orientation dispositions and performance trajectories: The roles of supplementary and complementary situational inducements. *Organizational Behavior and Human Decision Processes, 106*(1), 21-38.

Chi, M. T. (1978). Knowledge structures and memory development. In R. S. Siegler (Ed.), *Children's thinking: What develops?* (pp. 73-96). Hillsdale, NJ: Lawrence Erlbaum Associates.

Chi, M. T., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 7-75). Hillsdale: Lawerence Erlbaum.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*(2), 121-152.

Chi, M. T. H., Glaser, R., & Farr, M. J. (1988). *The nature of expertise*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cianci, A. M., Klein, H. J., & Seijts, G. H. (2010). The effect of negative feedback on tension and subsequent performance: The main and interactive effects of goal content and conscientiousness. *Journal of Applied Psychology, 95*(4), 618-630.

Colom, R., Rebollo, I., Palacios, A., Juan-Espinosa, M., & Kyllonen, P. C. (2004). Working memory is (almost) perfectly predicted by g. *Intelligence, 32*(3), 277-296.

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*(5), 769-786.

Conway, A. R. A. (1996). Individual Differences in Working Memory Capacity: More Evidence for a General Capacity Theory. *Memory, 4*(6), 577-590.

Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology, 79*(4), 347-362.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. [Comment/Reply]. *Behavioral and Brain Sciences, 24*(1), 87-185.

de Groot, A. D. (1978). *Thought and choice in chess. (2nd ed)*. Oxford, England: Mouton.

Diener, C. I., & Dweck, C. S. (1978). An analysis of learned helplessness: Continuous changes in performance, strategy, and achievement cognitions following failure. *Journal of Personality and Social Psychology, 36*(5), 451-462.

Diener, C. I., & Dweck, C. S. (1980). An analysis of learned helplessness: II. The processing of success. *Journal of Personality and Social Psychology, 39*(5), 940-952.

Duncker, K. (1945). *On Problem-Solving. (Psychological Monographs, No. 270.)*. Washington, DC: American Psychological Association.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95*(2), 256-273.

Elliot, A. J., McGregor, H. A., & Gable, S. (1999). Achievement goals, study strategies, and exam performance: A mediational analysis. *Journal of Educational Psychology, 91*(3), 549-563.

Elliott, E. S., & Dweck, C. S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology. Vol, 54*(1), 5-12.

Ericsson, K. A. (1996). The acquisition of expert performance: An introduction to some of the issues *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games* (pp. 1-50). Hillsdale, NJ: Lawrence Erlbaum Associates.

Ericsson, K. A. (2006). The Influence of Experience and Deliberate Practice on the Development of Superior Expert Performance *The Cambridge handbook of expertise and expert performance* (pp. 683-703). New York, NY: Cambridge University Press; US.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*(2), 211-245.

Ericsson, K. A., Krampe, R. T., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*(3), 363-406.

Feltovich, P. J., Prietula, M. J., & Ericsson, K. (2006). Studies of Expertise from Psychological Perspectives. In K. A. Ericsson, N. Charness, P. J. Feltovich & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 41-67). New York, NY: Cambridge University Press.

Friedrich, R. J. (1982). In Defense of Multiplicative Terms in Multiple-Regression Equations. *American Journal of Political Science, 26*(4), 797-833.

Geddes, B. W., & Stevenson, R. J. (1997). Explicit learning of a dynamic system with a non-salient pattern. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 50*(4), 742-765.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*(2), 155-170.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*(1), 1-38.

Gist, M. E., & Stevens, C. K. (1998). Effects of practice conditions and supplemental training method on cognitive learning and interpersonal skill generalization. *Organizational Behavior and Human Decision Processes, 75*(2), 142-169.

Glaser, R. (1987). Introduction: Further notes toward a psychology of instruction *Advances in instructional psychology* (Vol. 3). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hagmayer, Y., Meder, B., Osman, M., Mangold, S., & Lagnado, D. (2010). Spontaneous causal learning while controlling a dynamic system. *The Open Psychology Journal, 3*, 145-162.

Halford, G. S., Maybery, M. T., & Bain, J. D. (1986). Capacity limitations in children's reasoning: A dual-task approach. *Child Development, 57*(3), 616-627.

Hilbert, T. S., & Renkl, A. (2009). Learning how to use a computer-based concept-mapping tool: Self-explaining examples helps. *Computers in Human Behavior, 25*(2), 267-274.

Holmbeck, G. N. (2002). Post-hoc probing of significant moderational and mediational effects in studies of pediatric populations. *Journal of Pediatric Psychology, 27*(1), 87-96.

Holyoak, K. J. (2005). Analogy. *The Cambridge handbook of thinking and reasoning* (pp. 117-142). New York, NY: Cambridge University Press.

Jaeggi, S. M., Studer-Luethi, B., Buschkuehl, M., Su, Y.-F., Jonides, J., & Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning--Implications for training and transfer. *Intelligence, 38*(6), 625-635.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*(1), 23-31.

Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology, 93*(3), 579-588.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review, 9*(4), 637-671.

Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology, 74*(4), 657-690.

Kanfer, R., & Ackerman, P. L. (1996). A self-regulatory skills perspective to reducing cognitive interference *Cognitive interference: Theories, methods, and findings* (pp. 153-171). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kanfer, R., Ackerman, P. L., Murtha, T. C., & Dugdale, B. (1994). Goal setting, conditions of practice, and task performance: A resource allocation perspective. *Journal of Applied Psychology, 79*(6), 826-835.

Keil, C. T., & Cortina, J. M. (2001). Degradation of validity over time: A test and extension of Ackerman's model. *Psychological Bulletin, 127*(5), 673-697.

Kimball, D. R., & Holyoak, K. J. (2000). Transfer and expertise. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 109-122). New York, NY: Oxford University Press.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12*(1), 1-48.

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science, 15*(10), 661-667.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity? *Intelligence, 14*(4), 389-433.

Lee, F. J., & Anderson, J. R. (2001). Does learning a complex task have to be complex?: A study in learning decomposition. *Cognitive Psychology, 42*(3), 267-316.

Locke, E. A. (2000). Motivation, cognition, and action: An analysis of studies of task goals and knowledge. *Applied Psychology: An International Review, 49*(3), 408-429.

Locke, E. A., & Latham, G. P. (1990). A theory of goal setting & task performance. *Upper Saddle River, NJ: Prentice Hall*.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist, 57*(9), 705-717.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*(4), 492-527.

Luchins, A. S., & Luchins, E. H. (1959). *Rigidity of behavior: A variational approach to the effect of Einstellung*. Oxford: University Press.

Luchins, A. S., & Luchins, E. H. (1991). Task complexity and order effects in computer presentation of water jar problems. *Journal of General Psychology, 118*(1), 45.

Mace, C. (1935). Incentives: some experimental studies. *Reports from the Industrial Health Residency Board of London (Vol 72)*.

Mackintosh, N., & Bennett, E. (2003). The fractionation of working memory maps onto different components of intelligence. *Intelligence, 31*(6), 519-531.

Mawer, R. F., & Sweller, J. (1982). Effects of subgoal density and location on learning during problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*(3), 252-259.

Mayer, R. E., Mautone, P., & Prothero, W. (2002). Pictorial aids for learning by doing in a multimedia geology simulation game. *Journal of Educational Psychology, 94*(1), 171-185.

Miller, C. S., Lehman, J. F., & Koedinger, K. R. (1999). Goals and learning in microworlds. *Cognitive Science, 23*(3), 305-336.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review, 63*(2), 81-97.

Murphy, K. R., & Davidshofer, C. O. (1998). Psychological Testing (4th ed.).

Newell, A. (1973). Production systems: Models of control structures. In W. G. Chase (Ed.), *Visual information processing* New York: Academic Press.

Newell, A., Shaw, J., & Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychological Review, 65*(3), 151-166.

Newell, A., & Simon, H. A. (1961). Computer simulation of human thinking. *Science, 134*, 2011-2017.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Oxford: Prentice-Hall.

Novick, L. R., & Holyoak, K. J. (1991). Mathematical problem solving by analogy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(3), 398-415.

Osman, M. (2008). Observation can be as effective as action in problem solving. *Cognitive Science, 32*(1), 162-183.

Owen, E., & Sweller, J. (1985). What do students learn while solving mathematics problems? *Journal of Educational Psychology, 77*(3), 272-284.

Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84*(4), 429-434.

Paas, F. G., Camp, G., & Rikers, R. (2001). Instructional compensation for age-related cognitive declines: Effects of goal specificity in maze learning. *Journal of Educational Psychology, 93*(1), 181-186.

Paas, F. G., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*(1), 63-71.

Paas, F. G., & Van Merrienboer, J. J. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors, 35*(4), 737-743.

Paas, F. G., & Van Merrienboer, J. J. (1994a). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills, 79*(1, Pt 2), pp.

Paas, F. G., & Van Merrienboer, J. J. (1994b). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology, 86*(1), 122-133.

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58*(3), 193-198.

Piaget, J. (1928). *Judgment and reasoning in the child*. London: Routledge and Kegan Paul.

Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and Instruction, 12*(1), 61-86.

Raven, J. C., Raven, J., & Court, J. H. (1994). *Manual for Raven's progressive matricies and vocabulary scales. Section 4: The advanced progressive matricies*. Oxford: Oxford Psychologists Press.

Reder, L. M., Charney, D. H., & Morgan, K. I. (1986). The role of elaborations in learning a skill from an instructional text. *Memory & Cognition, 14*(1), 64-78.

Schnotz, W., & Kurschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review, 19*(4), 469-508.

Schunn, C. D., & Reder, L. M. (2001). Another source of individual differences: Strategy adaptivity to changing rates of success. *Journal of Experimental Psychology: General, 130*(1), 59-76.

Seijts, G. H., Latham, G. P., Tasa, K., & Latham, B. W. (2004). Goal Setting and Goal Orientation: An Integration of Two Different Yet Related Literatures. *Academy of Management Journal, 47*(2), 227-239.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84*(2), 127-190.

Simon, D. P., & Simon, H. A. (1978). Individual differences in solving physics problems *Siegler, Robert S [Ed] (1978) Children's thinking: What develops?* (pp. 325-348). xi, 371 pp. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc; England.

Simon, H. A., & Lea, G. (1974). Problem solving and rule induction: A unified view. In L. W. Gregg (Ed.), *Knowledge and cognition*. Oxford: Lawrence Erlbaum.

Simon, H. A., & Newell, A. (1971). Human problem solving: The state of the theory in 1970. *American Psychologist, 26*(2), 145-159.

Simon, R. I. (1970). Encoding effects on complex problem solving. *Journal of Experimental Psychology, 83*(2), 227-231.

Singley, M. K., & Anderson, J. R. (1989). *The transfer of cognitive skill.* Cambridge, MA: Harvard University Press; US.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science: A Multidisciplinary Journal, 12*(2), 257-285.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer.

Sweller, J., Chandler, P., Tierney, P., & Cooper, M. (1990). Cognitive load as a factor in the structuring of technical material. *Journal of Experimental Psychology: General, 119*(2), 176-192.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 2*(1), 55-89.

Sweller, J., & Gee, W. (1978). Einstellung, the sequence effect, and hypothesis theory. *Journal of Experimental Psychology: Human Learning and Memory, 4*(5), 513-526.

Sweller, J., & Levine, M. (1982). Effects of goal specificity on means-ends analysis and learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*(5), 463-474.

Sweller, J., Mawer, R. F., & Howe, W. (1982). Consequences of history-cued and means-end strategies in problem solving. *The American Journal of Psychology, 95*(3), 455-483.

Sweller, J., Mawer, R. F., & Ward, M. R. (1983). Development of expertise in mathematical problem solving. *Journal of Experimental Psychology: General, 112*(4), 639-661.

Sweller, J., van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251-296.

Tarmizi, R. A., & Sweller, J. (1988). Guidance during mathematical problem solving. *Journal of Educational Psychology, 80*(4), 424-436.

Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin, 133*(6), 1038-1066.

van Merrienboer, J. J. (1990). Strategies for programming instruction in high school: Program completion vs. program generation. *Journal of Educational Computing Research, 6*(3), 265-285.

Vollmeyer, R., & Burns, B. D. (1996). Hypothesis and goal specificity: Conditions that influence the learning and control of a complex system. *Zeitschrift fur Experimentelle Psychologie, 43*(4), 657-683.

Vollmeyer, R., & Burns, B. D. (2002). Goal specificity and learning with a hypermedia program. *Experimental Psychology, 49*(2), 98-108.

Vollmeyer, R., Burns, B. D., & Holyoak, K. J. (1996). The impact of goal specificity on strategy use and the acquisition of problem structure. *Cognitive Science, 20*(1), 75-100.

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content*. Oxford: Harvard University Press.

Wertheimer, M. (1959). *Productive thinking*. Oxford: Harper.

Winfred, A., & Woehr, D. J. (1993). A confirmatory factor analytic study examining the dimensionality of the Raven's Advanced Progressive Matrices. *Educational and Psychological Measurement, 53*(2), 471-478.

Winters, D., & Latham, G. P. (1996). The effect of learning versus outcome goals on a simple versus a complex task. *Group & Organization Management, 21*(2), 236-250.

Wirth, J., Kunsting, J., & Leutner, D. (2009). The impact of goal specificity and goal type on learning outcome and cognitive load. *Computers in Human Behavior, 25*(2), 299-305.

Woltz, D. J., Gardner, M. K., & Bell, B. G. (2000). Negative transfer errors in sequential cognitive skills: Strong-but-wrong sequence application. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(3), 601-625.

Wood, R. E., Mento, A. J., & Locke, E. A. (1987). Task complexity as a moderator of goal effects: A meta-analysis. *Journal of Applied Psychology, 72*(3), 416-425.

Zeitz, C. M. (1997). Some concrete advantages of abstraction: How experts' representations facilitate reasoning. In P. J. Feltovich, K. M. Ford & R. R. Hoffman (Eds.), *Expertise in context: Human and machine* (pp. 43-65). Cambridge, MA: MIT Press.