The Availability and Cost of Accessing Information from the Interface: Consequences for Memory, Planning and Learning

Samuel M. Waldron

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Thesis submitted to Cardiff University for the degree of Doctor of Philosophy

April, 2007

Declaration

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

Signed .... (candidate) Date 07/12/07

Statement 1

This thesis is being submitted in partial fulfilment of the requirements for the degree of Ph.D.

Signed .... (candidate) Date 07/12/07

Statement 2

This thesis is the result of my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by explicit references.

Signed .... (candidate) Date 07/12/07

Statement 3

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

Signed .... (candidate) Date 07/12/07

Statement 4

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loans **after expiry of a bar on access** approved by the Graduate Development Committee.

Signed .... (candidate) Date 07/12/07

Contents

CHAPTER 1: GENERAL INTRODUCTION

CHAPTER 2: EMPIRICAL SERIES 1
Designing information fusion for the encoding of visual-spatial information

CHAPTER 3: EMPIRICAL SERIES 2
Planning with information access costs in mind

CHAPTER 4: EMPIRICAL SERIES 3
Learning solution to the eight-puzzle with varying goal-state accessibility

CHAPTER 5: GENERAL DISCUSSION

Index of tables

Index of figures

Publications

Much of the work reported in Chapter 2 also appears in the following journal article:

Waldron, S. M., Patrick, J., Duggan, G. B., Banbury, S., & Howes, A. (in press). Designing information fusion for the encoding of visual-spatial information. *Ergonomics*.

Some of the data reported in Experiment 2.1 have also been published in the following conference proceedings article:

Waldron, S. M., Duggan, G. B., Patrick, J., Banbury, S., & Howes, A. (2005). Adaptive information fusion for situation awareness in the cockpit. *Proceedings of the 49th Annual Meeting of the Human Factors and Ergonomics Society*, 49-53.

Some of the data reported in Experiment 2.3 are also reviewed in the following journal article:

Waldron, S. M., Patrick, J., Morgan, P., & King, S. (in press). Influencing cognitive strategy by manipulating information access. *The Computer Journal.*

Much of the data reported in Experiment 3.1 also appear in the following conference proceedings article:

Waldron, S. M., Patrick, J., Howes, A., & Duggan, G. B. (2006). Planning with information access costs in mind. *Proceedings of the 28th Annual Meeting of the Cognitive Science Society*, 2335-2340.

Chapter 3 is currently under revision for publication as the following journal article:

Waldron, S. M., Patrick, J., & Duggan, G. B. (under review). The influence of information access cost during problem solving: Consequences for memory and planning. *Memory & Cognition.*

Some of the data reported in Experiment 4.1a have also been published in the following conference proceedings article:

Morgan, P. L., Waldron, S. M., King, S., & Patrick, J. (2007). Harder to access, better performance? The effects of information access costs (IACs) on strategy and performance in display-based tasks. *Lecture notes in Computer Science*, 115-125.

The work reported in Chapter 4 is being prepared for publication as the following journal article:

Waldron, S. M., & Patrick, J. (under preparation). Learning to solve the eight-puzzle with varying goal-state accessibility.

Many of the theoretical concepts discussed throughout this thesis are also reviewed in the following conference proceedings article:

Duggan, G. B., Banbury, S., Howes, A., Patrick, J., & Waldron, S. M. (2004). Too much, too little, or just right: Designing data fusion for situation awareness. *Proceedings of the 49th Meeting of the HFES*, 528-532.

Acknowledgements

I would firstly like to thank my current supervisor John Patrick for encouraging me to think long and hard about my work, and former supervisors Andrew Howes and Simon Banbury for invaluable advice and input along the way. Much appreciation is also indebted to office mates past and present - Geoff 'you're a tiger' Duggan, Phil Morgan, and Alastair McKenzie-Kerr for putting up with my intolerance for being too hot or too cold, and my occasional snoozes!

Gratitude is also indebted to all residents of the 64PP team for much encouragement over the years, 'Irish Steve' for programming assistance, and the football gang for much needed respite. I would like to say thank you to my parents for raising me to 'give it a go', and always being there for me no matter what. Close friends from Cardiff: Cerith, Parky, Suze and Tigger also deserve an accolade for putting up with my over-zealous complaints of exhaustion over the last three years, and significant others (Amy, Bing, Lammy, Ryan, Tara, Tom, Whitmoor) for helping me to enjoy many a good time.

A huge thank you also goes to a very special person, Joanna Barrett, for all her words of wisdom, support, extreme patience and caring.

This work was supported by two UK, MoD Data and Information Fusion, Defence Technology Centre grants: Designing integrated displays to support team situation awareness. (J. Patrick, S. Banbury, A. Howes & D. Jones, 2003-2006); Multivariate information management and exchange (J. Patrick & D. Jones, 2006-2009).

Summary

Recent developments in technology have meant that operators of complex systems, such as those found in the modern aircraft cockpit, now have access to an unprecedented volume of information. Significant attention within the Human Factors and Ergonomics community has therefore been focussed upon developing methods by which externally represented information can be made as accessible as possible within the interface. However, commensurate attention has not been paid to recent experimental work demonstrating that even millisecond changes to the accessibility of information provided within the interface can have surprisingly large consequences for the deployment of human memory during low-level routine interactive behaviour (Gray, Simms, Fu, & Schoelles, 2006). Therefore, a series of nine studies explored the impact of information accessibility upon more complex human behaviour, with particular emphasis placed upon learning, memory and planning. Three experiments contained within Chapter 2 point to caveats associated with the use of information fusion as a means of increasing the availability of onscreen information during a series of simulated flight navigation missions. Paradoxically, increasing information availability was found to lead to greater problems for retention of visual-spatial information. The cost of accessing problem solving information was manipulated across three further experiments reported in Chapter 3. Least-effort tradeoffs concerning the use of memory, previously observed during routine copying (Gray et al., 2006), were extended to problem solving and were found to also influence planning behaviour, as reflected by eye-tracker and verbal protocol data. The final three experiments constituting Chapter 4 demonstrated expedited learning during repeated problem solving when task-relevant information was harder to access. Performance deficits were observed when interactive behaviour was characterised by excessive reliance upon highly accessible externally represented information. The results are contrasted to similar work conducted in the training literature, and future directions for exploiting information access costs to facilitate learning, memory and planning are discussed.

Chapter 1

General Introduction

> *"Our situation seems paradoxical: more and more data is available in principle, but our ability to interpret what is available has not increased. On the one hand, all participants in a field of activity recognise that having greater access to data is a benefit in principle. On the other hand, these same participants recognise how the flood of available data challenges their ability to find what is informative or meaningful for their goals and tasks."* (Woods, Patterson, & Roth, 2002, p. 23)

1.1. The paradox of increased information availability

Recent developments in technology have brought with them a proliferation of data sensors and sources. As a consequence, operators of complex systems, such as those found in the modern aircraft cockpit, now have access to an unprecedented volume of information. However, the problem faced by the Human Factors and Ergonomics community is how to effectively organise and present this information to the human operator. The focus of the current thesis will be on how small changes to the accessibility of information provided within an external display affect human behaviour, and may be used to promote certain behavioural strategies above others. The term external display is intended to refer to an interface designed to provide supplementary information about an event that cannot be directly perceived. The accessibility of information is determined by the time, physical and mental effort required when attempting to retrieve and process information from ones' task environment.

In most socio-technical systems, it is not possible (or in fact desirable) to provide operators with direct observation of events and processes (Lintern, Waite, &

Talleur, 1999). However, each round of technological advances, whether in electrical engineering, computer science, or artificial intelligence, does bring the promise of a better external interface within which to represent more indirect information (Woods et al., 2002). Despite the fact that developed cultures now live in what has been termed an 'information rich world' (e.g., consider the sheer volume of information available within the World Wide Web), our ability as humans to sort through and integrate great quantities of information has not developed so rapidly (Pirolli & Card, 1999). Information is no longer a scarce resource. However, processing capacity and attention still is (Miller, 1956), and problems emerge when information processing limits are reached (see Figure 1.1).

Figure 1.1: An illustration of the information overload problem.

Note. This "Bizarro" cartoon is reprinted by permission of Dan Piraro.

The explosion of available information has brought with it many instances of information overload (Kirsh, 2000), whereby the quantity of information available exceeds the processing limitations of the operator (Simon, 1981). Not surprisingly, research has generally found decision quality to degrade as a function of information overload (e.g., Hwang & Lin, 1999), and a number of approaches to cognitive engineering (e.g., Ecological Interface Design – Rasmussen & Vicente, 1989, Configural Display Design – Woods, Wise, & Hanes, 1981; Proximity Compatibility Principle –Wickens & Carswell, 1995) have attempted to provide design solutions to increase the ease with which operators of a complex system can extract information required from the interface.

Although these approaches generally report reductions in information overload (see Wickens & Hollands, 2000 for a review), the current thesis argues that increasing the ease with which information can be extracted from an interface is not the complete answer, is not a generic solution, and may under certain conditions reveal unwanted side effects. Instead, it is proposed that harmonising intelligent display design with knowledge of the human information processing system may allow the interface to elicit the required behaviour on behalf of the operator, and circumvent the limited capacity of human memory processing boundaries. In order to achieve such goals, it is argued that particular emphasis must be placed upon the adaptive nature of cognition (Anderson, 1990; Payne, Howes, & Reader, 2001), and the interaction between human behaviour and the structure of one's socio-technical environment.

The remainder of Chapter 1 will be split into two conceptual halves. The first half will evaluate popular approaches to cognitive engineering that have attempted to ameliorate instances of information overload. The second half will review major lessons learned from the automation literature, and recent experimental research from the fields of cognitive and psychological science, in order to instantiate the claim that instead of simply designing displays to increase information accessibility, a more promising approach may be to design the interface so as to orient operator behaviour towards the selection of certain microstrategies, and away from others.

1.2. Contemporary solutions to information overload

The ever increasing volume of information to be represented within complex system displays has signified the demise of the single-sensor single-indicator (SSSI) display format. Pre-1980s, the majority of display design was limited to SSSI constraints, whereby each indicator represented the value of a single measurement (Goodstein, 1981; Woods, 1991). The major downfall centred upon the fact that several separate displays were often needed to represent the information required to make a single decision. For example, thirty spatially separate indicators were traditionally required to represent the pressure-versus-temperature relationship within a nuclear power plant control room (Vicente, Moray, Lee, Rasmussen, Jones, Brock, & Djemil, 1996). Operators working within a SSSI environment were required to sequentially gather the information needed from individual instruments (often spatially separate), maintain this information in memory, and mentally integrate the information collected to arrive at a decision (Vicente et al., 1996). These processes of information gathering and integration imposed high, sometimes overwhelming, cognitive demands upon the operator, taxing limited resources such as attention and memory (Bennett & Flach, 1992).

The following section briefly reviews some major attempts to improve upon the SSSI format. As will become evident, each approach attempts to increase the ease with which the operator can extract information from the interface. However, concern is raised that insufficient attention has been paid to the interaction between internal and external representations of the task space (Norman, 1993). In addition, these approaches are criticised on the basis that the application of each depends largely upon preconceived analyses of the task space (Kirwan & Ainsworth, 1992), and are thus often situation-specific and non-adaptive.

1.2.1. Integrated display design

Ecological interface design (EID), for example, is largely based upon conceptual tools developed by Rasmussen (1983; 1985), such as the skills, rules, knowledge taxonomy and the abstraction hierarchy. The abstraction hierarchy is used as a functional decomposition of the work environment to determine the kinds of information that should be provided within the interface. Based upon an analysis of cognitive workload determined by the likely deployment of skills, rules and knowledge requirements (representing automatic, rule-based and advanced processing), information is organised and presented within the interface so as to take advantage of human perception and psychomotor abilities.

The task of information extraction is made easier, in the hope that the human information processing system can then concentrate efforts on more complex tasks such as problem solving. EID proposes that complex relationships between variables can be made directly accessible to operators in a manner that allows effortless extraction of information from the interface using powerful perceptual capabilities (Pawlak & Vicente, 1996; Vicente & Rasmussen, 1992). Multiple system views are encouraged that provide both physical (traditionally contained within SSSI) and functional (higher-order) information. The facility to 'drill-down' into higher-order functional information is recommended in order to reveal lower-level physical data when required.

Although a number of case studies have found EID to be effective at supporting operator decision making (Marino & Mahan, 2005) and the detection of faults within a system network (Vicente et al., 1996), EID is not based upon knowledge of the human information processing system. It is argued here that insufficient attention has been paid to examining how such alterations to the external representation of one's task environment may impact upon the deployment of human cognition and the development of internal representations. This also applies to Lind's (1990) approach to 'multi-level flow modeling', which shares many of the aims of EID.

Similar criticisms can also be made of configural display design (CDD). Much emphasis is placed upon the mapping of information from multiple sources onto highly recognisable shapes used as foundations for emergent features (Bennett & Flach, 1992). By mapping multiple variables into a single geometric shape, high level visual properties such as closure and symmetry are provided (Woods et al., 1981) which can exploit humans' highly developed pattern recognition capabilities (Gibson, 1979), and thus increase the ease with which information is extracted from the interface (e.g., Buttigieg & Sanderson, 1991; Dinadis & Vicente, 1999; Jones, Wickens, & Deutsch, 1990). As with EID, CDD has also received much support. For example, in their comprehensive review of graphical displays, Bennett & Flach (1992) acknowledge that:

> "*There appears to be a clear consensus that performance can be improved by providing displays that allow the observer to utilize the more efficient processes of perception and pattern recognition instead of requiring the observer to utilize the cognitively intensive processes of memory, integration, and inference.*" (p. 514)

Again, however, the potential consequences to human cognition of such display design manipulations have not received commensurate attention. Even the most generic of the many approaches to integrated display design, the proximity compatibility principle (PCP), which is based more firmly upon an appreciation of what we know about the human information processing capabilities and boundaries (Wickens & Carswell, 1995), has yet to sufficiently scrutinise the effects of increasing information accessibility within the interface upon human interactive behaviour.

Although the principle of rendering displays relevant to a common task or mental operation close together in perceptual space (Carswell & Wickens, 1987) has gathered much support (Wickens & Hollands, 2000), PCP still requires some form of task analysis (Kirwan & Ainsworth, 1992) before it can be implemented. For this reason, PCP, along with EID and CDD, cannot be adaptive to the user unless one design can accommodate the diversity of operator requirements. In addition, Wickens & Carswell (1995) limit their discussion of PCP to the disadvantages that occur when multiple sources of information provided within an interface are positioned spatially far apart. They do not consider the potential advantages that could be predicted from the recent adaptive cognition literature (e.g., Gray & Fu, 2004).

1.2.2. Information fusion

Despite the concerns raised above, the popularity of integrating information from multiple sources as a means of minimising information overload has led to rapid developments in 'information fusion' technologies more generally. Recent developments in computer science and electrical engineering have provided interface designers with increased capability to fuse information traditionally acquired from a number of disparate sources (Dasarathy, 2001), and fusion technologies are fast becoming instantiated within complex system design (Xiong & Svensson, 2002). Again, however, the use of fusion does not consider the role of human information processing. Instead, fusion research is largely fuelled by disciplines such as computer science, and because still in its infancy, (as reflected by the relatively recent commencement of a major journal reporting developments within the field - 'Information Fusion' - 2000), has largely escaped evaluation from a Human Factors perspective. For example, much work has been conducted on areas such as the fusion of multi-sensor information relating to the tracking of objects during modern aircraft flight (e.g., Powell, Marshall, Milliken, & Markham, 2004), yet this work does not question the assumption that the human processor will inevitably benefit from the provision of highly accessible, integrated information. The role of the human information processing system has not been analysed as an integral part of the socio-technical system within which information fusion is anticipated to reside.

The current thesis, therefore, aims to provide the first in-depth examination of how small changes to the accessibility and availability of information provided within an interface will affect the manner in which operators interact with such information. In particular, a novel avenue of research will be pursued whereby small manipulations to information accessibility are used intelligently to encourage certain behavioural strategies on behalf of the user, yet discourage others. Investigation of this topic is not intended to extend to the often complex process by which more established approaches to integrated display design (e.g., EID, CDD, PCP) attempt to provide operators with higher-order functional information within a specific unified display. Rather, empirical assessment will be limited to the ramifications of small changes to the accessibility of information provided within an interface. Before reviewing relevant research from the fields of cognitive and psychological science to help guide such an investigation, major lessons learned from the use of automation to reduce operator workload will be addressed, as these may also inform the appropriate use of information fusion.

1.3. Lessons learned from the automation literature

In an attempt to reduce the cognitive workload experienced by operators of complex systems, many tasks have been automated over the last twenty years. However, a number of challenges have been faced in doing so. Defined as "the execution by a machine agent (usually a computer) of a function that was previously carried out by a human" (Parasuraman & Riley, 1997, p. 231), what is considered automation is continuously changing with time as technological advancements allow more and more tasks to be automated. The fact that computers are now being used to integrate information from a variety of sources to present the human operator with highly accessible fused information (Dasarathy, 2001; Bennett & Flach, 1992; Vicente, 2002) suggests that information fusion could be seen as a new addition to the array of automated technologies. It follows then, that a review of the challenges faced by the automation community will prove beneficial when considering the consequences of partially automating the assimilation of information from multiple sources in order to increase the accessibility of information provided within the interface.

1.3.1. 'Out of the loop' phenomenon

Researchers examining the interaction between automation and human behaviour have found that, under certain circumstances, the use of automation can leave the human feeling 'out of the loop' (Adams, Tenney, & Pew, 1995; Bainbridge, 1987; Norman, 1990; Sarter & Woods, 1995). By this it is meant that the human is left 'unaware' of the current state of an automated task and/or the processes responsible for a particular system status. Although very few high-level cognitive tasks (e.g., problem solving, decision making) are fully automated, the processes leading up to decision making often are. Within such scenarios, the human operator may find it difficult to resume control and make the correct decision based upon the information available. It seems quite plausible that similar situations may arise as a consequence of automating the gathering and assimilation of information. Operators presented with highly accessible fused information may also feel 'out of the loop', having not internally processed the information presented to them.

Separating humans from processing activities can also impair an operator's internal representation of a situation (Parasuraman & Riley, 1997). For example, air traffic controllers have reported problems maintaining a mental picture of the task space when automated conflict resolutions are used under times of high workload (Whitfield, Ball, & Ord, 1980). Too much automation has also been cited as leading to degradation in skills previously performed routinely by the operator (e.g., Lee & Moray, 1994). Although this does not generally pose too much of a problem (because the task is automated), technical failure will leave the operator in the unenviable position of resuming a task that is no longer routinely performed, and thus largely forgotten (Wickens & Hollands, 2002). Indeed, with respect to the provision of highly accessible information within the interface, Reising and Sanderson (2004) have highlighted similar difficulties caused by instrument failure when working with an EID display. In particular, they cited consequences to be more severe where information from multiple sources had been automatically integrated than when it had not.

1.3.2. Over-reliance & complacency

Over-reliance upon automated processes can also lead to problematic behaviour if the information provided as a result of an automated process is taken as definitive. Mosier, Skitka, Heers and Burdick (1997), for example, demonstrated that even experienced pilots tend to use automated cues as heuristic replacements for information seeking, and tend to rely upon these cues despite conflict between expected and actual automation performance. An irony of automation is that the more reliable an automated process is perceived as being, the more complacent the human operator becomes (Bainbridge, 1987; Parasuraman, Molloy, & Singh, 1993). It is argued that the information fusion literature would do well to heed such findings, as it seems quite plausible that making information too accessible within the interface may also lead to over-reliance upon fusion process and complacency on behalf of the operator, which may have negative ramifications.

Although reducing workload via automation has many benefits (Parasuraman & Riley, 1997), it is clear that if not used appropriately problems may arise. As already argued, many of the caveats associated with the use of automation may apply to methods of increasing information accessibility via information fusion. In an attempt to promote awareness of some of these issues, the current thesis will provide an examination of the effects small changes to the accessibility of information provided within an interface may have upon human behaviour.

1.4. Psychological consequences of manipulating the accessibility of information provided within an external display

In order to understand the link between interface design and human performance, it is important to fully appreciate the effect of the constraints of design upon the psychological processes selected to deal with particular design details. One perspective advocated by Garbis (2002) has emphasised the importance of considering the affordances provided by the human-machine interface. The 'cognitive artefacts' approach highlights how the processing and structure of tasks may be affected by the constraints and facilities available at the interface. The following evaluation of the potential impact of changes to information accessibility upon operator behaviour will follow from this perspective.

It is anticipated that increasing the accessibility of information provided externally will not always improve human performance, and under certain situations will elicit non-optimal behaviour. A review of the psychological literature that considers human behaviour within the context of environments requiring the support of external displays will ensue.

1.4.1. The adaptive character of thought

Borrowing from the title of Anderson's (1990) seminal work on adaptive cognition, this section will emphasise the typically adaptive nature of human interaction with a task environment. Importantly, the manner by which an individual interacts with an external display is not determined purely by the internal makeup of human cognition. Rather, human behaviour will often be affected by the design characteristics of the interface one is working with. According to Anderson's (1990) theory of the Adaptive Character of Thought (ACT), human behaviour can be explained by a series of cost-benefit tradeoffs, based upon the statistical properties constituting one's task environment. Human cognition is understood on the assumption that adaptation takes place to create optimal solutions to the information processing problems posed by particular environments.

Although Anderson repeatedly refers to adaptive cognition as an example of 'rational analysis', his conception of rationality is rather aligned with Simon's (1956) theory of 'bounded rationality'. Not only does the task environment affect human behaviour, but cognitive constraints also impose bounds upon underlying mechanisms deeming certain computations implausible. In addition, pragmatic constraints such as time and available information give rise to bounds of rationality. An updated revision of the ACT theory (ACT-R - Anderson & Lebiére, 1998) attempts to integrate Anderson's earlier work on the mechanistic properties of human cognition (Anderson, 1983) with his subsequent view of rational analysis (Anderson, 1990). Recent models of 'constraint satisfaction' have also focussed upon the integration of mechanistic and rational views of human cognition (see Howes, Vera, & Lewis, in press for further discussion on this topic).

Most relevant to the current discussion is the assertion made by Anderson and colleagues (Anderson, 1990; Anderson & Milson, 1989; Anderson & Lebiére, 1998; Anderson & Schooler, 1991; 2000) that human memory is an adaptive entity. The deployment of memory is argued to depend upon three factors. Firstly, for each retrieval from memory, there is argued to be an associated cost (C), which could entail time or mental effort. Secondly, if retrieval is successful there will be some gain (G), provided that the retrieval is rendered useful within the current context. Finally, a probability value (P) is assigned in advance to retrieving the memory according to whether it is likely to be relevant or not. Given these three components concerning the use of human memory, Anderson and colleagues have used the following formula to express when memory is not to be used, $PG < C$. The human memory system will stop considering memories when the probable use and gain of a particular retrieval are outweighed by the cost associated with retrieval.

Anderson and colleagues argue that the human memory system takes the form it does so as to make most available memories for information that are currently required (Anderson & Schooler, 2000). Indeed, when analysing the setting of goals during solution of the 'Tower of Hanoi' problem, Anderson & Douglass (2001) noted that since it was easy to reconstruct goals from the information provided within the external display, the price of memory failure was not high. For this reason, Anderson & Douglass (2001) argued that there was little motivation to engage in goal rehearsal during externally supported solution of the Tower of Hanoi task. With regard to display design, this suggests that when information is highly accessible within the interface, memory for visual-spatial information may be problematic due to a low probability of usage. In contrast, better memory may be predicted when information is less accessible within an interface due to the increased likelihood that memory for externally presented information will be beneficial. This major research question will receive particular attention within Chapters 2 and 4 of this thesis.

In addition to the adaptive nature of memory, ACT theory also states that selection between different problem solving strategies may also be determined by adaptive responses to constraints embedded within the task environment (Anderson, 1990). Rational analysis of problem solving, according to ACT, requires consideration of the goal of the system, the structure of the environment, and a set of computational constraints (e.g., memory span). Continued search of a problem space is argued to take place at increasing cost, and theories of optimality predict that partial plans will only continue to be generated until their expected benefits are outweighed by the costs of mental search (Payne et al, 2001). A good example of a factor associated with display design that may affect the cost-benefit analysis of planning has been termed the 'implementation cost' (O'Hara & Payne, 1998; 1999), that is, the cost of performing an operation upon the world.

One particular instantiation of the implementation cost explored by O'Hara & Payne (1998) was the cost associated with making each move during solution of an eight-puzzle problem. Essentially, O'Hara & Payne observed that as the cost of making moves during problem solving increased, so did levels of planfulness. Interpretation of these results led O'Hara & Payne to propose that the additional cost associated with making a move induced a shift in cost-benefit analysis, such that mental planning was perceived as more advantageous than display-based acting and evaluating (cf. Kirsh & Maglio, 1994). Evaluation of move sequences was argued to take place internally, rather then externally, when implementation costs were higher so as to reduce the overall time to complete the task.

Much display design currently strives to provide operators of complex systems with a direct manipulation interface in an attempt to reduce the information processing distance between the user's intentions and the facilities offered within the interface (see Hutchins, Hollan, & Norman, 1985 for a review). However, in light of the work cited above (O'Hara & Payne, 1998), this may not always be of benefit; especially if efficient planful behaviour is required of the user. Although the ease with which actions can be performed is not the focus of the current thesis, a comparison between increased implementation cost and increased information access cost during solution to an eight-puzzle-like task can be found in Experiment 3.1.

As detailed above in Sections 1.2 and 1.3, considerable emphasis has recently been placed within the fields of cognitive engineering and information fusion upon providing operators of complex systems with highly accessible integrated information. In light of the adaptive nature of memory and problem solving

(Anderson, 1990; Anderson & Milson, 1989; Anderson & Schooler, 1991; 2000; O'Hara & Payne, 1998; 1999; Payne et al, 2001), it is anticipated that increased accessibility to information within the interface will have consequences for how an operator uses such information. What follows is a more elaborate instantiation of this assertion within the context of a recent model of interactive behaviour (the 'soft constraints hypothesis' - Gray & Fu, 2004; Gray, Simms, Fu, & Schoelles, 2006) specifically implemented to address the cost-benefit tradeoffs associated with manipulating the accessibility of information within an external display.

1.4.2. Soft constraints hypothesis

Recent extensions of Anderson's (1990) work on rational analysis have begun to investigate how very small changes within the design of an interface can have surprisingly large consequences for the selection between microstrategies during interactive behaviour (e.g., Gray & Fu, 2004; Gray et al., 2006; Lohse & Johnson, 1996). As noted by Gray & Boehm-Davis (2000):

> "*Microstrategies develop in response to the fine-grained details of the interactive technology; that is, microstrategies focus on what most designers would regard as the mundane aspects of interface design: the ways in which subtle features of interactive technology influence the ways in which users perform tasks.*" (p. 322)

A number of authors have examined the role of external, as well as internal, representations in supporting human behaviour during a variety of tasks (e.g., Hutchins, 1995b; Larkin, 1989; Payne, 1991; Vallée-Tourangeau & Penney, 2005; van Nimwegen, van Oostendorp, & Tabachneck-Schijf, 2004; 2005; Zhang, 1997; Zhang & Norman, 1994). As discussed at length within the special issue of Cognitive

Science devoted to 'situated action' (Vol. 17, 1993), cognition can be said to be situated in-the-world as well as in-the-head (Norman, 1993), and changes within the external representation of information provided within a display can have substantial consequences for the use of internal cognition.

When alternative microstrategies can be applied during interactive behaviour, users tend to select the one that is most efficient given the task constraints (Gray & Boehm-Davis, 2000). This selection between microstrategies has been shown to be embodied at the level of milliseconds (Ballard, Hayhoe, Pook, & Rao, 1997; Gray & Boehm-Davis, 2000), and has recently been formalised within a theory of soft constraints (Gray & Fu, 2004; Gray et al., 2006). Hard constraints, according to Gray and colleagues, determine behaviour that is, or is not possible. Soft constraints, on the other hand, determine which of the microstrategies available is most likely to be chosen. It is assumed that when selection between microstrategies is non-deliberate or automatic (Gray & Fu, 2004), soft constraints are determined by least-effort tradeoffs.

To make predictions concerning an operator's use of information situated within an external display, the soft constraints hypothesis proposes that equal weighting be given to the time required for perception, action and memory retrieval (Gray & Fu, 2004). Gray et al (2006) have empirically demonstrated that the deployment of perceptual-motor and cognitive resources during interactive behaviour are adjusted based upon temporal cost-benefit tradeoffs, and not a general conservation of cognitive effort (e.g., memory) as maintained by previous authors (e.g., Cary & Carlson, 1999; 2001; Wilson, 2002). Gray and colleagues repeatedly eschew the unqualified assumption rife within the cognitive literature that the use of memory is preserved whenever possible (Ballard, Hayhoe, & Pelz, 1995), and instead argue that the deployment of memory is an adaptation to a rational analysis of the cognitive and task environment (Anderson, 1990; Anderson & Schooler, 1991).

Of particular interest to the current thesis is the finding that very small manipulations concerning the cost associated with accessing information from an interface (in the realm of milliseconds) will affect the adaptive user's selection between interaction- and memory-intensive strategies (Fu & Gray, 2000; Gray & Fu, 2001; 2004; Gray et al., 2006). Essentially, as the cost of accessing information from an interface is reduced, selection is observed to favour interaction-intensive perceptual-motor strategies. In contrast, as the cost of accessing information from an interface increases, memory-intensive strategies are used more often so as to reduce the overall time to complete the task. This effect has been shown reliably during low-level routine VCR programming (Gray & Fu, 2001; 2004) and copying behaviour (Fu & Gray, 2000; Gray et al., 2006), but has yet to be extended to more complex task domains such as problem solving (cf. Pfeiffer, 2004). It is worth noting at this point that although much emphasis within this thesis will be placed upon understanding the adaptive use of memory as a function of small changes to information accessibility, comparably less attention will be paid to the mechanistic properties underlying human memory. This is not an oversight of the thesis, but reflects the desire to study human use of memory as a response to the statistical properties of the task environment (Anderson & Milson, 1989). The experimental tasks used herein will all be supported by external representations, thus significant attempts will be made to assess the extent to which internal working memory (Baddeley, 1986) is used during task performance to support/replace/supplement externalised information. Related to the issues examined within this thesis are previous demonstrations of failures of working memory in familiar tasks where information was readily available. For example, Morton (1967) found that despite many years experience using a letter-based telephone dial, when tested, individuals could not recall accurately the location of letters on the dial. Nickerson and Adams (1979) found similarly poor performance when individuals were asked to remember the detailed appearance of a US penny. These findings demonstrate that even when information is readily available in the world, we do not always process it sufficiently to allow for accurate recollection of detail. While the coins may reflect a failure of attentional processes, this would not have been the case for the phone dialling, nor will it be a potential explanation of the performance observed in this thesis.

1.5. Focus of thesis

Given that very small changes to the accessibility of information provided within an external display have previously been found to have surprisingly large consequences for the use of perception, action and memory during low-level routine tasks (Gray & Fu, 2004; Gray et al., 2006), it is postulated that a new line of research is required to investigate the possible consequences of increasing information accessibility upon human behaviour and performance in more complex task domains. It is anticipated that further investigation of this topic will demonstrate that simply increasing information accessibility within an interface will not always be of benefit. A potentially more promising approach to display design will be examined whereby the artefacts of the interface are designed so as to exploit the adaptive nature of human cognition and orient behaviour towards desired microstrategies, and away from others.

Specifically it is hypothesised that one side effect of increasing the accessibility of information provided within an interface may be over-reliance upon display-based interaction-intensive strategies, and under-reliance upon internally-based memory-intensive strategies. Indeed, similar arguments can be found within the training literature. For example, learning of information and procedures has previously been found to improve as a function of difficulties encountered during training in a number of domains (e.g., VCR programming - Duggan & Payne, 2001; aircraft landing - Lintern, 1980; text comprehension - McNamara, Kintsch, Songer, & Kintsch, 1996, motor learning - Schmidt & Bjork, 1992; problem solving - Sweller, 1988).

The experiments reported within this thesis are distributed between three thematically aligned experimental chapters. The first of which (Chapter 2) contains three experiments assessing the effectiveness of information fusion technologies currently in development at supporting operator performance both during and after simulated flight missions. In particular, consequences of information fusion in affecting the *availability* of information provided within the external display were evaluated with respect to operator performance. An examination of how increasing the availability of information within an interface would affect memory for visual-spatial information was conducted.

In an attempt to address more directly, and extend Gray and colleague's soft constraints hypothesis (Gray & Fu, 2004; Gray et al., 2006) to the domain of problem solving, Chapter 3 examined the effect of small manipulations to the *cost of accessing* information from an external display on human interactive behaviour. Three experiments evaluated the extent to which least-effort tradeoffs observed during routine copying would extend to a transformation problem solving task. Eye-tracking data were collected in order to make fine-grained inferences concerning the distribution of cognition internally and externally (Norman, 1993). Furthermore, the collection of verbal protocol data (Ericsson & Simon, 1993) allowed for inferences to be made regarding the effect of different access costs upon the selection between different planning strategies (Davies, 2003). The final three experiments found within Chapter 4 investigated whether small differences associated with the *cost of accessing* goal-state information would affect rate of learning and transfer during repeated solution of an eight-puzzle-like problem solving task. Chapter 5 closes with a general summary of the effects of increasing the accessibility of information provided within an external display upon human behaviour. Methodological issues and future directions for research are also discussed.

Chapter 2

EMPIRICAL SERIES 1

Designing information fusion for the encoding of visual-spatial information

---

2.1. Introduction

The aim of Chapter 2 was to provide a case study examining the use of fusion to increase the availability of information provided within a simulated cockpit display. Operators of modern aircraft cockpits have access to an unprecedented volume of information, originating from a variety of on- and off- board sensors. Thus, information fusion techniques are used ever more frequently in an attempt to increase the accessibility of information provided within the task environment. In doing so, it is possible that the operator's task of information extraction and assimilation is made easier and more effortless (e.g., Vicente, Moray, Lee, Rasmussen, Jones, Brock, & Djemil, 1996).

Despite a wealth of knowledge indicating that human behaviour is adaptive and sensitive to the constraints of the environment they find themselves in (reviewed in Chapter 1), investigation of possible cost-benefit tradeoffs and boundary conditions associated with the use of fusion has been largely neglected. Chapter 2 aimed to assess the possibility that performance on certain task criteria may be degraded by increased information availability. Based upon work demonstrating that internal memory is used less often as function of information being made more available within the interface (e.g., Gray & Fu, 2004), it is argued that the manner in which information is presented within a fused environment may have negative consequences for the encoding of visual-spatial information.

2.1.1. Memory as an evaluation of display design

The general philosophy during assessment of display design is not to evaluate a display based upon the extent to which it promotes the internalisation of information presented within the external display, but rather, to focus upon the effectiveness of the display in supporting the operator during information extraction and decision making. It is argued here, however, that assessment of memory may be particularly important during the evaluation of display manipulations that intend to increase the ease with which information can be extracted from the interface. Given that increasing the ease with which information can be accessed will reduce the cognitive processing required on behalf of the operator, it seems logical that an unwanted side effect may be problematic retention of such information.

Limited work that has investigated the value of using memory as a methodology for evaluating display effectiveness (e.g., Vicente, 1992), has generally proposed that the use of memory to evaluate display effectiveness is better suited to analyses of semantically meaningful variables, as opposed to detailed visual information (Sperling, 1960). One study that has directly compared the use of visual and memory methods for the evaluation of display design found interesting differences according to the methodology used (Bennett, Payne, Calcaterra, & Nittoli, 2000). Results from the visual methodology suggested that integrated displays were effective in supporting tasks that required both low- and high-level properties. However, results obtained using the memory methodology differed: although the integrated display supported performance for tasks requiring the retention of higher-order properties, this was not the case for tasks requiring the retention of low-level data. Memory for visual-spatial information thus appears a valid measure of display effectiveness (provided the information to be retained in memory has some semantic property), and may reveal inadequacies with fused displays at supporting the retention of externalised low-level information.

What follows is a brief review of relevant psychological literature suggesting that reducing the cognitive processing required to extract information from a fused environment may have negative consequences for the encoding of such information. Firstly, work is cited highlighting the importance of display design in promoting active engagement (on behalf of the operator) in processes that make use of memory and inference making (e.g., Gray, Simms, Fu, & Schoelles, 2006; McNamara, Kintsch, Songer, & Kintsch, 1996). Processes such as memory and inference making contribute to the development of a robust internal representation of the task environment, and can thus be seen as 'transfer-appropriate processing' (Morris, Bransford, & Franks, 1977) when memory for visual-spatial information is important. Secondly, a proposal is made that the use of information fusion will, in some situations, reduce the extent to which operators of complex systems engage in such transfer-appropriate processing.

2.1.2. Reduced encoding of highly accessible externally represented information

Experimental research has shown that even very small changes to the design of an interface can significantly affect the extent to which internal memory is deployed during interactive behaviour. For example, by increasing the cost of accessing information from a simple eye movement to a head movement, Ballard, Hayhoe and Pelz (1995) induced a shift in participants' behaviour from a largely display-based strategy to one more reliant upon working memory. Work continuing in this theme has indicated that the lower the cost associated with accessing information from a display, the less likely participants are to employ memory-intensive strategies during routine interactive behaviour (Fu & Gray, 2000; Gray & Fu, 2004). If information is readily available in-the-world, a shift is observed from memory-intensive strategies to more display-based ones, often referred to as perceptual-motor strategies (Gray et al., 2006), where the display is used as an external memory source (O'Regan, 1992). It follows then, that the use of internal memory by operators of complex systems may decrease as the task of accessing and integrating information from the task environment is made easier via information fusion. A reduction in the use of memory during information extraction from a fused display is likely to have negative consequences for the retention of visually presented information. Indeed, the frequency with which different memory traces are called upon is integral to many theories of declarative memory (e.g., Anderson & Milson, 1989).

In their discussion of the adaptive nature of memory, Anderson and colleagues (Anderson, Fincham, & Douglass, 1999; Anderson & Milson, 1989; Anderson & Schooler, 1991) have argued that activation of a memory trace is, at least in part, determined by retrieval practice. They propose that the human memory system has the form it does so as to make more available memories that are used more often in the past (the practice effect). It has also been suggested that memory's most apparent deficit, forgetting, may in fact be an adaptive response to the need to focus on currently available information (Anderson & Milson, 1989; Bjork & Bjork, 1992). Functional decay theory proposes that when a task requires memory to be updated frequently, decay must occur to prevent interference with later memories (Altmann & Gray, 2002; Venturino, 1997). It is quite probable, therefore, that a reduction in the use of memory during information extraction (as a result of increased information accessibility) will have negative consequences for the internal encoding and retention of externally represented information. In contrast, the more frequent use of memory required when information is less accessible externally, may result in the development of a more robust internal representation of the task space.

There is also a wealth of knowledge suggesting that relevant internal processing improves subsequent task performance relative to a passive reliance upon equivalent information provided within the environment (e.g., Duggan & Payne 2001; McNamara et al., 1996; Palmiter, Elkerton, & Bagget, 1991). The idea that the encoding of information can be improved via relevant processing can be dated back to the seminal 'levels of processing framework' developed by Craik and Lockhart (1972). Also, the inserted questions literature (e.g., Glover, 1989) and the text comprehension literature (e.g., McNamara, et al., 1996) have demonstrated that prompting participants to make task-related inferences whilst reading text can aid the comprehension and retention of text. Using the everyday task of programming a VCR, Duggan and Payne (2001) improved participants' retention of instructional information by prompting them to adopt a chunked instruction-following procedure (reliant upon memory). By reading and then performing several steps (in temporal order) of the programming cycle at a time (compared to one step at a time), participants were engaged in more internal processing during the training phase and consequently outperformed the 'one step at a time' group at test.

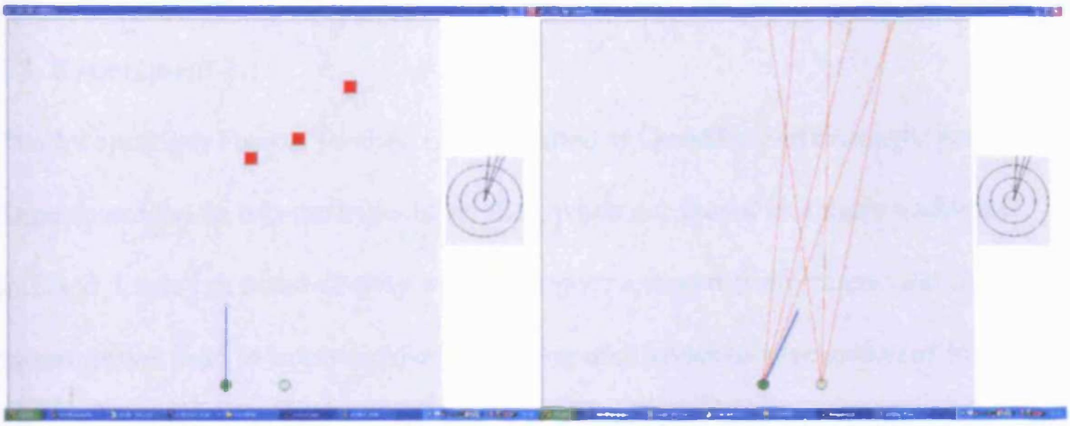
Engaging individuals in task-relevant inference making and memory-intensive strategies during task performance can be seen as an example of what Morris et al (1977) coined 'transfer-appropriate processing'. In essence, transfer-appropriate processing theory states that matching the cognitive demands experienced during learning with those at retrieval gives rise to better retention than mismatched learning and retrieval conditions. For example, if information is not permanently available during a task, or is presented in a random/variable fashion, the participant will receive practice during the task of forgetting and subsequently retrieving this information. Consequently, retrieval mechanisms will have been practised and therefore available at test. Relatedly, it is not always the case that a manipulation that maximises performance during the task will also benefit the retention of task-related information over time. In fact, manipulations that degrade the ease of acquisition during the task can often support the long-term retention of this information (Schmidt & Bjork, 1992). Following from 'cognitive load theory' (Chandler & Sweller, 1996; Sweller 1988), it has been recognised that so-called 'germane cognitive load' (van Merriënboer, Schuurman, de Croock, & Paas, 2002) can facilitate learning. Therefore it could be predicted that, within the limits of total available cognitive capacity, increased processing load associated with the extraction of information presented within an external display will facilitate the encoding and therefore retention of visual-spatial information.

Based upon the assumption that reducing the cost of accessing information within an interface will reduce the use of memory-intensive strategies during information seeking (e.g., Gray et al., 2006), and the assumption that relevant internal processing is required to maintain and update memory during interactive behaviour (e.g., Anderson & Milson, 1989; McNamara et al., 1996), it is predicted that under certain circumstances the use of information fusion may lead to impoverished encoding and poor retention of information presented within an external display. In order to test this hypothesis, the effectiveness of a fused and unfused cockpit display will be evaluated in terms of supporting operator performance, both during simulated flight missions, and when recalling mission information. A high-fidelity flight simulator was used in Experiment 2.1 to assess and compare the effectiveness of fusion technologies currently under development against more traditional unfused displays. Further studies (Experiments 2.2 and 2.3) were conducted using a low-fidelity simulation in an attempt to replicate and differentiate, under laboratory conditions, between competing explanations for the results observed in Experiment 2.1.

2.2. Information Fusion Testbed

Flight missions were used where the goal of the task was to navigate the aircraft within an area of interest and estimate the position of a number of locations (1, 2, or 3). For all experiments, the area of interest was represented by a 70 x 70 nautical mile (nm) terrain map. Although participants were never informed of the actual position of the location(s), dynamic indirect estimates of location position were provided within the interface in order to guide estimation. Following completion of each flight mission, participants' memory for location position(s) was tested. Contained within every square nm of the terrain map were a number of landmarks. Thus, memory for location information involved semantic properties and relational information.

During each flight mission a confederate aircraft flew alongside the participant's aircraft maintaining a constant separation of 10 nm, enabling information to be shared between the two platforms. The extent to which this information was fused within the interface was the focal point of Experiments 2.1 and 2.2, with Experiment 2.3 concentrating on manipulations to the temporal availability of fused information. Both fused and unfused displays contained radar information indicating direction but not range of location(s) relative to the participant's aircraft (see Figure 2.1). The representation and availability of dynamic location information projected onto the terrain map, however, differed according to whether fusion was, or was not present.



(a) Fused (b) UnFused

Figure 2.1: Representation of (a) fused and (b) unfused displays.

Note. Screenshots taken from a simulation of the IFT with three locations, omitting maps used. The intersections of unfused spokes furthest from both aircraft represent the best estimates of location.

Within the fused display, shared location estimates were displayed onscreen as small squares (see Figure 2.1a), which were permanently available onscreen (representing an estimation of both direction and range of location positioning). In the unfused display, information from the radar display from both the pilot and the confederate aircraft was presented as individual spokes onscreen (see Figure 2.1b). Each relevant intersection between corresponding spokes represented an estimation of both direction and range of location position(s). Consistent with the radar display, these spokes were available onscreen for two seconds, with an eight second interval between each display (during which no information was available onscreen). These dynamic onscreen location estimates were updated throughout each trial and were dependent upon the relationship between the pilot and confederate aircraft and the location(s). Participants were informed that the onscreen estimates would never be

100% accurate, would move around, and that they could use this information to guide their location estimations.

2.3. Experiment 2.1

The Information Fusion Testbed (IFT) located at QinetiQ, Farnborough, was used in Experiment 2.1 to test the hypothesis that, when compared to a more traditional unfused display, a fused display would support operator performance during flight missions, yet lead to impoverished encoding and subsequent retention of location information. Specifically, it was predicted that the use of fusion to provide operators with permanently available onscreen information (in the form of small squares) would lead to over-reliance upon the external display, and a lack of transfer-appropriate processing. In contrast, the semi-permanent nature of information provided within the unfused display (in the form of spoke intersections) was predicted to promote the use of memory and inference making during flight missions, thus improving the encoding and retention of location information.

2.3.1. Method

Participants Six male pilots between 30 and 50 years of age participated in the study, each with a minimum of ten years flight experience.

Materials The IFT simulated a high fidelity future jet cockpit including Head-Up Display, aircraft controls and interactive touchscreen display. Contained within the touchscreen display was the radar and terrain display. Information provided within the radar display became active when the participant's aircraft flew within 55 nm of the location(s). The nature of information provided within the terrain display depended upon whether fusion was present or not.

Design The representation and availability of information provided within the terrain display was manipulated within-subjects. The fusion algorithm continuously integrated information from the two aircraft, whilst taking into account previous estimations in order to provide permanent onscreen location information. The Fused display became active when both aircraft flew within 55 nm of the location(s), and location estimates were in the form of small squares (see Figure 2.1). In the UnFused condition, however, only the sharing of semi-permanent radar information between the two platforms was possible. As with the radar display, UnFused spokes became active the moment either of the respective aircraft flew within 55 nm of the location(s), and location estimates were determined by the relevant intersections between corresponding spokes. Consistent with the radar display, these spokes were displayed for two seconds with an eight second interval between each display. The number of locations (1, 2, or 3) was also manipulated within-subjects to create three workload conditions. On occasions where more than one location was to be identified, each of the locations was situated within 10 nm of one another. The starting location of both aircraft and the actual position of each location varied systematically across trials. Each participant received two trials from each treatment combination in a different randomised order.

Procedure Prior to commencement of the twelve experimental trials, two practice trials were completed to familiarise participants with the task and display formats. Each trial consisted of one flight mission, retention interval and recall test. The route flown during each mission was under the control of participants at all times, and because performance was likely to vary over the duration of flight, measures were taken at several points to observe both the development of location accuracy over time and any interaction with fusion. At distances of 50, 40, 35, 30 and 20 nm from the location, the prompt 'RESPOND NOW' appeared onscreen at which point participants were required to touch the screen to record each location estimate. Onscreen location information was always visible at prompt times in the Fused condition, but due to the semi-permanent nature of the radar information, was only visible during 20% of prompts in the UnFused condition. Each flight mission terminated once a response had been made to the final prompt, and recall of location(s) was measured five minutes following the completion of each flight mission. A paper map was given to participants to mark their location estimates. During the five minutes preceding this memory test, participants were required to complete a subjective assessment of their performance on the preceding flight. (See Appendix G for task instructions.)

2.3.2. Results

A number of rules were implemented during the analysis of these data. Firstly, if the number of responses made by the participant ever exceeded the number of responses required, the response(s) nearest the actual location(s) was taken. Secondly, any 'no-responses' resulted in missing data points. Two methods were used to measure participants' location estimates. Measure 1 matched participants' location estimate(s) to the actual location(s) in such a way so as to minimise the total mean distance error. Measure 2 matched the centre of participants' location estimate(s) to the centre of the actual location(s). Minimal differences were observed as a function of these two measures, and thus, only results obtained via Measure 1 will be reported. Upon violations of sphericity, Greenhouse-Geisser corrected degrees of freedom are reported, and non-transformed data are presented in tabular and graphical format throughout Chapter 2. The effects of fusion and workload on location accuracy during flight missions will be reported first, followed by the effects of fusion on the retention of location information.

Location accuracy For prompts 1 and 2, over 50% of locations were not correctly identified. Thus, to avoid empty cells only data from prompts 3, 4 and 5 were analysed. A log transformation was performed on the data in order to correct for differences in variance between the fusion conditions, and a 3 (Prompt 3/4/5) x 2 (Fused/UnFused) x 3 (Locations 1/2/3) within-subjects ANOVA was computed on the transformed data. The Fused condition yielded significantly more accurate estimations than the UnFused condition, $F\ (1, 5) = 47.36, p < .001, MSE = 0.04$, and estimation accuracy was affected by the number of locations, $F\ (2, 10) = 12.20, p < .01, MSE = 0.02$. Furthermore, a significant fusion x locations interaction was observed, $F\ (2, 10) = 11.60, p < .001, MSE = 0.04$, with simple main effects revealing an advantage for Fused over UnFused when there were two or three locations (*ps* <.05), but not when there was one, $F\ (1, 5) = 0.99, p > .05$ (see Table 2.1).

Table 2.1

The Effect of Fusion and Workload on Location Accuracy (Experiment 2.1).

	Fused		UnFused	
Locations	Mean Error	SD	Mean Error	SD
1	2.48	2.85	2.57	0.80
2	1.33	0.60	6.95	3.21
3	3.12	2.27	4.99	2.02

Note. Values are given in nautical miles. Mean Error is the difference between the actual and estimated location.

No main effect of prompt was observed, *F* (2, 10) = 0.60, *p* >.05, *MSE* = 0.03, and participants in each of the fusion conditions were equally accurate in identifying the correct number of locations on each trial, as indicated by identical proportional means (Fused: *Mean* = 0.83, *SD* = 0.21; UnFused: *Mean* = 0.83, *SD* = 0.24).

Location retention Delayed location recall was compared against the immediate location estimations provided at the final prompt in order to examine the retention of location information following the retention interval. A reciprocal transformation was performed on the data in order to correct for differences in variance between the fusion conditions, and a 2 (Fused/UnFused) x 2 (Immediate/Delayed) x 3 (Locations 1/2/3) within-subjects ANOVA was computed on the transformed data. Overall, the Fused condition yielded significantly more accurate estimates than the UnFused condition, *F* (1, 5) = 20.08, *p* <.01, *MSE* = 0.05, and accuracy deteriorated during the retention interval, *F* (1, 5) = 6.55, *p* <.05, *MSE* = 0.10. Importantly, a retention interval x fusion interaction was found (see Figure 2.2), *F* (1, 5) = 7.76, *p* <.05, *MSE* = 0.07, and simple main effects indicated that only at the final prompt (immediate) were estimations more accurate in the Fused, compared to the UnFused condition, *F* (1, 5) = 20.30, *p* <.01. No significant difference was observed between Fused and UnFused at delayed recall, *F* (1, 5) = 1.22, *p* >.05. Simple main effects also indicated that significant deterioration in accuracy occurred in the Fused, *F* (1, 5) = 7.42, *p* <.05, but not in the UnFused condition, *F* (1, 5) = 0.31, *p* >.05. A separate one way ANOVA computed on proportional data found no significant differences between the two fusion conditions in terms of recalling the correct number of locations for each trial, *F* (1, 5) = 0.17, *p* >.05, *MSE* = 0.01.

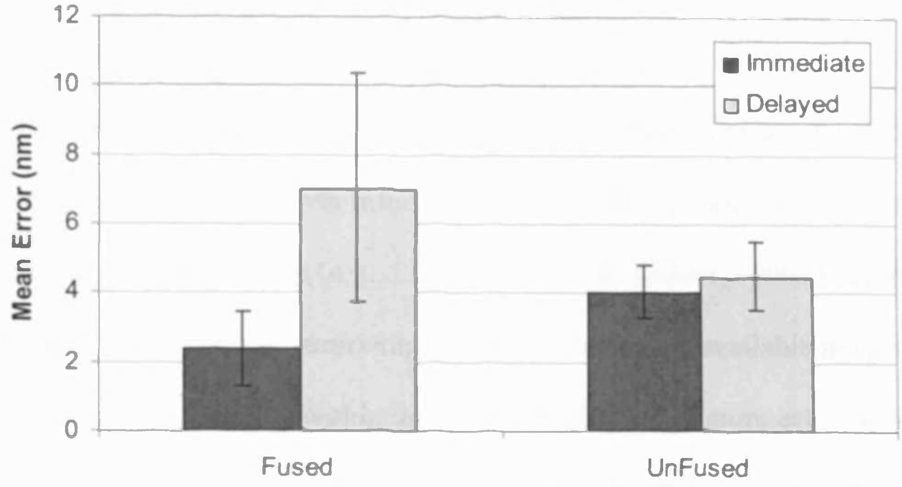


Figure 2.2: The effect of fusion and retention interval on location accuracy (Experiment 2.1).

Note. Mean Error is the difference between actual and estimated location. Error bars represent ±1 standard error.

2.3.3. Discussion

The two main predictions were supported. Firstly, during flight missions participants were more accurate in identifying location position(s) when fusion was present, compared to when it was not. Secondly, deterioration of accuracy following the retention interval was greater in the Fused, compared to the UnFused condition. Interpretation of the data is based upon the premise that more cognitive processing was required to extract information from the UnFused, compared to the Fused display. It is proposed that differences in the cognitive processing associated with extracting information from the two displays explains both the superiority of the Fused display at supporting the accurate estimate of locations during flight, and the impoverished retention of location information derived from the Fused display. It is argued that differences in the manner by which participants chose to complete the task, rather than design constraints, explain the differences in memory recall observed between the Fused and Unfused conditions.

Given the growing body of knowledge suggesting that making information more accessible within the interface via information fusion can facilitate information extraction and decision making (e.g., Lintern, Waite, & Talleur, 1999; Vicente et al., 1996; Woods, 1991), it is not surprising that the permanently available integrated location information provided within the Fused display led to more accurate location estimation than the provision of semi-permanent non-integrated location information provided within the UnFused display. With increasing workload, the Fused display resulted in superior location accuracy during flight, presumably reflecting the competing demands for limited resources, a finding consistent with the workload literature (Wickens, 2002). It is somewhat counterintuitive, however, that by increasing the availability of information provided within the interface, the Fused display also yielded a disproportionate rate of forgetting, when compared to the UnFused display. There are a number of potential explanations for this.

Firstly, the permanent nature of onscreen information provided within the Fused display may have increased pilots' reliance upon the interface as an external memory source (Gray & Fu, 2004; Gray et al., 2006; O'Regan, 1992), and in doing so, rendered internal processing involving memory and inference making redundant. Similar arguments can be found within the training literature whereby the provision of concurrent visual feedback within an external display can act as a temporary crutch to performance, and subsequently lead to a decrement in retention of skill over time (e.g., Patrick & Mutlusoy, 1982; Schmidt & Wulf, 1997). Continuous feedback within a learning environment is often found to be effective during the learning phase because it guides the individual towards the required responses and reduces errors.

However, many studies have also found that performance gains during practice are seldom observed at transfer tests when augmented feedback is withdrawn (see Stammers & Patrick, 1975). Although the onscreen information provided within the Fused display provided pilots with guidance information rather than feedback relating to performance, similar mechanisms are proposed to account for the problematic retention of Fused information in the current study. Over-reliance upon display support may be detrimental to retention.

Secondly, the semi-permanent nature of location information provided within the UnFused display may have encouraged the use of internal memory and inference-making strategies in order to maintain an understanding of location position(s) during times when location information was not available onscreen. As previously emphasised, the use of internal processes such as memory and inference making during task performance are integral to the effective retention of task-relevant information (Anderson & Milson, 1989; McNamara et al., 1996; Schmidt & Bjork, 1992).

A further issue concerns the additional cognitive processing necessary to integrate and distinguish between meaningful and coincidental spoke intersections provided within the UnFused display. The task of integrating and distinguishing between meaningful and coincidental intersections will necessarily have oriented pilot behaviour towards inference making, which may have acted as transfer-appropriate processing (McNamara et al., 1996; Morris et al., 1977) when considering the demands of the recall task. Whilst this might contribute to the superiority of the UnFused display at retention, this may also have handicapped estimates of location during the flight.

The limitation of Experiment 2.1 is primarily that it is an applied study and as a consequence there are potentially confounding differences between the Fused and UnFused conditions, some of which have already been discussed in the context of different explanations for the results. There is a lack of representational equivalence (Larkin & Simon, 1987) between the conditions, although from an applied perspective this is to some extent inevitable because a fused display for obvious reasons is never likely to adopt the physical characteristics of a radar display. In addition, there are small differences in the nature of the algorithms underlying the provision of information between the two displays. Therefore, the overall goal of Experiment 2.2 was not only to remove these algorithmic differences between the Fused and UnFused displays, but also to attempt to replicate the main results of Experiment 2.1 under laboratory conditions using a low-fidelity simulation of the IFT. Experiment 2.3 investigated the importance of the temporal availability of displayed information, which has been discussed above as a factor that may affect cognitive processing. Given that the permanence of information in the fused display may have reduced cognitive processing and thus recall, Experiment 2.3 attempted to mitigate this effect by varying the temporal availability of fused information.

2.4. Experiment 2.2

Small, but significant discrepancies between the Fused and UnFused displays in Experiment 2.1 may have led to subtle differences between the conditions with regard to the accuracy of onscreen information at different points in time. For example, the algorithm used to produce the onscreen location information provided within the Fused display in Experiment 2.1 integrated previous information regarding the positioning of a particular location when updating information pertaining to that location. The onscreen estimates provided within the UnFused display, however, relied completely upon radar information, and consequently did not take into account previous location information when updating position estimates. There will also have been slight inherent delays in the onset of onscreen location information in the Fused condition. This was because the Fused algorithm required both the pilot and confederate aircraft to be within 55 nm of the location(s) before onscreen estimates could be made available within the interface. In contrast, as soon as either aircraft in the UnFused condition flew within 55 nm of the location(s), information in the form of a single spoke was emitted from that aircraft (regardless of the status of the other aircraft).

Although the slight time delay and integration of location history information was inherent to the fusion technology under investigation in Experiment 2.1, it is important that both are removed in order to ensure the effects observed were attributable to the cognitive processing required to extract information from the interface, rather than differences between the fusion conditions in terms of accuracy of onscreen location information and the point at which such information became active. Therefore, Experiment 2.2 was conducted under laboratory conditions using a low-fidelity simulation of the IFT task (IFTsim).

2.4.1. Method

Participants 42 Cardiff University students took part in the study. Each was paid £5 or received course credit for their participation.

Materials IFTsim was written in *Visual Basic 6.0* and was presented to participants via a 12 x 13 inch high-resolution monitor and Pentium IV 2 Ghz PC. Both aircraft were set to travel at a constant speed (simulating the average speed in the IFT), and the participant's aircraft was guided via left and right arrow keys on the keyboard. On each occasion an arrow key was pressed, the participant's aircraft would change its current heading by 22.5° in the corresponding direction. The confederate aircraft was always positioned 10 nm west of the participant's aircraft. The starting location of both aircraft and the position of the fixed location(s) were varied systematically across trials.

The same algorithm was used to produce the onscreen location information in both fusion conditions. The nature of the algorithm meant that onscreen location information became more accurate as the aircraft grew closer to the actual location(s), and became active the moment the participant's aircraft flew within 55 nm of the actual location(s). Each onscreen estimate was updated in a unique fashion, four times per flight mission at distances of 50, 40, 30, and 20 nm from the actual location(s). Participants were prompted in the same manner as Experiment 2.1 to make their responses, although responses were made via mouse clicks on the screen rather than touchscreen responses, and only four were required per flight mission.

Design and procedure These were the same as Experiment 2.1, with the exception that the number of locations was either one or three. At the recall stage of each trial, the map was presented on the computer screen and responses were made via mouse clicks. Each participant received one of four randomised trial orders so as to minimise the contribution of any idiosyncratic order effects. (See Appendix H for task instructions.)

2.4.2. Results

Location accuracy A 4 (Prompt 1/2/3/4) x 2 (Fused/UnFused) x 2 (Locations 1/3) within-subjects ANOVA was computed upon the non-transformed data. As in

Experiment 2.1, participants in the current experiment revealed more accurate location estimations when working with the Fused, compared to the UnFused display, $F(1, 41) = 5.446, p < .05, MSE = 6.08$, and when there was one, compared to three locations to-be-identified, $F(1, 41) = 41.02, p < .001, MSE = 4.43$. Simple main effects examining a fusion x prompt interaction, $F(3, 123) = 6.10, p < .001, MSE = 17.92$, revealed that the only benefit of Fused over UnFused occurred at the final prompt, $F(1, 41) = 35.92, p < .001$ (Fused *Mean Error* = 2.79, *SD* = 0.94; UnFused *Mean Error* = 4.60, *SD* = 2.61). Simple main effects were also used to explore a fusion x location interaction, $F(1, 41) = 22.10, p < .001, MSE = 7.43$, and indicated that estimates in the UnFused condition became less accurate as the number of locations increased, $F(1, 41) = 40.26, p < .001$, whereas the Fused condition was relatively unaffected by the number of locations to-be-identified, $F(1, 41) = 2.70, p > .05$, (see Table 2.2).

Table 2.2

The Effect of Fusion and Workload on Location Accuracy (Experiment 2.2).

	Fused		UnFused	
Locations	Mean Error	SD	Mean Error	SD
1	5.82	3.61	5.27	3.30
3	5.87	2.93	7.31	3.33

Note. Values are given in nautical miles. Mean Error is the difference between the actual and estimated location.

Again, no significant differences were found between the two fusion conditions in terms of identifying the correct number of locations on each trial, $F(1, 41) = 2.29, p > .05, MSE = 0.05$.

Location retention A 2 (Fused/UnFused) x 2 (Immediate/Delayed) x 2 (Locations 1/3) within-subjects ANOVA compared accuracy of delayed recall with immediate accuracy at the final prompt for both fusion conditions. A log transformation was required to correct for differences in variance between the fusion conditions. As with Experiment 2.1, the Fused display was found to yield more accurate estimations than the UnFused, F (1, 41) = 25.06, p <.001, MSE = 0.02, and accuracy deteriorated during the retention interval, F (1, 41) = 56.48, p <.001, MSE = 0.05. Importantly, a retention time x fusion interaction was found, F (1, 41) = 11.80, p <.001, MSE = 0.03, (see Figure 2.3), and again simple main effects indicated that estimations derived from the Fused display were superior when immediate estimations were taken at the final prompt, F (1, 41) = 39.26, p <.001, but that there was no difference between the two fusion conditions at delayed recall, F (1, 41) = 0.03, p >.05. In contrast to Experiment 2.1, however, a decrement to location accuracy was witnessed as a function of retention interval for both the Fused and UnFused conditions (ps <.01).

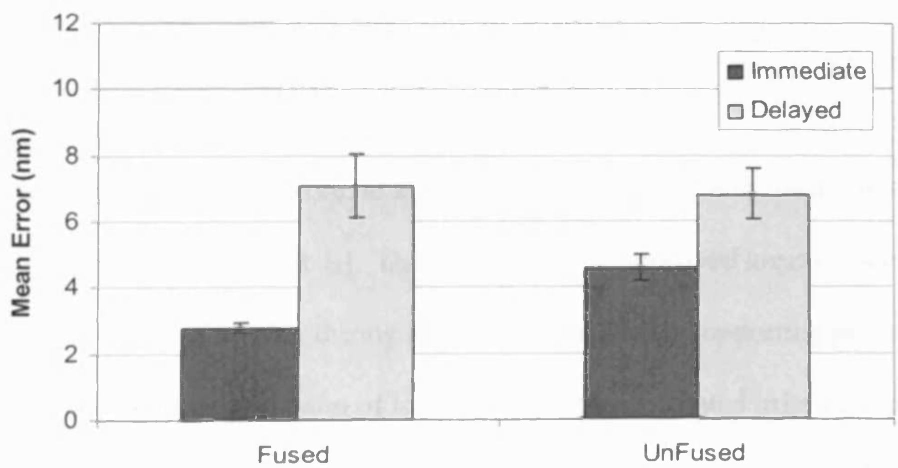


Figure 2.3: The effect of fusion and retention interval on location accuracy (Experiment 2.2).

Note. Mean Error is the difference between actual and estimated location. Error bars represent ±1 standard error.

Similar values observed at recall for the two fusion conditions may imply that a floor effect could be distorting the data. However, further examination in accordance with criterion proposed by Cohen (1995) suggested this was not the case. The average recall values for both the Fused and UnFused conditions were sufficiently different to the average maximum recall error, indicating that performance was not at a floor. In addition, the considerable difference in the simple effect F values for the effect of retention interval on location accuracy was over twice that for the Fused condition, compared to the UnFused condition, suggesting a greater deterioration over the retention interval in the Fused condition. This interpretation is confirmed by an analysis of effect size that indicated a substantially larger effect in the Fused, compared to the UnFused condition (*partial eta squared*, 0.54 versus 0.34 respectively).

Again, no significant differences were observed between the two fusion conditions in terms of recalling the correct number of locations on each trial $F(1, 41) = 1.89$, $p > .05$, $MSE = 0.07$.

2.4.3. Discussion

Manipulation of fusion and workload affected task performance in much the same way as observed in Experiment 2.1. The Fused display improved location accuracy relative to the UnFused display during flight missions, again supporting previous work suggesting that the provision of highly accessible integrated information can support operator decision making (e.g., Lintern et al., 1999; Vicente et al., 1996; Woods, 1991). However, unlike Experiment 2.1 where there was no decrement in UnFused accuracy over the retention interval, in this experiment both Fused and UnFused information deteriorated. Nevertheless, there appeared to be a greater decrement in the Fused condition, as indicated by the interaction effect and the difference in effect sizes. Location accuracy also reduced as the number of locations to-be-identified increased from one to three.

Naive participants were used in Experiment 2.2, in contrast to the experienced pilots in Experiment 2.1. Although previous work (e.g., Mosier, Skitka, Burdick, & Heers, 1996) has found student and experienced pilot samples to be equally susceptible to automation bias (over-reliance upon automated information), participants in the current experiment found the estimation of three locations particularly difficult when working with an UnFused display. In contrast, the effect of workload on location accuracy during Experiment 2.1 was approximately equivalent across fusion conditions, and may reflect pilot expertise associated with distinguishing between meaningful and coincidental UnFused intersections.

Given the superiority of the Fused display at supporting the accurate estimation of locations during flight missions, it is striking that this did not translate to better memory for location positioning following the retention interval. Indeed, the disproportionate rate of forgetting observed in the Fused condition suggests that there is scope for improving the retention of fused information by orienting participant interaction with the Fused display towards the use of internal processes (such as memory and inference making). From an applied perspective the important goal is to find a means of mitigating this effect and supporting the retention of fused information. Interpretations of Experiments 2.1 and 2.2 have suggested that the decrement in Fused recall may be due to a lack of cognitive processing via an over-reliance upon the display as an external memory source (O'Regan, 1992). In contrast, the semi-permanence of information in the UnFused display may have resulted in memory-intensive processing. This perspective is consistent with Gray and colleague's (Fu & Gray, 2000; Gray & Fu, 2004) distinction between external display-based versus internal memory-based interaction strategies, and their assertion that such shifts in strategy are based upon consideration of the temporal availability of information provided within an external display (Gray et al., 2006). Therefore, paradoxically, Experiment 2.3 investigated a method for improving the retention of fused information that involves reducing its availability within the interface.

2.5. Experiment 2.3

Reducing the temporal availability of location information provided within the interface was explored in the final experiment as a possible method for improving memory for Fused information in the current task. There are two possible explanations why this might be effective. Firstly, as discussed previously, when information is less available within an interface, stronger reliance upon internal processing is induced, including memory and inference making (Duggan & Payne, 2001; Gray & Fu, 2004; Gray et al., 2006; McNamara et al., 1996; O'Regan, 1992). Secondly, the on/off-set of visually presented information may lead to attentional capture, as seen in visual monitoring tasks (Sutcliffe, 1995; Yantis, 1993; Yantis & Jonides, 1996). The former explanation predicts that the *duration* with which fused information is made unavailable will determine its retention, as opposed to the latter explanation, that predicts improvements will be a function of on/off-set *frequency*.

In order to evaluate these competing interpretations of any effect of reducing the availability of fused information, four conditions were employed in Experiment 2.3. Three conditions in which Fused information was provided in a semi-permanent fashion within every ten second cycle (on2off8, on8off2, on1off4) were compared to a Fused condition in which information was permanently provided within the interface (on10). If reducing the availability of Fused information does have the desired effect, these conditions allow the relative influence of the duration and frequency with which onscreen information was provided within the Fused display to be investigated.

2.5.1. Method

Participants Eighty Cardiff University students took part in the study. Each was paid £5 or received course credit for their participation. One participant in the on2off8 condition was excluded due to obtaining standard deviations in excess of ± 3 from the mean.

Design The temporal availability of onscreen location information was manipulated between-subjects in order to remove any possible contamination due to carry-over effects (Poulton, 1982), and workload was again manipulated within-subjects. In order to improve comparability between location estimates provided within the different Fused conditions, onscreen location information was also removed during prompt times regardless of its temporal availability. Table 2.3 provides a schematic representation of the temporal availability of information provided within the four Fused conditions. The 'on2off8' Fused condition replicated the temporal availability of location information provided within the UnFused display (Experiments 2.1 and 2.2). The 'on8off2' condition reversed the temporal cycle of the on2off8 condition. In doing so, the duration for which location information was unavailable within the display was reduced to two seconds, yet the frequency with which Fused information flashed on/off within each ten second cycle was held constant. The extent to which attentional factors affected performance within the current task were evaluated with respect to the 'on1off4' condition, during which location information was available for a total of two seconds (on two separate one second episodes) for every ten second cycle. The on1off4 condition thus provided participants with onscreen information for the same duration as the on2off8 condition, but reflected an on-/off-set frequency ratio of 2:1.

Table 2.3

A Schematic Representation of the Temporal Availability of Fused Information (Experiment 2.3).

Condition	On-/off-set period									
on10	On									
on2off8	on		Off							
on8off2	off		On							
on1off4	on	off				on	off			
	1	2	3	4	5	6	7	8	9	10
	→ Time (seconds) →									

These four conditions allow for selected critical comparisons to be made on the recall data. Firstly, each of the semi-permanent Fused conditions (on2off8, on8off2, on1off4) can be compared to the 'on10' condition, in which Fused information was permanently available onscreen, in order to assess the extent to which reducing the availability of Fused information improves retention. Secondly, if the relative importance of duration takes precedence, one would expect the on2off8 and on1off4 conditions (both of which have a total off duration of 8 seconds in every 10 second cycle) to yield better retention of location information than the on8off2 condition (which has a total off duration of 2 seconds in every 10 second cycle). Finally, if the relative importance of frequency outweighs that of duration, one would expect the on1off4 condition (with an on-/off-set frequency of two per 10 second cycle) to yield better retention of location information than both the on2off8 and on8off2 conditions (each with an on-/off-set frequency of one per 10 second cycle).

Materials and procedure The materials and procedure were identical to those employed during Experiment 2.2, with the exception that participants experienced six experimental trials, rather than twelve (three with one location, three with three locations). In addition, the retention interval was reduced to 1.5 minutes in order to verify the resilience of the effect with a shorter retention time that may be important in operational contexts. (See Appendix I for task instructions.)

2.5.2. Results

Location accuracy A 4 (Prompt 1/2/3/4) x 2 (Locations 1/3) x 4 (Temporal Availability of Fused Information) ANOVA was computed with the first two factors manipulated within-subjects and the final factor manipulated between-subjects. A main effect was found for the availability of information, F (3, 75) = 3.30, p <.05, MSE = 9.01 (see Table 2.4), with Bonferroni-corrected post hoc analyses indicating that only the on8off2 condition significantly improved location accuracy relative to the on10 condition (p <.05).

Table 2.4

The Effect of Information Availability on Location Accuracy (Experiment 2.3).

Information Availability	Mean Error	SD
on10	6.16	4.04
on2off8	5.56	3.64
on8off2	5.17	3.30
on1off4	5.34	3.97

Note. Values are given in nautical miles. Mean Error is the difference between the actual and estimated location.

The presence of a main effect of prompt, F (1.76, 131.99) = 409.47, p <.001, MSE = 6.44, indicated that participants' location estimations became more accurate over time. Although no main effect of locations was found, F (1, 75) = 0.03, p >.05, MSE = 4.92, simple main effects examining the locations x prompt interaction, F (1.91, 143.28) = 133.73, p <.001, MSE = 10.15, indicated that accuracy decreased as the number of locations increased at prompts two, three and four (*ps* <.01), but increased at prompt one (p <.001).

Location retention A 2 (Immediate/Delayed) x 2 (Locations 1/3) x 4 (Temporal Availability of Fused Information) ANOVA was computed upon the immediate and delayed recall data with the first two factors manipulated within-subjects and the final factor between-subjects. A reciprocal transformation was required to correct for differences in variance between the two fusion conditions. Accuracy of estimates decreased during the retention interval, F (1, 75) = 84.21, p <.001, MSE = 0.10, and was affected by the temporal availability of onscreen fused information, F (3, 75) = 5.76, p <.001 MSE = 0.14. Again, a retention time x availability of information interaction was found, F (3, 75) = 3.36, p <.05, MSE = 0.10, (see Figure 2.4).

Although simple main effects pointed to an effect of availability of information at the final prompt (immediate), F (3, 75) = 5.40, p <.01, but not at delayed recall, F (3, 75) = 1.76, p >.05, planned comparisons revealed that participants in the on2off8 and on1off4 conditions exhibited significantly more accurate delayed recollection when compared to the on10 condition (*ps* <.05). No differences were observed in recall between the on8off2 versus on10, the on8off2 versus on2off8, or the on1off4 versus on2off8 comparisons (*ps* >.05).

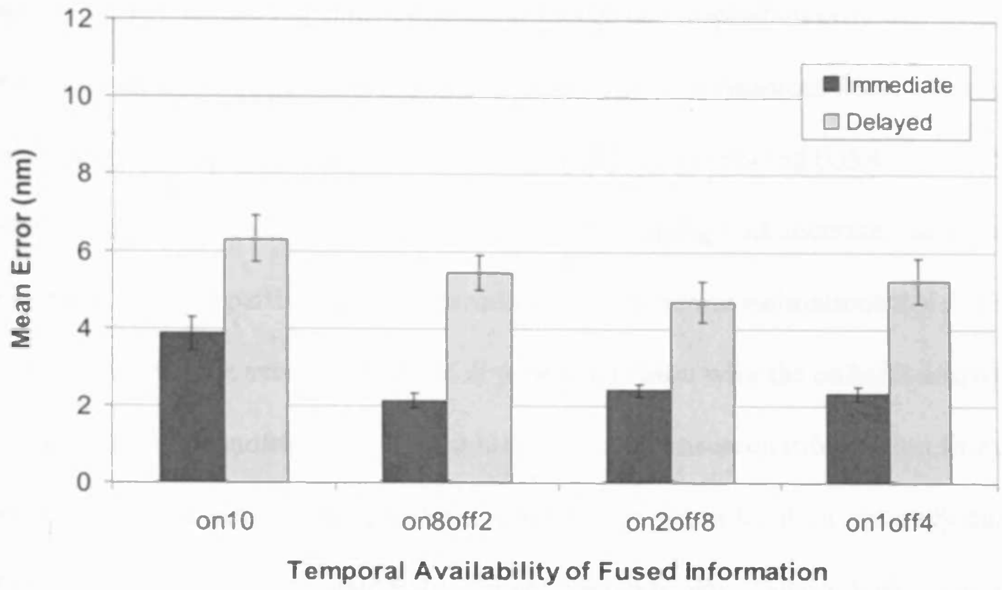


Figure 2.4: The effect of availability of onscreen information and retention interval on location accuracy (Experiment 2.3).

Note. Mean Error is the difference between actual and estimated location. Error bars represent ±1 standard error.

2.5.3. Discussion

Reducing the availability of onscreen information within the Fused display not only led to increased memorability of location information, but also improved participants' location accuracy during flight missions. Specifically, the on8off2 semi-permanent condition improved location accuracy during flight, and the on2off8 and on1off4 semi-permanent conditions improved memory after flight, relative to the permanent Fused condition.

Although not predicted, it is perhaps not surprising that the on8off2 condition improved location accuracy during flight missions when compared to the on10 condition. Mosier, Skitka, Heers, and Burdick (1997) pointed out that pilots tend to use automated cues as a heuristic replacement for information seeking, and tend to rely upon these cues despite conflict between expected and actual automation performance. Perhaps the semi-permanent nature of information provided within the on8off2 Fused condition minimised over-reliance on the onscreen location estimates (never 100% accurate), and promoted internal inference making (Glover, 1989; McNamara et al., 1996). Provided that inference making was accurate, such a strategy may have allowed participants to formulate more accurate estimations than those provided within the external display. A possible reason why the on2off8 and on1off4 semi-permanent conditions (both of which removed onscreen information from eight seconds of every ten) did not also lead to improvement in location accuracy during flight may be due to the considerably larger demands placed upon working memory (Baddeley, 1986). Importantly, however, no decrement in performance during flight missions was witnessed in these conditions relative to the on10 condition, in which Fused information was permanently provided within the external display.

Support was found for the prediction that reducing the temporal availability of information provided within the external display would improve the encoding of fused information. Location recall was found to be superior in both of the semi-permanent Fused conditions that removed onscreen information for eight seconds of every ten second cycle (on2off8, on1off4), when compared to the Fused condition where onscreen information was permanently available within the interface (on10). However, an exception was that when information was removed for only two seconds of every ten second cycle (on8off2), recall measures did not differ to when information was continually available throughout each ten second cycle. This suggests that a threshold value for the unavailability of Fused information may exist in order to improve retention, even though there is not a significant difference between the on2off8 and on8off2 conditions.

The next issue concerns the possible explanation for improved recall for semi-permanent Fused information. The fact that the on1off4 condition did not improve retention relative to the on2off8 condition (despite a doubled on-/off-set frequency, with duration held constant), supports the proposition that attentional capture (Sutcliffe, 1995; Yantis, 1993; Yantis & Jonides, 1996), at least within the on/off-set frequencies examined here, is unimportant. However, the improved recall observed in the on2off8 and on1off4 conditions (both of which remove information for eight seconds of every ten second cycle), relative to the on10 condition, strongly suggests that it is the duration with which Fused information is made temporarily unavailable that is responsible for improved recall. Hence it is the process of inference making (McNamara et al., 1996; Palmiter et al., 1991), and the use of memory when information was unavailable during flight missions (Fu & Gray, 2000; Gray & Fu, 2004; Gray et al., 2006) that improved the encoding of fused information in the current task (Anderson & Milson, 1989; Anderson & Schooler, 1991; Duggan & Payne, 2001; Schmidt & Bjork, 1992).

Again, it is emphasised that the improved encoding of Fused information observed as a function of reducing the temporal availability of information provided within the external display did not lead to an associated decrement in location accuracy during flight performance. In addition, the lack of an interaction between location and temporal availability of Fused information during flight performance suggests that each of the Fused conditions was affected in much the same way by changes in workload.

2.6. General discussion

Evidence is provided supporting the concern outlined in the introduction that, under certain circumstances, the provision of highly accessible information within a Fused interface may lead to impoverished encoding and problematic retention of visual-spatial information. Experiments 2.1 and 2.2 indicated that the provision of permanently available integrated information in the Fused display improved location accuracy during flight missions, supporting much work conducted within the field of cognitive engineering (Bennett & Flach, 1992; Lintern et al., 1999; Vicente, 2002; Wickens & Carswell, 1995; Woods, 1991). However, the retention of location information derived from the Fused environment deteriorated disproportionately when compared to the UnFused condition (in which onscreen information was not integrated and only provided for two seconds of every ten second cycle). Experiment 2.3 demonstrated that memory for fused information within the current task could be improved by reducing operator reliance upon the interface as an external memory source (Gray et al., 2006; O'Regan, 1992), and encouraging transfer-appropriate processing (Morris et al., 1979), such as inference making (McNamara et al., 1996; Palmiter et al., 1991) and the use of internal memory (Anderson & Milson, 1989).

Essentially, Chapter 2 demonstrates that a sensitive balance exists between reducing an operator's cognitive workload by making the required information more accessible (Kirsh, 2000), and ensuring that extraction of information is not made so effortless that information is promptly forgotten. The deployment of internal memory during information seeking reduces as information becomes more accessible within the interface (Ballard et al., 1995; Fu & Gray, 2000; Gray & Fu, 2004; Gray et al., 2006), which can have negative consequences for the activation of memory traces in the future (Anderson, Fincham, & Douglass, 1999; Schmidt & Bjork, 1992). The current series of experiments provides a topical, perhaps counterintuitive example, whereby actually *reducing* the availability of Fused information provided within the external display improves subsequent performance. It is not suggested that the Fused display imposed constraints meaning that participants would never be able to remember visually presented information, but rather, the Fused display encouraged reliance upon the interface as an external memory source, and thus did not encourage the internal processing required for effective retention of information. It should be noted that the superior retention of information observed in the UnFused condition occurred within an incidental, as opposed to intentional learning task (Postman, 1964). It remains an empirical question whether it would also be found with intentional learning, but this is perhaps unlikely.

2.6.1. Task constraints

Task performance criteria will dictate whether memory encoding is beneficial. Some may argue that efficacious encoding and retention of information provided within task environments such as the one chosen for the current set of studies is not necessary. Such displays are designed so as to allow the operator instant access to the information needed, thus memory often becomes redundant. However, humans will always be required to monitor, supervise, adjust, and maintain augmented displays (see Bainbridge, 1987) which on occasion will fail. Indeed, Reising and Sanderson (2004) have highlighted the consequences of instrument failure within a display influenced by Ecological Interface Design and stated that "the more an arrangement of parts adds information beyond that in the parts alone – the more devastating the impact of a faulty sensor might be" (p. 317).

It is anticipated that display manipulations designed to promote the development of a robust internal representation of the information presented externally will be easier in some situations than others. For example, memory for locations assessed in the current set of studies will have had semantic and relational properties with regard to landmarks situated within the terrain map (of which there were many). The mental organisation of such information has previously been shown to be influenced by the manner in which a map is learnt (Curiel & Radvansky, 1998), and long-term working memory for visual representations of natural scenes is surprisingly robust and long-standing (Hollingworth, 2005). Whether reducing operator reliance upon an interface representing detailed physical information would have similar effects to those observed in Experiment 2.3 is yet to be seen, but is likely to prove more difficult (Vicente, 1992).

It is fully acknowledged that the paramount function of any external display is to provide operators with the information needed, when needed. Indeed, it would be defeatist to provide the operator with the necessary information for the majority of the task, only to remove critical information when it is required most. Therefore, much research is required in order to develop effective methods and guidelines by which active encoding of highly accessible information can be promoted, without compromising the presentation of such information in a timely fashion. Adaptive task allocation (Parasuraman, Mouloua, & Molloy, 1996) may provide a means of 'refreshing' an operator's memory for fused information, and has previously been used to improve operators' mental picture of automated processes (Parasuraman, 1993). As has been recognised in the automation literature (Parasuraman, 2000), it is argued that the use of fusion is not necessarily an 'all-or-none' concept. Instead, it is proposed that the use of fusion be adaptive to differing task demands and work in harmony with what we already know about the human information processing system (Wickens & Hollands, 2000).

2.6.2. Adaptive information fusion

Adaptively deploying fusion capabilities may well lead to improvements in human performance when compared to the static use of fusion. The use of fusion to provide operators with highly accessible information within the interface could be based upon boundary conditions and tradeoffs concerning the effectiveness of fusion under different conditions, and could be present within the interface more or less depending upon task criterion such as operator workload. Benefits of such a methodology have often been cited in the automation literature (e.g., Hilburn, Jorna, Byrne, & Parasuraman, 1997), particularly when lower-order sensory and psychomotor functions such as information acquisition have been under investigation (Kaber, Wright, Prinzel, & Clamann, 2005).

Lintern (1980) provides a good example of the benefits of adaptively providing supplementary visual cues during a simulated aircraft landing task. When compared to conditions in which visual cues were either not provided at all, or were presented continuously during training, transfer of skill was found to be superior when cues were provided adaptively within the external display. According to Lintern (1980), participants relied too heavily upon extrinsic visual cues when provided continuously, and were in need of assistance when they were not provided at all. Although training is not the focus of the current thesis, it is envisaged that similar methods based upon evaluations of operator workload (e.g., Gregoriades & Sutcliffe, 2006; Kaber et al., 2005) and sensitive to limited cognitive resources (Wickens, 2002) may improve the design of a variety of fused environments. Indeed, it is important that future research begins to assess the effect that making display information temporarily unavailable has on other cognitive resources, such as concurrent attention in multi-tasking environments. Although workload was examined to some extent in Chapter 2, multi-tasking was not (Adams, Tenney, & Pew, 1995; Monsell, 2003).

A number of further limitations have to be acknowledged with respect to the experiments reported in Chapter 2. The balance between investigation of a topic within an applied context and within controlled laboratory conditions is often problematic to effect. Experiment 2.1 had the benefit of using highly skilled pilots (albeit only a few in number) and a high-fidelity flight simulator, whereas Experiments 2.2 and 2.3 utilised naive students as participants, using a low-fidelity simulation. Whilst Experiment 2.1 presented a rich and realistic context with experienced pilots, it was not practically feasible to continue using this expensive resource in order to disentangle all the potential confounding factors in subsequent experiments. Consequently, Experiment 2.2 attempted to eliminate some of these variables, and replicate the results of Experiment 2.1. Although student participants would not have developed the domain specific encoding structures of experienced pilots, it is expected that they would have been familiar with information typically provided within standard maps. Also, the design of Experiment 2.3 was deliberately not intended to explore the possible contribution of representational equivalence (Larkin & Simon, 1987) between conditions, but rather, to pursue an important applied issue of how to improve the retention of fused information, as configured in our task environment.

The paradoxical assertion that memory for fused information may be improved by reducing the temporal availability of information provided within the interface is also somewhat limited to the scenario currently under investigation. In order to evaluate the generalisability of this finding, research is required in different settings. Indeed, it is expected that such a method will suit some scenarios yet not others. Where the scheduled removal of onscreen information is not appropriate, germane cognitive load of a different nature may be necessary in order to improve the encoding and subsequent retention of visual-spatial information (see Paas & Kester, 2006 for a review). It is anticipated, nevertheless, that the underlying principles discussed throughout Chapter 2 extend to a variety of task domains, and are worthy of careful consideration during the design of fused environments of the future.

CHAPTER 3

EMPIRICAL SERIES 2

Problem solving with information access costs in mind

3.1. Introduction

The previous chapter pointed to potential caveats associated with increasing the availability of information within the interface. As well as increasing the availability of displayed information to the operator, technologies such as information fusion also aim to reduce the cost associated with accessing task-relevant information from the interface. Therefore, the present chapter will witness a shift in emphasis from that of the *availability* of information, to that of the *cost of accessing* information.

The time, physical and mental effort associated with accessing a particular piece of information are all examples of what will hereafter be referred to as the information access cost (IAC). We experience access costs on a regular basis during interaction with computer-based environments, and IAC can fall within various heterogeneous categories. For example, when preparing a manuscript using a word processing package, one may face the task of inputting relevant data within a table. Although the location of this data may already be known, the accessibility of the data may vary considerably. On occasion, the corresponding data may be held within a minimised window, and therefore accessible at a click of a button. On other occasions, the corresponding spreadsheet may take considerable time, physical and mental effort to access. The purpose of Chapter 3 is to examine the consequences of changing the cost associated with accessing information from an interface to human problem solving behaviour.

From the perspective of display design, IAC is an important, yet under-researched issue. Facilities such as 'drill-down' (often provided within data-rich fused environments) allow the operator to access low-level data by drilling into the higher-level data provided by default within integrated/fused displays (as discussed in Chapter 1). Although now often a requirement in modern system displays, the process of drilling down takes time and effort; and as a consequence the corresponding information has a higher access cost. (For further discussion of the theoretical and practical consequences of drill-down see Duggan, Banbury, Patrick, Howes, & Waldron, 2004.)

Associated with the current Chapter's shift in emphasis away from information availability towards the cost of accessing information from the interface, will be a corresponding shift in the unit of analysis from performance measures to behavioural strategies. As identified in Chapter 1, human behaviour can be characterised as largely adaptive to one's task environment (Anderson, 1990), with soft constraints based upon cost-benefit analysis often biasing the selection between candidate microstrategies (Gray & Fu, 2004). In order to assess human behaviour with respect to changes in the accessibility of information within an interface, an interactive paradigm will be adopted within which the distribution of cognition between the individual and the task environment can be evaluated (Hutchins, 1995a; Norman, 1993).

3.1.1. Effect of IAC upon strategy selection during routine interactive behaviour

Work by Gray and colleagues (Fu & Gray, 2000; Gray, Simms, Fu, & Schoelles, 2006) has begun to examine the behavioural consequences of changes in IAC during interactive routine copying behaviour. Using the Blocks World Task (originally developed by Ballard, Hayhoe, & Pelz, 1995), a series of studies conducted by Gray et al (2006; Fu & Gray, 2000) have demonstrated that human interactive behaviour is highly sensitive to small changes in IAC, with as little as one second making an appreciable difference at the level of strategy selection. Before reviewing such work, a brief description of the Blocks World Task (BWT) is necessary.

The aim of the BWT is to recreate the pattern of blocks dictated by a Target Window in a Workspace Window (see Figure 3.1). This is achieved by clicking on and dragging blocks one at a time from a Resources Window into the Workspace Window. Throughout the Fu & Gray (2000) study, all three windows were covered by grey boxes, of which only one could be uncovered at a time. The Resources and Workspace windows could be uncovered by moving the mouse cursor into the window to-be-opened. Each window was then covered again upon removal of the mouse cursor.

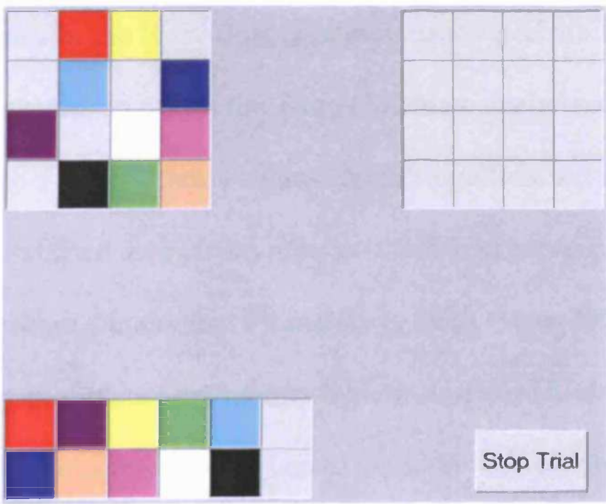


Figure 3.1: An example of a Blocks World Task start-state in the Low IAC condition. Target Window is top-left, Resources Window is bottom-left, and Workspace Window is top-right.

The cost of uncovering the Target Window varied between three IAC conditions (Low-Cost, Control Condition, High-Cost). In the Low-Cost condition, participants were required to hold down a function key. The Target Window then remained uncovered until either the function key was released or the mouse cursor entered either of the other two windows. In the Control condition, the Target Window was uncovered during times when the mouse cursor resided within the window. This was also the case in the High-Cost condition, with an additional lock-out time of one second per uncovering. Once participants were confident that the Target Pattern had been recreated in the Workspace Window, they were instructed to click the 'Stop Trial' button. If the Target and Workspace patterns matched, participants were taken to the next trial. If they did not, an error was recorded and participants were required to correct their mistake(s) before proceeding to the next trial (of which there were forty in total, each yielding a different Target Pattern).

Fu & Gray (2000) found that when compared to the Low-Cost and Control conditions, participants in the High-Cost condition uncovered the Target Window less frequently, yet spent more time with the Target Window uncovered, and copied more blocks per visit to the Target Window. These differences indicate that as IAC increased, behaviour shifted away from reliance upon interactive perceptual-motor strategies, towards reliance upon internal memory. Fu & Gray (2000) interpreted these findings within the framework of cost-benefit analysis (Anderson, 1990). When IAC was low, strategy selection was argued to avoid the use of internal memory because the relative costs of encoding and retrieval (memory is assumed to be somewhat volatile) are deemed higher than relying upon the display as an external memory source (O'Regan, 1992). As a consequence, participants in the Low-Cost condition relied heavily upon the perceptual-motor strategies required to sequentially copy one block per uncovering of the Target Window. In contrast, Fu & Gray (2000) conceptualised higher costs associated with uncovering the Target Window as prompting rather different outcomes from cost-benefit analysis. Rather than rely upon the interaction-intensive strategies required to copy one block per uncovering of the Target Window, Fu & Gray argued that participants in the High-Cost condition chose to commit to memory a number of blocks per uncovering of the Target Window, and subsequently copy each of these blocks from memory before returning to the Target Window to encode another chunk of blocks (Chase & Simon, 1973). The increased time spent uncovering the Target Window as a function of an increase in IAC was seen as a consequence of the additional time needed to encode several blocks at a time, as opposed to one. However, given the increased time cost associated with accessing the Target Window, this chunking strategy was likely used as a means of reducing overall time on task.

To be specific, Fu & Gray (2000) reported an increase in the number of blocks copied per uncovering of the Target Window from below 2 in the Low-Cost condition to 2.5 in the Control condition and nearly 3.5 in the High-Cost condition. Although this effect dissipated over the 40 trials, a fine-grained analysis of the copying strategies used by participants provided further support for increased reliance upon memory as IAC increased. The most common block copying strategy promoted by both the Low-Cost and Control conditions included a single visit to the Target Window, followed by a visit to the Resources Window, and concluded with a visit to the Workspace Window (required to place the respective block). In contrast, the most common block copying strategy promoted by the High-Cost condition did not necessitate a visit to the Target Window, implying that the majority of blocks copied when the IAC was High were copied from memory.

It is worth noting that this shift in emphasis upon the use of internal memory as access costs increased was despite an increase in the number of errors made by participants in the High-Cost condition, although no differences in time to complete the task were observed between the IAC conditions. Essentially, the results from Fu & Gray (2000) indicated that when IAC was High, the cost of uncovering the target window following each block copied was deemed higher than the cost of encoding and retrieving up to four chunks of three to four blocks per uncovering (thus reducing the number of visits to the Target Window).

Similar soft constraints upon the use of interaction-intensive and memory-intensive strategies contingent upon IAC have also been reported using a simulated VCR programming task (Gray & Fu, 2001; 2004). During this task, participants were required to set the VCR to record a number of different programs, each with several components constituting program information. As the cost associated with accessing program information increased, participants chose to access such information less frequently, and as a consequence relied more upon memory during VCR programming. In addition, the Low-Cost condition was replaced by a new condition in which all information was permanently visible, and thus eye-tracking data was required. In this instance, the only cost associated with accessing externally represented information was the saccadic eye movements needed to bring the required information to the centre of focus. As expected, such Low access costs promoted the use of highly interaction-intensive microstrategies (Gray & Fu, 2001; 2004).

Although consequences of variations in IAC have been empirically evaluated during routine interactive behaviour (Duggan & Payne, 2001; Fu & Gray, 2000; Gray & Fu, 2001; 2004; Gray et al., 2006), very little work has attempted to extend these findings to more complex human behaviour. Fu & Gray (2006) have provided the only attempt to date, investigating information seeking strategies during a map navigation task. Using a train map, participants were given start and destination stations and asked to travel from the start to the destination. At each respective intervening station (of which there were many), there were a number of different route options available. The speed of these different routes (referred to as tracks hereafter) varied, as did the cost of accessing speed information for different tracks. The relative speed of only one track at a time could be viewed, and the cost of each information-seeking action was operationalised by either a mouse-click (Low IAC), or a mouse-click plus an additional 1 second lock-out (High IAC). The utility of the information was also manipulated, but is not relevant to the current discussion.

Despite the fact that Fu & Gray's (2006) train navigation task was set up so that more information seeking behaviour would more often than not improve efficiency on the task, it was found that information-seeking behaviour reduced as IAC increased. Because substantial exploration of the relative speed of various route options was required in order to find a fast solution, premature cessation of information seeking behaviour promoted by higher access costs left major parts of the task space unexplored, and thus reduced efficiency in terms of travelling time.

The High IAC prompted participants to stabilise at a suboptimal level of information seeking. Although Fu and Gray (2006) again couched their interpretation of these results within the rational analysis framework (Anderson, 1990), they also adapted a Bayesian Satisficing Model to describe how under certain conditions a local decision rule (such as the decision to rely more upon memory when access costs are High) can lead to suboptimal global performance. Although Fu and Gray (2006) used the term 'problem solving' as a descriptor for their task, the focus was primarily upon information seeking behaviour, rather than the complex processes of problem solving.

Following from the work conducted by Gray and colleagues (e.g., Gray et al., 2006) exploring the consequences of changes in IAC during routine interactive behaviour, the primary goal of Chapter 3 was to evaluate the extent to which similar variations in IAC affect the selection between candidate microstrategies during more complex higher-level human cognition, such as problem solving. Will the shifts in strategy selection contingent upon IAC observed during routine interactive behaviour (e.g., Gray & Fu, 2004; Gray et al., 2006) extend directly to an externally supported problem solving task, or will an increase in cognitive demands manifest differential behavioural responses to changes in IAC? In order to answer such questions, it is important to begin with an understanding of problem solving within the context of an external display.

3.1.2. Externally supported adaptive problem solving

Newell and Simon (1972) historically defined problem solving as getting from an initial problem state to a desired goal state, without knowing exactly what operators are required to get there. The resources available and structure of the problem will vary from situation to situation. In order to facilitate comparison between the work of Gray and colleagues investigating the role of access costs during routine copying behaviour (e.g., Fu & Gray, 2000; Gray et al., 2006) and the current extension to problem solving, Chapter 3 will be limited to investigation of well-structured problems situated within the context of an external display. By definition, the start-state of a problem, the goal-state, the available operators and the constraints upon operator selection are known in advance during solution to well-structured problems (Ormerod, 2005).

Much problem solving is conducted within the context of an external display (Payne, 1991), and perhaps Larkin (1989) should be credited with providing the first comprehensive account of display-based problem solving. The main feature of DiBS (Display-Based Solver) emphasised that all consequences of problem solving actions are immediately reflected by a change in display, and most importantly, posited that in many tasks an external display may act as an important information storage resource. Zhang (1997) went on to discuss the nature of external representations during problem solving with respect to what he called 'representational determinism'. By this, Zhang meant the form of a representation can determine the information that can be perceived, what processes can be activated, and what structures can be discovered from the specific representation.

> "*The directly perceived information from external representations and the directly retrieved information from internal representations may elicit perceptual and cognitive biases, respectively, on the selection of actions.*" (Zhang, 1997, p. 186)

There are obvious parallels to be drawn between what Zhang referred to as representational determinism and what Gray and colleagues have referred to as soft constraints (Gray & Fu, 2004). Although perhaps the fundamental mechanism underlying each differed: Zhang spoke of a central control allocating and switching attention between internal and external representations, whereas Gray and colleagues were more specific in their assertion that the selection between internal and external microstrategies is at least in part determined by a series of least-effort tradeoffs based upon cost-benefit analysis (Anderson, 1990).

Although the problem solving literature has yet to consider the influence of information access costs, work examining the problem solving process within the constraints of an external display has demonstrated that problems solvers are sensitive to other costs situated within the task environment. For example, Knowles & Delaney (2005) have demonstrated lasting reductions in the number of illegal moves made during the River-Crossing problem following an increase in their cost. Knowles & Delaney argued that this was primarily due to an increase in cautiousness brought about by the increased cost associated with making illegal moves. Similarly, O'Hara & Payne (1998) have demonstrated that a problem solver's propensity to plan ahead during solution to the 'eight-puzzle' increases in conjunction with the implementation cost (IC): when the cost of making each move increased from pressing a function key to typing a string, the number of moves required to complete the task reduced significantly. Evidence of planfulness came mainly from increased latencies between each move. Verbal protocol analysis also suggested more planful behaviour as a function of increased IC.

In a similar vein to Gray and colleagues (e.g., Gray et al., 2006), O'Hara & Payne (1998) used cost-benefit analysis (Anderson, 1990) as a foundation for interpretation of their results, suggesting that participants chose to plan more as the IC increased in order to combat higher costs imposed upon implementing each move. The authors state that "problem solving search strategies are chosen so as optimize performance within the constraints of a particular situation" (O'Hara & Payne, 1998, p. 34). In this instance problem solving efficiency appeared of utmost importance (time and moves to solution), and increased planfulness associated with increased IC facilitated such optimisation. O'Hara & Payne (1999) have also extended these findings showing similar effects associated with an increased cost of using an 'undo' function during solution to a slide-jump puzzle and a system lockout delay following each move made during solution to both an eight-puzzle and a non-puzzle-like office administration type task.

3.1.3. Strategy selection based upon IAC during interactive problem solving?

Recent research suggests that problem solvers are often sensitive to costs situated within the task environment. Therefore, it seems important to investigate how problem solvers will react to changes in the cost associated with accessing task-relevant information from the external display. It is easy to envisage an array of applied problem solving situations within which task-relevant information is not always readily available to the operator and information may come at a variety of costs (Gronlund, Dougherty, Durso, Canning, & Mills, 2005). In such situations, would selection between microstrategies contingent upon changes in IAC highlighted during routine copying behaviour (Fu & Gray, 2000; Gray et al., 2006) map directly onto problem solving behaviour?

On the one hand, it may be the case that the least-effort tradeoffs reported during low-level routine behaviour are so entrenched in the human cognitive system that a change in task from routine copying to problem solving may make no difference to the selection between microstrategies. On the other hand, changes to the cognitive processes required to complete a problem solving task may distort the selection between microstrategies based upon IAC. For example, least-effort tradeoffs contingent upon the time costs of IAC may begin to take into consideration the deployment of other cognitive processes, such as planning (O'Hara & Payne, 1998). Higher demands upon working memory often present during problem solving may also moderate the effect of IAC upon the selection between candidate microstrategies.

Although no study to date has attempted to investigate the extent to which memory and/or planning are effected by changes in IAC during a problem solving task, a limited number of studies have suggested that the deployment of internal memory during problem solving can be seen as an adaptive response to costs situated within the task environment (e.g., Cary & Carlson, 1999; 2001). For example, Pfeiffer (2004) compared problem solving performance during solution to a series of Balls and Boxes problems under two conditions: 1) when information regarding the current-state in the problem space was permanently available, and 2) when it was only available upon request. Pfeiffer (2004) found that reliance upon memory during problem solving tended to increase as a function of a reduction in the availability of current-state information. When current-state information was not always available, participants seemed content to make up to nine moves per current-state request.

Unfortunately, however, no data were reported by Pfeiffer regarding the deployment of planning behaviour. Although it is worth noting that participants were found to be no worse at completing the Balls and Boxes puzzle when current-state information was only available upon request, compared to when current-state information was continuously available within the external display. Adopting a cost-benefit analysis perspective (Anderson, 1990), one may expect an increase in IAC to lead to an increase in planful behaviour. O'Hara & Payne (1998) argued that increased IC experienced during solution to the eight-puzzle (described previously) resulted in elevated levels of planning due to a shift in the selection between microstrategies based upon least-effort tradeoffs. Essentially, as the cost of making each move increased, participants were argued to engage in more planning in order to combat the cost (by reducing the number of moves and time required to solve the problem).

Although it seems plausible that problem solvers may also choose to engage in planful behaviour in order to combat high information access costs, the well-established relationship between memory and planning (see Gilhooly, 2004 for a review) may complicate matters somewhat. For example, Kotovsky, Hayes, & Simon (1985) found that increasing the load on working memory during solution to the Tower of Hanoi prevented even minimal planning from occurring. Similarly, Phillips, Wynn, Gilhooly, Della Salla, & Logie (1999) found reduced levels of planning as secondary tasks, designed to tax working memory, were introduced whilst participants attempted to solve versions of the Tower of London task. Phillips et al (1999) went on to highlight the importance of memory in formulating, retaining, implementing, and revising plans online. Therefore, the aim of Chapter 3 was to answer two rather related questions. Firstly, how will small changes in IAC affect the use of internal memory during an externally represented well-structured problem solving task? Secondly, how will small changes in IAC affect the deployment of planning during an externally represented well-structured problem solving task?

Following from work detailing the adaptive nature of memory contingent upon IAC during routine interactive behaviour (Fu & Gray, 2000; Gray & Fu, 2001; 2004; Gray et al., 2006) and information availability during problem solving (Pfeiffer, 2004), it is predicted that small increments in IAC will increase reliance upon memory during problem solving. However, predicting the potential interaction with, and ramifications for, planning behaviour is less straightforward. The second question will therefore be explored with no prior predictions and IC will be manipulated in the first experiment so as to provide a comparison to IAC.

3.2. Blocks Problem Solving Task

In order to simultaneously evaluate the consequences of changes in IAC upon memory and planning during problem solving behaviour, the Blocks Problem Solving Task (BPST) was developed for the purpose of Chapter 3 (see Figure 3.2). The aim of the BPST is to move the blocks in the Current-State Window so that the Goal-State is met. The rules of the task determine that only one coloured block in the Current-State Window can be moved at a time into an adjacent empty (white) space.

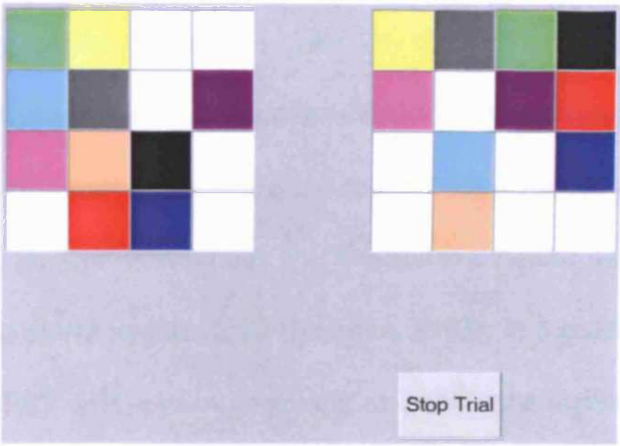


Figure 3.2: An example of a Blocks Problem Solving Task start-state in the Low IAC condition. Goal-State Window is left, Current-State Window is right.

The task is conceptually similar to the BWT (used by Fu & Gray, 2000; Gray et al., 2006), with the exception that blocks are not dragged from the Resources Window into the Workspace Window in order to meet the Target Pattern. But rather, are moved one at a time within the Current-State in order to meet the Goal-State. Emphasis is thus shifted from routine copying to finding the most efficient move sequences. Consideration of many alternative routes through the problem space can be demanding, and the extent to which this is performed internally or externally (Kirsh & Maglio, 1994) is ultimately up to the discretion of the problem solver. The task parameters can be seen as a hybrid bringing together the BWT and the eight-puzzle (Ericsson, 1974a), and thus allow for independent manipulation of both the cost associated with accessing task-relevant information (IAC) and making each move (IC). Using the BPST, the effects of the IAC can be observed during problem solving behaviour, and can be contrasted to behaviour promoted by IC. As already mentioned, higher ICs are known to increase planfulness during problem solving (O'Hara & Payne, 1998; 1999). Therefore, analysis of the effect of IC on BPST problem solving may provide a useful baseline from which to compare the effect of IAC. The demands placed upon working memory (Baddeley, 1986) by the BPST will likely vary according to IAC. Higher access costs are expected to lead to stronger reliance upon working memory during problem solving. Although such demands will not be measured explicitly, qualitative methods will be used to evaluate the distribution of memory between the world and the head (Norman, 1993). In a general sense, completion of the BPST will require, to greater or lesser extent, encoding of the rules of task engagement, encoding of the Goal-State configuration of coloured blocks, encoding of move sequences and relations between problem solving states (schemas).

Each study contained within Chapter 3 used an eye-tracker, whereby the frequency and duration with which participants' eyes visited each window were recorded in an attempt to provide fine grained insight into the strategic distribution of memory between the world and the head. Measures of the time taken between each move, and in some cases verbal protocol, were also used as indicators of the planning strategies conducted. In addition, the number of moves and time required to complete each problem were used throughout as two key measures of problem solving proficiency. Although the term 'move' could be used to describe both a single block movement and a chunked movement of several blocks (Chase & Simon, 1973), use of the term move in Chapter 3 refers to single bock movements only. This was deemed the only practical approach in absence of information relating to participants' sub goals (Anderson & Lebiére, 1998). Such analysis, although theoretically interesting, was beyond the scope of the current investigation.

As an overview, Experiment 3.1 manipulated IAC within the BPST described above in order to examine the extent to which the least-effort tradeoffs highlighted during routine copying behaviour (Fu & Gray, 2000; Gray et al., 2006) would extend to a relatively simple problem solving task. As expected, significant effects of IAC were observed with regards to the use of memory during problem solving. IC was also manipulated in the same manner as O'Hara & Payne (1998), and striking differences were found between these two variables. Based upon analyses of these differences, Experiments 3.2 and 3.3 attempted to further elucidate the role of IAC in determining the selection between different problem solving and planning strategies with varying task demands.

3.3. Experiment 3.1

The aim of Experiment 3.1 was to assess the extent to which least-effort tradeoffs contingent upon changes in IAC observed by Gray et al (2006; Fu & Gray, 2000) during low-level routine copying behaviour extend to a relatively simple problem solving task. Of primary importance was provision of an appropriate test of the primary hypothesis: How will small changes in IAC affect the use of internal memory during an externally represented well-structured problem solving task? As discussed previously, participants may adopt a number of strategies to cope with higher access costs. On the one hand, a higher IAC may prompt them to try to remember bigger chunks of the target pattern (Fu & Gray, 2000; Gray et al., 2006), and not worry about the number of moves made in the problem space. On the other hand, they may try more planning in the hope that fewer moves are required (O'Hara & Payne, 1998).

3.3.1. Method

Participants Seventy-two Cardiff University undergraduate Psychology students participated in the study for course credit and were randomly assigned to one of six conditions.

Apparatus/Materials The experiment was written in *Microsoft Visual Basic 6* and was conducted using a 2 Ghz Pentium IV PC connected to a *Tobii 1750* 34 x 27 cm eye-tracker monitor[1], extended keyboard, and mouse. All eye movements were recorded at a rate of 15 frames per second, with time-stamp accuracy of +/-3 ms. Due to rapid fixation switches observed, no limit was set on the duration of a fixation in order to be classified as a fixation. This avoided the loss of disproportionate rates of data from the Low IAC conditions when fixation filters were employed. Gaze estimation was within 1 degree of accuracy, even across large head movements. Mouse movements and key presses were also recorded and saved.

The Goal- and Current-State Windows were the same size, and ten coloured blocks and six empty spaces resided within each 4 x 4 grid. No colours were used twice, and the empty spaces were always white. The rules of the task determined that coloured blocks could only be moved into adjacent (horizontal or vertical) empty spaces.

---

[1] The *Tobii 1750* eye-tracker does not require head mounts.

Design Both IAC and IC were manipulated between-subjects in order to negate possible contamination via asymmetric transfer (e.g., Poulton, 1982).

IAC was manipulated on three levels: both windows were permanently uncovered when IAC was Low. In contrast, both windows were covered by grey masks when IAC was Medium or High, and could only be uncovered by placing the mouse cursor over the window to-be-opened. The grey masks then reappeared the moment the mouse cursor left the respective window. There was an additional 2.5 second lockout associated with uncovering the Goal-State when IAC was High[2].

IC was manipulated on two levels: when IC was Low, participants were required to select the block they wished to move via a left mouse click and then press the corresponding arrow key on the keyboard depending upon the direction they wished to move the block. When IC was High, participants were required to select the block they wished to move via a left mouse click and then type a string "move_left/right/up/down_".

Independent manipulation of IAC and IC resulted in six between-subject treatment conditions: Low IAC/Low IC; Low IAC/High IC; Medium IAC/Low IC; Medium IAC/High IC; High IAC/Low IC; High IAC/High IC. Based upon data collected during pilot testing pertaining to the average number of moves required to solve a number of arbitrarily produced BPST problems, task difficulty was also manipulated on two levels. Six problems were identified as 'Easy' due to all pilot participants requiring between thirty and forty moves to solve these problems. An additional six problems were identified as 'Hard' due to all solutions during pilot testing requiring between forty and fifty moves.

[2] 2.5 seconds was chosen because pilot studies revealed this to be analogous to the increase in time taken to make each move when IC was high.

Several dependent measures were taken throughout Experiment 3.1. Firstly, the frequency with which participants chose to inspect the Goal-State was measured. Necessarily, data from the Low IAC condition came from the eye-tracker (consecutive fixations within the Goal-State Window were collapsed and counted as one). In accordance with Gray & Fu (2001; 2004), data from the Medium and High IAC conditions came from the *Visual Basic* program (which measured the frequency with which the Goal-State Window was uncovered). The eye-tracker also provided data pertaining to the time participants spent viewing the Goal-State when the IAC was Low or Medium. However, when IAC was High, eye-tracker data was not used to measure Goal-State viewing time due to the fact that participants were able to view the area defined by the eye-tracker analysis software as representing the Goal-State whilst enduring the 2.5 s lockout[3].

The *Visual Basic* program also measured the time participants chose to uncover the Goal-State Window in the Medium and High IAC conditions. The eye-tracker was used to measure the time participants spent viewing the Current-State in all IAC conditions. The time participants spent looking down at the keyboard when making moves (more common in High IC conditions) naturally did not contribute to this data, and thus a precise measure of the time participants spent viewing the Current-State Window was obtained. In addition, the *Visual Basic* program recorded the number of moves made, the time taken to complete each trial, inter-move latencies (likely to reflect time taken to mentally evaluate the efficiency of different routes through the problem space), and the frequency with which participants clicked the stop button

---

[3] Despite the fact that the contents of the Goal-State were masked during the 2.5 s lockout, participants still chose to view the Goal-State area whilst waiting for the lockout to time out because there was nothing else of use to look at during this time. As it was not possible to instruct the eye-tracker analysis software not to record eye fixations during these lockout times, data contamination would likely occur in the High IAC condition if this measure was used.

when in fact the two patterns did not match (errors). By dividing the number of moves by the number of Goal-State visits for each trial, an estimate of the number of moves made per Goal-State inspection was obtained.

Procedure Participants were seated approximately 50 cm away from the eye-tracker and handed an instruction sheet (see Appendix J). Two practice trials then followed a 16-point eye-tracker calibration. Both practice trials were in the format of the experimental condition. Different block configurations were used for each of the twelve experimental trials (see Appendix A), and each participant within each of the six conditions received one of twelve different randomised orders of trials.

3.3.2. Results

A 3 (IAC Low/Medium/High) x 2 (IC Low/High) between-subjects ANOVA conducted upon each of the dependent variables revealed quite different results for IAC and IC (summarised in Tables 3.1 & 3.2 respectively). Only one interaction between the two independent variables existed, and will be addressed prior to separate discussion of the IAC and IC, summaries of the correlations between variables, and any practice effects incurred during completion of the twelve trials. Goal-State inspection frequency and uncover time data (the latter of which applies only to the Medium and High IAC conditions) were log transformed in order to attain homogeneity of variance. In addition, a reciprocal transformation was performed on the moves per Goal-State inspection data for the same reason. Non-transformed data are presented in tabular and graphical format throughout Chapter 3. The eye-tracker failed to record eye fixations for four participants. On such occasions, missing data points were removed from all analyses (Tabachnick & Fidell, 2001). No effects were found for task difficulty (Easy vs. Hard), and thus the levels of this variable were collapsed during all analyses and will not be discussed further.

Inter-relationships The only interaction between IAC and IC concerned the frequency with which participants inspected the Goal-State Window, F (2, 62) = 11.64, p <.001, MSE = 0.02, (see Figure 3.3). Simple main effects indicated that the effect of IAC was significant at both Low, F (2, 62) = 109.58, p <.001, MSE = 0.02, and High levels of IC, F (2, 62) = 149.99, p <.001, MSE = 0.02, with Bonferroni-corrected pairwise comparisons revealing inspection frequencies to decrease as IAC increased (ps <.001). However, although simple main effects indicated significantly more inspections as IC increased at both the Low, F (1, 62) = 8.23, p <.01, MSE = 0.02, and Medium levels of IAC, F (1, 62) = 43.49, p <.001, MSE = 0.02, no effect of IC was observed when IAC was High, F (1, 62) = 0.52, p >.05, MSE = 0.02. The data suggest that this is likely to be due to a floor effect (Cohen, 1995).

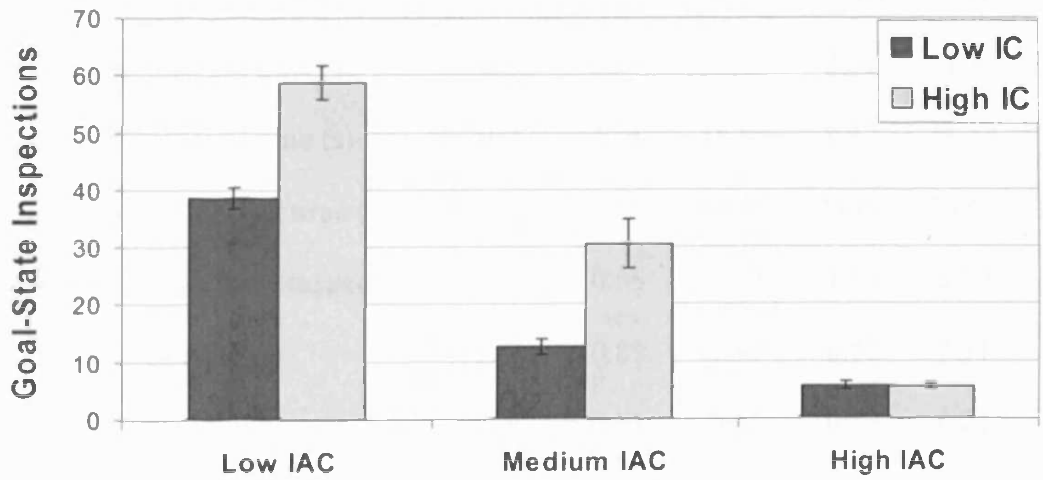


Figure 3.3: The interaction between IAC and IC for number of Goal-State inspections per trial (Experiment 3.1).

Note. Error bars represent ± 1 standard error.

The effect of IAC Analysis of the IAC x IC interaction reported above indicated that the frequency with which participants chose to inspect the Goal-State decreased as IAC increased at both levels of IC, F (2, 62) = 244.67, p <.001, MSE = 0.02, with post hoc comparisons (Bonferroni-corrected) revealing significant differences between all IAC conditions (ps <.001). Although the time participants chose to spend viewing the Goal-State (as measured by the eye-tracker) did not increase from Low to Medium, F (1, 40) = 0.46, p >.05, MSE = 21.16, the time participants chose to uncover the Goal-State (as measured by the *Visual Basic* program) did increase significantly as IAC rose from Medium to High, F (1, 44) = 4.86, p <.05, MSE = 0.02, (see Table 3.1).

Table 3.1

The Effect of IAC during Solution to a series of BPST problems (Experiment 3.1).

	Low IAC		*Medium IAC*		*High IAC*	
Measure	Mean	SD	Mean	SD	Mean	SD
Goal-State inspection frequency	46.65	12.54	21.59	14.24	5.64	1.99
Goal-State viewing time (s)	16.35	5.09	17.17	8.01	N/A	N/A
Goal-State uncover time (s)	N/A	N/A	18.06	8.37	21.13	8.74
Current-State viewing time (s)	55.98	23.72	66.26	24.84	73.67	25.74
Moves per Goal-State inspection	0.91	0.35	2.53	1.60	8.27	2.65
Inter-move latency (s)	1.91	0.83	1.99	0.83	2.11	0.92
Errors	0.06	0.11	0.13	0.13	0.21	0.30
Moves-to-completion	38.04	6.05	40.16	5.82	42.16	5.59

Note. Goal-State uncover time excludes lock-out delays incurred via High IAC.

Increasing IAC not only affected Goal-State inspection strategies, but also had consequences for the time spent viewing the Current-State, F (2, 62) = 3.56, p <.05, MSE = 291.08, with Bonferroni-corrected post hoc analyses revealing longer viewing times in the High than the Low IAC conditions (p <.05). In addition, the number of problem solving moves made per inspection of the Goal-State was significantly affected by IAC, F (2, 62) = 149.23, p <.001, MSE = 0.05, with Bonferroni-corrected post hoc analyses revealing the number of moves made per Goal-State inspection to rise in accordance with IAC (ps <.001). When IAC was Low, participants generally chose to interleave a minimum of one visit to the Goal-State per move made in the Current-State. When the IAC was High, however, participants were prepared to make up to eight moves in the Current-State per Goal-State inspection.

With regard to measures deemed to reflect planning behaviour and problem solving proficiency, no differences were observed between IAC conditions with respect to average inter-move latencies, either when all time spent viewing the Goal-State was included in the analysis, F (2, 66) = 0.75, p >.05, MSE = 0.32, (see Table 3.1) or when the analysis was limited to the Medium and High IAC conditions and time spent uncovering the Goal-State was excluded, F (1, 44) = 0.94, p >.05, MSE = 0.32, (Medium IAC: *Mean* = 1.54, SD = 0.75; High IAC: *Mean* = 1.59, SD = 0.73). The number of errors made was affected by IAC, F (2, 66) = 3.58, p <.05, MSE = 0.04, and Bonferroni-corrected post hoc analyses indicated that more errors were made in the High than the Low IAC condition (p <.05). Trial duration was also affected by IAC, F (2, 66) = 6.25, p <.01, MSE = 801.07, with Bonferroni-corrected post hoc analyses revealing longer times in the High than the Low IAC conditions (p <.01) (Low IAC: *Mean* = 125.91, SD = 66.79; Medium IAC: *Mean* = 135.79, SD = 76.06; High IAC: *Mean* = 154.37, SD = 70.27). Finally, a small, but significant effect of IAC was also found for the number of moves required to solve each trial, $F(2, 66) = 6.81$, $p < .01$, $MSE = 14.89$, with Bonferroni-corrected post hoc analyses pointing to more moves required in the High than the Low IAC conditions ($p < .001$).

In sum, higher access costs experienced during solution to a series of BPST problems led to increased reliance upon memory (as evidenced by a reduction in the frequency with which Goal-State information was inspected and an increase in the number of moves made per Goal-State inspection), did not affect the key measure of planning (inter-move latencies), but led to reductions in problem solving proficiency (as evidenced by increases in trial duration, error frequency, and the number of moves required to solve each problem).

The effect of IC Increasing the cost associated with making each move invoked comparatively different strategy selection. Firstly, analysis of the interaction between IAC and IC Goal-State inspection frequency reported earlier indicated that although higher IC led to significantly more inspections at the Low, $F(1, 62) = 8.23$, $p < .01$, $MSE = 0.02$, and Medium levels of IAC, $F(1, 62) = 43.49$, $p < .001$, $MSE = 0.02$, no effect of IC was observed when IAC was High, $F(1, 62) = 0.52$, $p > .05$, $MSE = 0.02$.

As observed with an increase in IAC, higher implementation costs increased the time spent viewing the Goal-State (as measured by the eye-tracker), $F(1, 40) = 52.47$, $p < .001$, $MSE = 21.16$, and the time spent uncovering the Goal-State (as measured by the *Visual Basic* program), $F(1, 44) = 43.79$, $p < .001$, $MSE = 0.02$. Also in line with an increase in IAC were longer Current-State viewing times (as measured by the eye-tracker) as a function of higher IC, $F(1, 62) = 75.31$, $p < .001$, $MSE = 291.08$. However, rather than increasing the moves per Goal-State inspection ratio, the High

IC reduced the number of moves participants were prepared to make per inspection of the Goal-State, $F(1, 62) = 61.95$, $p < .001$, $MSE = 0.05$, (see Table 3.2).

Table 3.2

The Effect of IC during Solution to a series of BPST problems (Experiment 3.1).

	Low IC		*High IC*	
Measure	Mean	SD	Mean	SD
Goal-State inspection frequency	18.97	15.11	28.25	23.25
Goal-State viewing time (s)*	12.11	3.21	22.42	5.53
Goal-State uncover time (s)**	14.18	3.65	25.01	8.76
Current-State viewing time (s)	48.42	12.78	85.46	21.68
Moves per Goal-State inspection	4.56	3.64	3.55	3.69
Inter-move latency (s)	1.36	0.30	2.65	0.73
Errors	0.18	0.25	0.09	0.14
Moves-to-completion	44.45	5.14	35.79	2.78

Note. Goal-State uncover time excludes lock-out delays incurred via High IAC.

* Excluding data from High IAC condition

** Excluding data from Low IAC condition

In addition, increasing IC had an effect upon the key measure of planning: significant rises in inter-move latency were observed as a function of IC, both when all time spent viewing the Goal-State was included, $F(1, 66) = 93.83$, $p < .001$, $MSE = 0.32$, (see Table 3.2) and when analyses were limited to the Medium and High IAC conditions and the measure of inter-move latency was restricted to time spent working in the Current-State Window (Low IC: *Mean* = 1.09, *SD* = 0.25; High IC: *Mean* =

2.03, *SD* = 0.75), $F(1, 44) = 32.83$, $p < .001$, $MSE = 0.32$. Increasing IC also reduced the frequency with which errors were made, $F(1, 66) = 4.13$, $p < .05$, $MSE = 0.04$, and the number of moves required to solve each problem, $F(1, 66) = 90.62$, $p < .001$, $MSE = 14.89$. As with IAC, trial times increased as a function of IC, $F(1, 66) = 369.34$, $p < .001$, $MSE = 801.07$, (Low IC: *Mean* = 74.59, *SD* = 19.53; High IC: *Mean* = 202.79, *SD* = 37.81).

In sum, increasing the cost of making each move from pressing an arrow key to typing a string decreased reliance upon memory during problem solving (evidenced by an increase in the frequency with which participants chose to inspect Goal-State information and a reduction in the number of moves made per Goal-State inspection), significantly affected the key measure of planning (reliable increases in inter-move latencies), and improved problem solving proficiency when considering errors and moves made, but not when considering time-to-solution.

Correlations Because separate univariate ANOVAs were performed upon each dependent variable, it is important to examine how these variables may interrelate. Therefore, Spearman's rho correlation was computed on all non-transformed data in order to assess the potential inter-relationships. Correlations were computed separately within each of the six between-subject treatment combinations, and also across the complete data-set. The only significant correlation consistently found within each of the six treatment combinations and across the entire data-set indicated that as Goal-State inspection frequencies increased, the number of moves made per Goal-State visit decreased (-.94, $p < .01$). Despite similar correlations, this relationship varied with IAC (as illustrated by Figure 3.4).

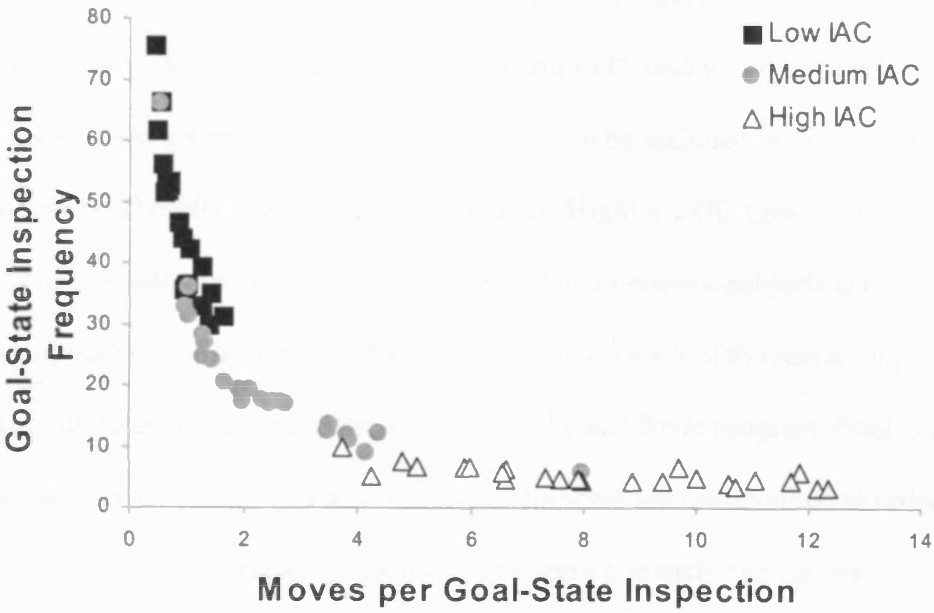


Figure 3.4: A scatter-plot of the relationship between Goal-State inspection frequency and moves per Goal-State inspection as a function of IAC (Experiment 3.1).

For example, the number of moves made per Goal-State inspection varied substantially in the High IAC condition (on average four to twelve), despite each participant reliably viewing the Goal-State fewer than ten times per trial. In contrast, the number of moves per Goal-State inspection remained relatively stable in the Low IAC conditions (fewer than two), despite substantial variation in Goal-State inspection frequency.

Other significant correlations across the data-set included those found between Goal-State inspection frequency and moves-to-solution (-.39, p <.01), moves per Goal-State inspection and moves-to-solution (+.51, p <.01), and inter-move latency and moves-to-solution (-.59, p <.01). These relationships between the aforementioned variables indicated that frequent inspections of the Goal-State and long latencies between moves are associated in some way with a reduction in the number of moves required to solve each BPST problem.

Practice effects Finally, the contribution of any practice effects within the current experiment were observed as a function of trial and IAC. Due to the manner in which the eye-tracker was set up, eye fixation data could not be included in the evaluation of practice effects. Therefore, a 3 (IAC Low/Medium/High) x 2 (IC Low/High) x 12 (Trial) ANOVA with the first two factors manipulated between-subjects and the final factor manipulated within-subjects was conducted upon each of the remaining dependent variables using data gathered from the *Visual Basic* program. Goal-State Window uncover frequency and uncover time data were log transformed in order to attain homogeneity of variance. In addition, a reciprocal transformation was performed on the moves per Goal-State inspection data for the same reason. Where violations of sphericity occurred, the Greenhouse-Geisser corrected degrees of freedom are reported accordingly.

Overall, problem solving proficiency (as measured by moves and time to solution) was found to improve as a function of practice (*ps* $<.001$). In terms of strategy selection, this meant that the frequency with which participants chose to inspect the Goal-State reduced in linear fashion as a function of trial, $F(1, 44) = 15.15, p < .001, MSE = 0.02$, as did the time participants chose to spend uncovering the Goal-State, $F(1, 44) = 10.60, p < .001, MSE = 0.03$. However, neither the number of moves participants chose to make per Goal-State inspection nor inter-move latency data was affected by practice on task (*ps* $>.05$), thus suggesting that the effect of IAC upon the use of memory was robust and consistent across trials. Finally, the absence of any interaction between IAC, IC and trial (*ps* $>.05$) indicated that the observed effects of practice were consistent across all conditions.

3.3.3. Discussion

First and foremost, the results of Experiment 3.1 indicated that the selection between microstrategies contingent upon changes in IAC previously observed during routine copying behaviour (Fu & Gray, 2000; Gray et al., 2006) extend to a simple problem solving task. As the cost of accessing Goal- and Current-State information increased, so did participants' propensity to rely upon memory-, rather than interaction-intensive problem solving search strategies. Secondly, this appears to be despite difficulties encountered during problem solving in the High IAC condition, and is in contrast to the more efficient use of interaction-intensive, planful microstrategies promoted by an increase in IC. Importantly, the effect of IAC upon strategy selection was also consistent over time, despite improvements in problem solving efficiency resulting from practice (see Anzai & Simon, 1979; Glaser & Bassok, 1989; Matthews, Davies, Westerman, & Stammers, 2000 for useful reviews).

The rational analysis framework (Anderson, 1990) has previously been adapted to explain the tradeoff between interaction- and memory-intensive strategies contingent upon changes in IAC during routine copying tasks (Fu & Gray, 2000; Gray et al., 2006) and between planning and acting moderated by changes in IC during solution to the eight-puzzle (O'Hara & Payne, 1998; 1999). Both IAC and IC were simultaneously manipulated in the current study, and support was found for the adaptive use of memory contingent upon changes in IAC, and the adaptive deployment of planning as a consequence of changes in IC. Surprisingly, very little interaction between IAC and IC was observed. Overall it was found that as IAC increased, so did reliance upon internal memory. It is proposed that when access costs were High, accessing Goal-State information following each move made in the Current-State Window was considered more costly than using memory (which would reduce the number of times Goal-State information needed to be inspected). Similarly, as the cost of making each move increased, it is argued that the benefit of planning began to outweigh the cost of acting without planning (thus reducing the number of moves required to complete each problem).

The shift in task from routine copying to problem solving appears to have had little, if any, effect upon the adaptive use of memory (Anderson & Milson, 1989; Anderson & Schooler, 1991), thus providing strong support for the finding that as IAC increases so does the use of memory- as opposed to display-based interactive strategies (Fu & Gray, 2000; Gray & Fu, 2001; 2004; Gray et al., 2006). As IAC increased, participants chose to inspect the Goal-State less frequently and made considerably more moves in the Current-State per inspection of the Goal-State. However, the small but significant rise in the number of moves required to solve the series of BPST problems associated with an increase in IAC indicated that participants found these problems harder to solve efficiently when access costs were High.

There are a number of possible reasons why an increase in IAC may have led to a small rise in moves-to-solution. Firstly, the associated increment in reliance upon internal memory is naturally more error-prone than reliance upon the external display (Anderson & Lebiére, 1998). The hard constraints of the task meant that participants in the Medium and High IAC conditions were unable to view both windows simultaneously at any point, and thus, error-monitoring will have been considerably more difficult. Indeed, the frequency with which errors occurred (pressing the stop button when in fact the two patterns did not match) did increase in conjunction with an increase in IAC. However, the rise in errors made as IAC increased does not fully explain the corresponding rise in moves-to-solution. When error-correction trial data were removed from the analysis, the increase in moves-to-solution remained.

Secondly, it is possible that an increase in IAC resulted in a rise in the number of errors made during problem solving (e.g., mistakenly moving a block to the incorrect position), that were subsequently corrected prior to pressing the Stop Trial button. However, no measure of the number of 'corrected errors' can be provided due to complications associated with distinguishing between corrected errors and unblocking moves (movement of a block in the opposing direction to the destination cell in order to make space for the movement of another block).

A third explanation offered, upon which further efforts will be concentrated, centre's upon the possibility that the increase in reliance upon memory (as a function of IAC) may have made planning difficult. The close-knit relationship between memory and planning (see Gilhooly, 2004 for a review) suggests that changes in memory load will often have consequences for planning proficiency. Although in a related study Pfeiffer (2004) did not find this to be the case, low demands placed upon working memory may have accounted for this. Work examining problem solving with higher internal memory load has highlighted substantial difficulties encountered during planning (e.g., Kotovsky et al., 1985; Phillips et al., 1999). The fact that participants in the Medium and High IAC conditions of the current experiment were only able to view one window at a time will have meant that any planning conducted will necessarily have had to rely upon memory for either the Goal- or the Current-State (both of which contain 10 coloured blocks and 6 empty spaces). In order to further explore the relationship between IAC and planning within the current context, it is first necessary to compare and evaluate the selection between microstrategies based upon changes in IAC and IC.

There appear to be two critical differences with regard to the selection between candidate problem solving search strategies based upon increments in IAC and IC. Firstly, increasing IAC did not affect inter-move latencies, whereas increasing IC did (see O'Hara & Payne, 1998; 1999 for similar findings regarding IC). Secondly, rises in IAC promoted more memory-dependent internal strategies (extending the work of Gray et al., 2006), whereas the increment in IC encouraged more display-based problem solving (extending O'Hara & Payne's, 1998 interpretation of the effects of IC upon planning behaviour). With regard to planning therefore, it would seem that small increments to IAC promoted rather different strategies to those induced via small increments to IC. Davies's (2003) recent assessment of the planning literature will be used as a conceptual tool for further instantiating these differences.

According to Davies (2003), human planning during problem solving generally takes one of two overarching forms; 'initial' versus 'concurrent'. Initial planning is said to be characterised by mental evaluation of the problem space prior to the manipulation of any external operators. Concurrent planning, on the other hand, is defined as 'on-line' planning conducted during solution to a problem. Although Davies's (2003) original definition of initial planning will be adjusted somewhat for the purpose of the current investigation (see below), it is crucial to note that the extent to which planning is memory- or display-based, and the length of inter-move latencies, are both argued to represent major defining differences between initial and concurrent planning styles (Davies, 2004).

Essentially, initial planning invokes stronger reliance upon memory and shorter inter-move latencies, whereas concurrent planning generally elicits longer inter-move latencies and less reliance upon internal memory. The ability to inspect and re-inspect externalised information is seen as a prerequisite to concurrent planning, yet the vast majority of initial planning is observed to occur largely without the use of external representations (Davies, 2004).

Concurrent planning is proposed to be characterised by 'partial-order' evaluation of problem solving routes (cf. Ratterman, Spector, Grafman, Levin, & Harward, 2001), whereby the processes of planning and acting are largely interleaved (Anderson, 1990). When indulging in concurrent planning, we do not plan long sequences before we perform them, but rather, evaluate our plans as we progress through the problem space (Atwood & Polson, 1976; Delaney, Ericsson, & Knowles, 2004; Simon & Reed, 1976). Concurrent planning is said to be 'opportunistic' (Hayes-Roth & Hayes-Roth, 1979) in the sense that plans can be easily revised in the face of new information. Based upon this assessment, it appears that the strategies evoked by an increase in IC (longer inter-move latencies and regular inspection of the Goal-State between moves) map rather well onto Davies's (2003) conceptualization of concurrent planning, thus extending O'Hara & Payne's (1998) consideration of the nature of the planning promoted by a rise in IC.

The strategies associated with an increase in IAC, however, appear less aligned with concurrent planning, and for the following reasons seem more akin to an initial planning style. As the cost associated with accessing task-relevant information increased in Experiment 3.1, participants reliably chose to inspect such information less frequently. Therefore, a concurrent planning strategy would not have been of benefit. Instead, it is argued that the increase in reliance upon internal memory during problem solving would more likely promote an initial planning strategy.

Bearing in mind the limited number of times participants in the High IAC condition seemed prepared to inspect the Goal-State, and the boundaries of visual short-term memory (Luck & Vogel, 1997), it would seem logical to presume that upon each inspection of the Goal-State, participants in High IAC conditions would engage in some form of initial planning. This may be as rudimentary as committing a chunk of 3-4 blocks to memory (Chase & Simon, 1973), or as complex and demanding as formulating a plan based upon memory for the Current-State. Either way, this process will be entirely internal, will not involve the manipulation of any external representations, and could therefore be seen as an incidence of initial planning (unless inspection of the Goal-State was purely motivated in order to evaluate actions previously made in the Current-State, which could also contribute to a concurrent planning style).

Although Davies (2003) originally defined initial planning to be limited to 'total-order' mental evaluation of the problem space prior to the manipulation of any external operators (cf. Ratterman et al., 2001), it is argued here that small access costs situated appropriately within the task environment may promote a series of initial planning 'episodes' or 'stages' throughout the problem solving process. As described above, access costs could be used to ensure that an initial planning episode still be determined by internal, rather than external, consideration of the problem space. Indeed, such proposed episodes of initial planning would likely lead to relatively long followed by relatively short inter-move latencies, and thus would remain aligned with Davies's (2003) high-level conceptualisation of initial planning.

3.4. Experiment 3.2

In an attempt to provide a truer representation of the planning styles elicited by small changes in IAC, the BPST was modified for Experiment 3.2 so as to reduce the demands placed upon working memory during problem solving in the Medium and High IAC conditions. To this end, the Current-State Window remained visible at all times allowing participants to engage in the formulation of plans whilst simultaneously viewing both the Current- and the Goal-State Window (thus reducing demands upon working memory, and therefore the perceived cost of planning).

IAC, however, was still manipulated on the Goal-State, and an in-depth analysis of the distribution of inter-move latencies aimed to further evaluate the effect of IAC upon on the nature of planning. Following from the previous discussion, it was predicted that small increments in IAC would promote an initial planning style (both prior to the manipulation of any external representations, and throughout the entire problem solving process), yet discourage concurrent planning. The use of first- as well as inter-move latencies to inform the assessment of how IAC affected the selection between initial and concurrent planning styles was based upon previous exploration of the use of initial and concurrent planning during solution to the Tower of Hanoi (Davies, 2003). Within this analysis, relatively long first-move latencies (relative to problem solving task) coupled with a series of subsequent short inter-move latencies were identified as reflections of initial planning, whereas regular medium-sized inter-move latencies (and shorter first-move latencies) were argued to be representative of concurrent planning.

It was also predicted that the experimental manipulation would not affect the influence of IAC upon the selection between memory- and interaction-intensive microstrategies, but would encourage planning behaviour in the Medium and High IAC conditions, and when compared to Experiment 3.1, reduce the number of moves required to solve each problem in the Medium and High IAC conditions. For analyses that would benefit from Low IAC data, the Low IAC/Low IC condition from Experiment 3.1 was included because this condition would not be affected by the BPST modification currently under investigation.

3.4.1. Method

Participants Twenty-four Cardiff University undergraduate Psychology students participated in the study for course credit, and were randomly assigned to either the Medium or the High IAC condition.

Apparatus/Materials The experimental and recording equipment used were identical to those employed during Experiment 3.1, with the exception that the Current-State Window was always visible.

Design & Procedure When IAC was Medium, the Goal-State Window could be uncovered by placing the mouse cursor over the window. This was also applicable when IAC was High, with an additional 2.5 second lockout delay. No new Low IAC data was collected because this condition was not affected by leaving the Current-State open at all times. To make each move, participants were required to select the block in the Current-State Window they wished to move via a left mouse click, and press the corresponding arrow key depending upon the direction they wished to move the block (equivalent to the Low IC in Experiment 3.1). IC was not manipulated in Experiment 3.2, but the experimental procedure was identical to that employed during Experiment 3.1 (see Appendix J for task instructions).

In addition to the dependent variables taken in the previous experiment, the eye-tracker was also used to measure the frequency with which saccadic eye movements were made between the two windows during periods when participants chose to cease making moves and view both windows simultaneously. Given that any time spent dedicated to memorising the Goal-State pattern would not require saccadic eye movements back and forth between the Goal- and Current-State, this may provide a measure of the extent to which participants chose to engage in planful behaviour whilst viewing both windows simultaneously (see Goldberg, Stimson, Lewenstein, Scott, & Wichansky, 2002).

3.4.2. Results & discussion

Firstly, the effect of leaving the Current-State Window open upon the Medium and High IAC conditions will be evaluated with respect to the hypothesis that it will encourage planful behaviour and improve problem solving proficiency. Table 3.3 provides a comparison between the problem solving strategies selected when IAC was manipulated on the Current-State Window (Current-State IAC) to when it was not (Current-State Open). Secondly, a fine-grained analysis of the distribution of inter-move latencies will be reported in an attempt to provide quantitative insight into the time allocated to mental evaluation of the problem space between each move made in the Medium and High IAC conditions from the current experiment, and the Low IAC/Low IC condition from the previous experiment.

Evidence for planning In order to ascertain the effect of leaving the Current-State Window permanently open during solution to a series of BPST problems, a 2 (IAC Medium/High) x 2 (Current-State IAC/Open) ANOVA was performed upon each of the dependent variables. Goal-State uncovering time and error data was log transformed in order to meet assumptions of homogeneity of variance. Neither the frequency with which participants chose to uncover the Goal-State Window, $F(1, 44) = 0.35$, $p > .05$, $MSE = 9.13$, or the number of moves made per uncovering, $F(1, 44) = 0.54$, $p > .05$, $MSE = 4.64$, were affected by leaving the Current-State Window open at all times. Thus, as expected, the distribution of memory according to IAC followed much the same pattern as reported in Experiment 3.1, and during routine copying behaviour (Fu & Gray, 2000; Gray et al., 2006).

Table 3.3

The Effect of Current-State Accessibility during Solution to a series of BPST problems (Experiments 3.1 & 3.2).

	Current-State IAC				Current-State Open			
	Medium IAC		High IAC		Medium IAC		High IAC	
Measure	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Goal-State inspection frequency	**12.60**	4.49	**5.72**	2.07	**13.33**	3.26	**4.67**	1.21
Moves per Goal-State inspection	**3.65**	1.54	**8.89**	2.34	**3.31**	0.83	**9.52**	3.17
Goal-State uncover time (s)	**11.68**	2.35	**16.67**	2.96	**23.33**	6.29	**31.39**	18.01
Current-State viewing time (s)	**47.53**	11.17	**52.00**	6.54	**43.09**	8.89	**49.39**	11.17
Time-to-solution	**68.04**	14.92	**91.70**	19.60	**63.74**	12.02	**85.09**	17.11
Errors	**0.14**	0.10	**0.28**	0.39	**0.01**	0.03	**0.10**	0.16
Moves-to-solution	**44.85**	3.92	**46.59**	4.33	**41.92**	4.51	**41.19**	3.92

Note. Goal-State uncover time excludes lock-out delays incurred via High IAC.

Although leaving the Current-State Window open at all times did not affect the frequency with which participants chose to uncover the Goal-State Window, or the number of moves made per Goal-State inspection, it did appear to encourage participants to uncover the Goal-State Window for longer periods of time, $F\ (1, 44) = 10.03$, $p < .001$, $MSE = 0.09$. This was true for both the Medium ($p < .01$) and High ($p < .001$) IAC conditions. Although marginally faster solution times were observed when the Current-State was permanently visible (see Table 3.3), this difference was not significant, $F(1, 44) = 1.37$, $p > .05$, $MSE = 261.09$, and participants in Experiment 3.2 did not compensate for increased Goal-State uncover times by spending less time viewing the Current-State Window (as measured by the eye-tracker), $F(1, 44) = 1.01$, $p > .05$, $MSE = 140.00$. This suggests, again as expected, that participants took the opportunity to view both windows simultaneously.

Support for the hypothesis that participants would choose to view both windows simultaneously in order to engage in planful behaviour was also obtained by dividing the number of saccadic eye fixations to the Goal-State per uncovering (see Figure 3.5).

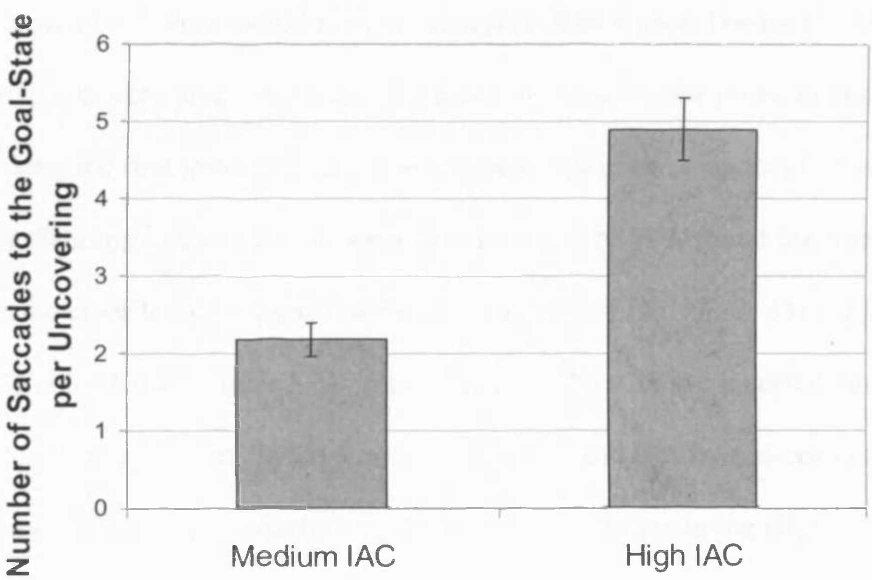


Figure 3.5: Frequency of Goal-State saccades per uncovering (Experiment 3.2).

Note. Error bars represent ± 1 standard error.

A separate one-way ANOVA indicated that the increase in the number of saccades to the Goal-State per uncovering as a function of IAC was indeed significant, $F(1, 22) = 25.45$, $p < .001$, $MSE = 2.17$. Furthermore, providing participants in the Medium and High IAC conditions with the option to view both the Goal- and the Current-State simultaneously when not making moves also improved problem solving proficiency (also as predicted). The number of moves required to solve the same set of BPST problems was significantly lower when IAC was High in the current experiment, when compared to the Low IC/High IAC condition in Experiment 3.1, F (1, 44) = 10.01, p <.01, MSE = 17.45. A significant reduction was also observed in the number of errors made in the High IAC condition, F (1, 44) = 5.61, p <.05, MSE = 0.01.

Distribution of inter-move latencies In order to assess the distribution of inter-move latencies across all three IAC conditions, data from the Low IAC/Low IC condition from Experiment 3.1 were included in the analyses. Based upon Davies (2003), the time participants were prepared to spend mentally evaluating the problem space before making the first move was taken as a measure of initial planning before any concurrent planning could occur. A separate one-way ANOVA found log transformed first-move latencies to differ significantly as a function of IAC, F (2, 33) = 16.65, p <.001, MSE = 0.03, (see Figure 3.6). Importantly, all 2.5 s delays incurred during the High IAC condition were excluded from this analysis, and Bonferroni-corrected post hoc analyses revealed significantly longer first-move latencies in the High, when compared to the Medium and Low IAC conditions (*ps* <.01).

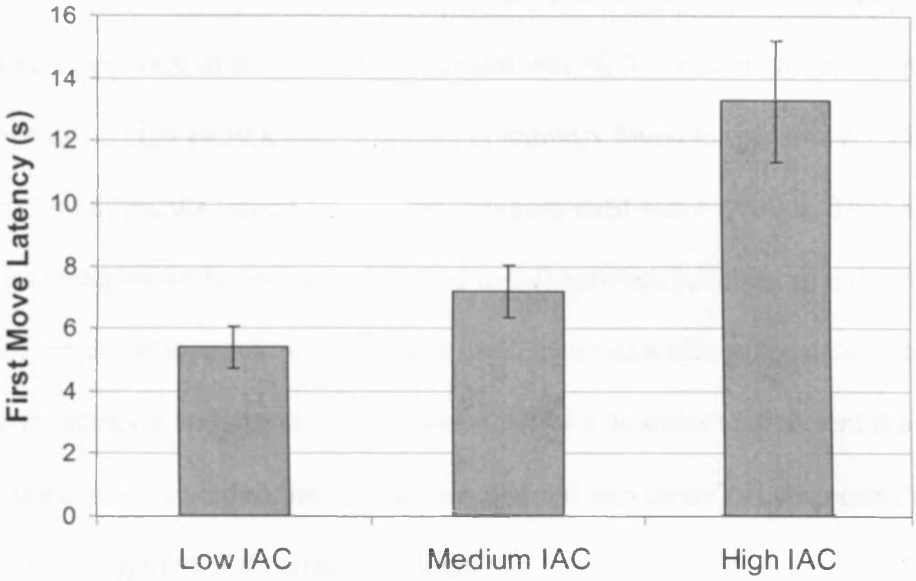


Figure 3.6: The effect of IAC on first-move latencies (Experiment 3.2).

Note. Error bars represent ± 1 standard error. Low IAC data taken from Experiment 3.1.

In an attempt to examine the extent to which episodes of initial planning were promoted by an increase in IAC throughout the remainder of the problem solving process, all inter-move latencies for every participant during solution to the first trial were initially determined as falling within a series of 0.25 s time frames, ranging from =0.25 to >29.75 =30 s. No specific work could be found upon which to govern the development of inter-move latency time categories that would represent periods of initial and concurrent planning. Thus, a number of 'categories of interest' were generated iteratively until it was possible to represent the natural trends observed within the current data, consistent with Davies's (2003) assertions concerning initial and concurrent planning within inter-move latency data.

According to Davies (2003) initial planning is typically characterised by relatively long and relatively short inter-move latencies (representing extended mental evaluation of move sequences, and the time to make pre-planned moves respectively).

As no inter-move latencies were found to be shorter than 0.25 s in duration, the shortest category used in the current assessment was =0.5 s. Although inter-move latencies rose as high as 30 s, they were not commonly found to rise above 7 s in duration. Therefore, the largest time frame category used was >7 =30 s. The vast majority of inter-move latencies were found to fall between 0.5 and 1 s, and thus a third time frame category of >0.5 =1 s was used. The remainder of the data included inter-move latencies with durations between 1 and 7 s. In order to represent the range of these data, this six second time frame was divided into three 2 s categories (>1 =3, >3 =5, >5 =7). Figure 3.7 illustrates the data.

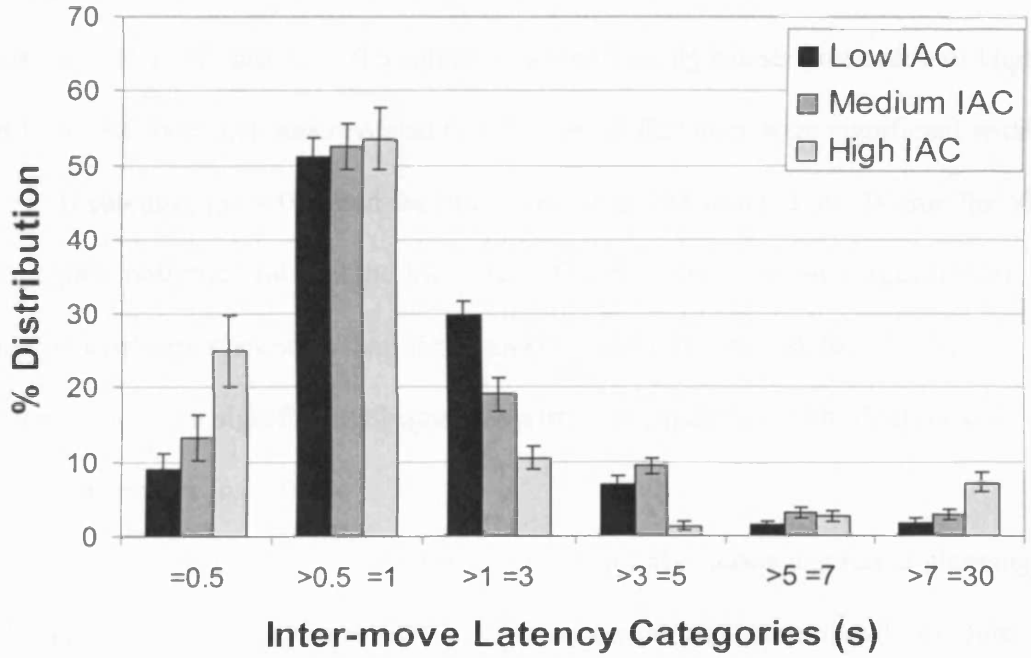


Figure 3.7: The distribution of inter-move latencies during solution to the first BPST as a function of IAC (Experiment 3.2).

Note. First-move latencies were excluded from this analysis. Error bars represent ± 1 standard error. Low IAC data taken from Experiment 3.1.

A 2 (IAC Low/Medium/High) x 6 (Category) ANOVA with the first factor manipulated between-subjects and the second factor manipulated within-subjects was performed on the log transformed data. An interaction between IAC and Category, F (4.05, 66.84) = 7.58, p <.001, MSE = 0.02, revealed that the inter-move latency categories were affected differentially by IAC. Simple main effects indicated that the effect of IAC was significant at the =0.5, >1 =3, >3 =5, and >7 =30 categories (*ps* <.01), but not the >0.5 =1 or >5 =7 categories (*ps* >.05).

In order to reveal the nature of the differential distribution of inter-move latencies between the three IAC conditions, Bonferroni-corrected post hoc analyses were performed upon each of the categories at which IAC was found to have had a significant effect. Within the =0.5 category, a significantly higher proportion of High than Low IAC latencies was revealed (p <.01). All differences were significant within the >1 =3 category (*ps* <.01), and the order went High>Medium>Low. Within the >3 =5 category, both the Low and the Medium IAC conditions revealed a significantly higher proportion of latencies than the High (*ps* <.001). In contrast, the >7 =30 category revealed a significantly higher proportion of High than both Medium and Low IAC latencies (*ps* <.001).

The emerging pattern of data supports the main hypothesis that initial planning will be promoted by an increase in IAC, whereas concurrent planning will not. Inter-move latencies provided by participants in the High IAC condition were significantly more likely to fall within the shortest (<0.5) and longest (>7 =30) categories than the latencies provided by participants in the Medium and Low IAC conditions. The long inter-move latencies are interpreted to be indicative of an initial planning episode (cf. Davies, 2004), and the very short inter-move latencies are thought to represent time between moves that were largely pre-planned, and thus implemented from memory

(Phillips, Wynne, McPherson, & Gilhooly, 2001). Compared to the Low and Medium IAC conditions, the paucity of inter-move latencies obtained from the High IAC condition falling within the >1 =3 and >3 =5 categories is argued to represent a reluctance to engage in concurrent planning.

It is also worth noting at this point that the two categories within which no effects of IAC were observed (>0.5 =1, >5 =7), contained the majority and minority of inter-move latencies respectively. Understanding quite why above 50% of all inter-move latencies fell within the >0.5 =1 category, yet inter-move latencies of >5 =7 s in duration were so rare is somewhat peripheral to the goals of the study, and is thus left open to speculation. What is important, given the aim of this investigation, is that the longest and shortest inter-move latencies were both predominantly made up of data from participants in the High IAC condition, whereas the middle-sized inter-move latencies were largely made up of data provided by participants in the Low IAC condition.

In sum, Experiment 3.2 demonstrated that participants in the Medium and High IAC conditions did make use of the facility to engage in the formulation of plans whilst viewing both windows simultaneously. This did not affect the manner by which participants chose to use internal memory, but did lead to an improvement in problem solving proficiency during solution to a series of BPST problems. Detailed quantitative analyses of participants' inter-move latencies suggested that an increase in IAC resulted in problem solving represented by very short and relatively long inter-move latencies, whereas lower access costs encouraged problem solving characterised by a variety of medium-sized inter-move latencies. Following Davies's (2003) analysis of planning style based upon inter-move latencies, it is argued that the data collected in Experiment 3.2 provide preliminary evidence that small increases in IAC promote an initial planning style, yet discourage a concurrent planning style.

However, attention should be paid to two substantial limitations associated with Experiment 3.2. Firstly, one must always be cautious when making inferences relating to psychological processes (such as planning) based upon quantitative data such as inter-move latencies. Although Experiments 3.1 and 3.2 have provided invaluable insight into the distribution of memory between the head and the world, it is anticipated that collection of verbal protocol (Ericsson & Simon, 1993) would provide further important insight into the nature of planning occurring within each of the IAC conditions. Secondly, due to there being six empty spaces residing within the Current-State at all times, the complexity of the planning required to complete the series of BPST problems used in Experiments 3.1 and 3.2 is relatively low when compared to other tasks used in the literature to uncover the processes underlying human problem solving and planning (e.g., Tower of Hanoi). Therefore, a final experiment will be conducted in order firstly to seek qualitative support for the distinction between concurrent and initial planning proposed to be contingent upon changes in IAC, and secondly to evaluate the extent to which the results obtained thus far extend to a more complex problem solving task in which planning is critical for successful solution. Although there are no strong grounds to predict that the relationship between IAC, memory and planning will change when more complex problem solving is required, it seems important to check that the expected longer solution times and more complex planning do not prompt participants to change their problem solving strategy (e.g., memorise the Goal-State prior to problem solving so as reduce demands on working memory).

3.5. Experiment 3.3

In an attempt to extend the results gained thus far to a more complex problem solving task, and to allow for the collection of verbal protocol (Ericsson & Simon, 1993), the BPST used in Experiment 3.2 was modified for the purposes of Experiment 3.3 in such a manner so as to mimic the problem solving demands of the much researched eight-puzzle (see Ericsson, 1974a,b,c). As with the BPST, the eight-puzzle requires the problem solver to rearrange a configuration of blocks to reach a Goal-State. However, because there are only eight blocks and one empty space included within the eight-puzzle, planning is crucial during solution to this task. Prior to discussion of the development of novel coding categories to differentiate between initial and concurrent planning during solution to the eight-puzzle-like BPST, a brief discussion of the costs and benefits of verbal protocol will ensue.

Verbal protocol In order to provide qualitative insight into the nature of planning contingent upon IAC, verbal protocol will be collected during solution to the eight-puzzle-like BPST in line with the guidelines developed by Ericsson & Simon (1993) and more recently Patrick & James (2004). Verbal protocol has been used extensively as a method of tracing complex cognitive processes during problem solving behaviour (e.g., Anzai & Simon, 1979; Ericsson, 1975; Fleck & Weisberg, 2004; Hayes-Roth & Hayes-Roth, 1979; Newell & Simon, 1972; O'Hara & Payne, 1998; Ratterman et al., 2001). Essentially, the method requires the participant to verbalise their thoughts concerning the solution to a particular problem, and thus much insight can be gained regarding the processes underlying behaviour (see Ericsson & Simon, 1993 for an extensive description of this method).

Inevitably, some speculation has arisen over the years regarding the potential impact of verbal protocol methodology upon task performance (e.g., Ahlum-Heath & Di Vesta, 1986; Berardi-Coletta, Buyer, Dominowski, & Rellinger, 1995; Gagne & Smith, 1962; McGeorge & Burton, 1989; Schooler, Ohlsson, & Brooks, 1993). The major focus has been upon whether verbalisation whilst solving a problem changes problem solving performance. For example, it has been argued that the act of verbalisation can improve performance on a task by encouraging participants to evaluate more extensively current and future situations (Berardi-Coletta et al., 1995). In contrast, Schooler et al (1993) have suggested that on occasion, instructing participants to 'think aloud' during problem solving may lead to *verbal overshadowing*. By this, Schooler et al (1993) refer to situations where the processes involved in solving a problem are not readily verbalisable. If prompted to 'think aloud' under such conditions, the problem solver may focus disproportionate attention upon the processes easier to verbalise, and performance may be impaired. Also, it is possible that the protocol obtained may not reflect the best representation of the processes used to solve the problem. Obtaining verbalisations that best reflect the underlying cognitive processes is of course of paramount importance.

These concerns, however, have largely been centered upon non-verbal domains (Schooler et al., 1993). Several controlled experiments more directly evaluating the impact of 'thinking aloud' upon problem solving processes have provided converging evidence that verbal protocol does not affect the nature of processing underlying task performance (e.g., Brinkman, 1993; Fleck & Weisberg, 2004; Newell & Simon, 1972). Importantly, Ericsson & Simon (1993) argued that instructing participants to think aloud will not disrupt participants' thoughts, so long as only items naturally attended to by participants are verbalised and it is made explicitly clear to participants that they should concentrate on talking to themselves and should not at any point attempt to explain things to the experimenter.

Recent evaluations of verbal protocol methods have suggested that concurrent verbal protocol has advantages over retrospective verbal protocol (van Gog, Paas, van Merriënboer, & Witte, 2005), provided that verbal reports only reflect information directly used during problem solving that is either stored in short-term memory, or has an enduring trace in long-term memory (Patrick & James, 2004). Therefore, protocol will be collected in a concurrent manner, with a large emphasis placed within Experiment 3.3 upon ensuring participants receive adequate instruction concerning the think aloud method, with sufficient practice both using standard examples provided by Ericsson & Simon (1993), and within the context of the eight-puzzle-like BPST. In addition, all codings of verbal protocol will be conducted in tandem whilst viewing video recordings of the problem solving process with eye-tracker data overlaid.

Distinguishing between initial & concurrent planning Several studies have used protocol analysis to evaluate the effectiveness and nature of initial planning during a preparation period (e.g., Gilhooly, Phillips, Wynn, Logie, & Della Sala, 1999) and opportunistic planning during solution to everyday tasks (e.g., Hayes-Roth and Hayes-Roth, 1979). However, no study to date has attempted to use verbal protocol to distinguish between the deployment of initial and concurrent planning. Therefore, the use of verbal protocol analysis to evaluate the extent to which instances of initial and concurrent planning are used during solution to the eight-puzzle-like BPST will require a novel approach.

Ericsson (1975) has provided a useful framework for the analysis of verbal protocol collected concurrently during solution to the eight-puzzle. In short, Ericsson deemed the following five headings to adequately capture the different categories of verbalisations during solution to the eight puzzle: Plans, Intentions, Cognitions, Evaluations, and Move Descriptions. However, Ericsson provided no means of distinguishing between initial and concurrent planning styles. Although Ericsson's five headings will be used as a foundation for the current verbal protocol analysis, a number of modifications will be required in order to distinguish between initial and concurrent planning styles.

According to Davies (2003; 2004), initial and concurrent planning differ in the sense that initial planning involves strictly mental evaluation of the problem space and occurs prior to solution, whereas concurrent planning occurs 'on-line' during solution to a problem and thus includes external manipulation of the problem space. As in Experiment 3.2, the definition of initial planning employed here will differ slightly to that put forward by Davies (2003; 2004). The use of the term initial planning during analysis of the verbal protocol will not be limited to planning prior to making any moves, but will encompass all episodes of internal planning conducted whilst not making moves. Therefore initial planning will be characterised by segments of protocol coded as representing planning (Ericsson, 1975) that do not involve the movement of a single block. In contrast, concurrent planning will be characterised by segments of protocol coded as representing planning (following the same criteria) that include the movement of at least one block. This dichotomy between initial and concurrent planning will also be applied to Ericsson's (1975) conceptualisation of intentions (see Section 3.5.1 for precise category definitions).

Ericsson's (1975) definition of cognition will also be modified in order to successfully distinguish between initial and concurrent planning. Concurrent planning is said to rely more upon display-based interaction-intensive strategies, and thus perceptually available information should play an important role (Davies, 2004). Initial planning, on the other hand, is thought to rely more upon memory and thus be determined less by perceptually available information. For the current analysis, Ericsson's (1975) category entitled cognition will be changed to 'perceptual description' and the corresponding definition will be limited to the *description* of perceptually available information.

Regular and frequent evaluations of actions are also thought to be important during concurrent planning (Hayes-Roth & Hayes-Roth, 1979). In order to measure the extent to which participants evaluated their plans during the current task, Ericsson's (1975) rather non-specific category entitled evaluations will be modified so that by definition it only encompasses evaluations specific to a particular action (move/plan/intention). In fact, general evaluations regarding task performance not connected to a previous action will be excluded from this category, and will come under a new category entitled 'performance evaluation'. Ericsson's move description category was deemed suitable for the current analysis, as extended move descriptions may be representative of display-based problem solving. The number of moves made in each semantically related move description will also be recorded as this may reflect periods of trial and error behaviour.

3.5.1. Method

Participants Thirty-six Cardiff University undergraduate Psychology students participated in the study for course credit, and were randomly assigned to one of three IAC conditions.

Apparatus/Materials The experimental and recording equipment used were identical to those employed during Experiment 3.2, with the exception that the number of blocks residing within each window was reduced from sixteen to nine. Eight of these blocks were coloured, and one was empty (white). The windows remained the same size as in the two previous experiments. Therefore, the size of each block was increased proportionally. A wireless microphone was clipped to the item of clothing located nearest the participant's mouth in order to record the verbal protocol.

Design The manipulation of IAC on the Goal-State Window, visibility of the Current-State Window, the visual location of the two windows, and the method of moving blocks was identical to Experiment 3.2. However, because of the increase in task complexity, participants were limited to completing six different eight-puzzle-like BPST problems. In addition to the dependent variables measured in both previous experiments, the *Visual Basic* program was modified in order to record the number palindromic move sequences made, and the number of moves contained within each. A palindromic move sequence referred to any sequence of moves in which the second half of the sequence is the mirror image of the first (e.g., Red, Green, Yellow, Yellow, Green, Red). Such reflective moves sequences are seen in the planning literature as evidence of backtracking through the problem space (O'Hara & Payne, 1998), and thus may represent poor planning. The eye-tracker was also set up so as to allow the segmentation of eye fixation data by trial.

Categorisation scheme Verbal protocols were collected, transcribed and segregated into semantically related segments. Each segment was then coded according to the following category headings.

Initial Planning. The activity of plan development and consequence evaluation performed whilst not making any moves. Verbal statements considered to be plans must contain specific information about how the participant is going to attain the desired state or property. Words such as "if", "when" and "then" are useful indicators of verbal statements corresponding to planning activity.

Concurrent Planning. The activity of plan development and consequence evaluation performed whilst making a minimum of one move. Verbal statements considered to be plans must contain specific information about how the participant is going to attain the desired state or property. Words such as "if", "when" and "then" are useful indicators of verbal statements corresponding to planning activity.

Initial Intentions. Segments of protocol made whilst the participant was not making any moves which suggest that the participant was trying to attain a particular state or property, but without specific indication of how these states or properties are to be achieved. For example, "I need to get the green into the correct position".

Concurrent Intentions. Segments of protocol made whilst the participant was in the process of making a minimum of one move which suggest that the participant was trying to attain a particular state or property, but without any specific indication of how these states or properties are to be achieved. For example, "I need to get the yellow into the correct position".

Perceptual description. A description of perceptually available information. For example, "The green, yellow and red make up the bottom row". This description must not be connected to a previous plan or intention, but may act as a catalyst to the next action taken.

Action evaluation. An evaluation of a previous action. For example, "that worked", "I messed that up", "the green is now in the right place".

Performance evaluation. A general evaluation of one's task performance. For example, "this is really hard", "I think I may be near to finishing".

Move Description. A description of the move(s) a participant is just about to make, or is currently making. The description must match the moves recorded at that time.

Other. If a verbal statement appeared not to fall into either of the above eight categories, it was coded as 'other'.

Procedure Participants were first given an instruction sheet explaining the eight-puzzle-like BPST, and were then given verbal protocol instructions in line with Ericsson & Simon (1993) in order to familiarise themselves with the process of thinking aloud (see Appendix K). Participants were instructed that they were to verbalise every thought, that they were to think aloud constantly, and that they were not to plan what to say, or try to explain what they were saying. It was emphasised that the experimenter wanted the participant to act as if they were alone in the room, speaking to oneself. They were also informed that if they fell silent for long periods of time, the experimenter would prompt them to think aloud[4]. In accordance with Ericsson and Simon (1993), practice was given at thinking aloud by asking

[4] The experimenter only prompted participants to verbalise following approximately fifteen seconds of silence.

participants to think aloud whilst completing the following sum: 215 + 318. Having successfully completed the first example, participants were also required to again think aloud whilst working out how many rooms there were in their home.

Once participants were comfortable with the think aloud process, they were seated approximately 50 cm away from the eye-tracker and experienced a 16-point eye-tracker calibration. Following successful calibration, participants were required to complete a short block movement practice task that required the movement of each of the blocks in a sequence. This allowed participants in each of the IAC conditions to familiarise themselves with the method of moving blocks, accessing the Goal-State (dependent upon IAC), and also thinking aloud whilst moving the blocks. Different block configurations were used for each of the six experimental trials (see Appendix B). Each participant within each of the three conditions received one of twelve different randomised orders of trials.

3.5.2 Results & Discussion

The results for Experiment 3.3 will be split into two sections. Firstly, an overview of the effect of IAC upon problem solving strategy selection and performance will be provided in order to evaluate the extent to which problem solving search strategies and task performance mimicked those observed during previous BPST studies. Secondly, the results of the verbal protocol analysis will be presented in order to assess the qualitative impact of IAC upon problem solving and planning in particular. Only participants who successfully solved all six problems within a 60 minute limit were included in analyses. This resulted in the rejection of incomplete data from one participant in the Low IAC condition, five participants in the Medium IAC condition, and three participants from the High IAC condition. (The experiment ran until twelve participants within each IAC condition had completed all six trials within the time limit.)

Effect of IAC upon performance measures Before analysing the verbal protocol data, it is important to assess the extent to which behaviour observed during solution to the eight-puzzle-like BPST problems (see Table 3.4) was in line with that reported in Experiment 3.2, where an identical task structure was employed but less planning was required to complete each BPST problem. Therefore, a 3 (IAC Low/Medium/High) x 6 (Trial) ANOVA with the first factor manipulated between-subjects and the second factor manipulated within-subjects was conducted upon each of the dependent variables.

Overall, the effects of IAC upon strategy selection and performance were largely akin to the results reported in Experiment 3.2. No interactions were found between IAC and trial, and thus the relative effects of each will be summarised separately. Time-to-solution, Goal-State inspection frequency, and moves per Goal-State inspection data all required log transformation in order to attain homogeneity of variance. Greenhouse-Geisser corrected degrees of freedom are reported for variables that did not meet assumptions of sphericity.

Strategy selection. With regard to the effects of IAC upon the selection between memory- and interaction-intensive strategies, main effects of IAC were found for Goal-State inspection frequency, $F(2, 33) = 71.42, p < .001, MSE = 0.20$, and moves per Goal-State inspection, $F(2, 33) = 66.78, p < .001, MSE = 0.01$. As with previous experiments, participants again responded to an increase in IAC by choosing to access the Goal-State Window less frequently, and making considerably more moves per

Goal-State inspection. Bonferroni-corrected post hoc analyses revealed significant differences between each of the IAC conditions for both of these measures (*ps* $<.001$).

The effect of IAC upon the number of saccadic eye movements between the Goal- and Current-State when both were visible, $F(1, 22) = 26.63, p <.001, MSE = 5.12$, again suggests that participants in Experiment 3.3 made use of the facility to view both windows simultaneously to aid problem solving more as IAC increased. The fact that the values reported here are higher than those observed in Experiment 3.2 may reflect an increase in the complexity and necessity of planning required to complete the eight-puzzle-like BPST problems.

Performance measures. In line with the strategy data, performance on the eight-puzzle-like BPST was also affected by IAC in much the same way as observed using the technically identical, but less complex BPST in Experiment 3.2. In short, IAC did not affect, time-to-solution, $F(2, 33) = 1.87, p >.05, MSE = 0.18$, (Low IAC: *Mean* = 305.23, *SD* = 135.89; Medium IAC: *Mean* = 370.33, *SD* = 136.79; High IAC: *Mean* = 332.41, *SD* = 111.86), the total number of moves required to solve each of the problems, $F(2, 33) = 1.95, p >.05, MSE = 8880.95$, or the number of palindromic move sequences made, $F(2, 33) = 0.51, p >.05, MSE = 214.59$ (see Table 3.4).

However, again in line with Experiment 3.2, IAC did affect first-move latencies. A 3 (IAC Low/Medium/High) x 6 (trial) ANOVA with the first factor manipulated between-subjects and the second factor manipulated within-subjects revealed IAC to have a significant impact upon the length of first-move latencies, $F(2, 33) = 10.61, p <.001, MSE = 64.22$. Bonferroni-corrected post hoc analyses again revealed longer first-move latencies in the High than the Low IAC condition ($p <.001$), suggesting that participants may have been engaging in more initial planning prior to movement of the first block as a function of an increase in IAC (Davies, 2003).

Table 3.4

The Effect of IAC during Solution to a series of eight-puzzle-like BPST problems (Experiment 3.3).

	Low IAC		Medium IAC		High IAC	
Measure	Mean	SD	Mean	SD	Mean	SD
Goal-State inspection frequency	**84.04**	34.53	**48.54**	21.00	**13.00**	3.85
Moves per Goal-State inspection	**1.43**	0.60	**2.78**	1.27	**8.03**	3.65
Goal-State saccades per uncovering	**N/A**	N/A	**3.41**	1.43	**8.17**	2.86
First-move latency (s)	**11.72**	7.28	**19.75**	11.90	**26.78**	17.57
Palindromic sequences	**12.92**	4.06	**16.40**	5.96	**15.07**	6.84
Moves-to-solution	**92.81**	29.01	**107.92**	36.30	**100.47**	46.16

Note. Goal-State uncover time excludes lock-out delays incurred via High IAC.

Furthermore, a 3 (IAC) x 6 (category) ANOVA was computed upon the remaining log transformed inter-move latency data from the first trial, with the first factor manipulated between-subjects and the second factor manipulated within-subjects. The same six time-frame categories were used as those in Experiment 3.2, with the exception that the longest inter-move latency category had to be extended so as to accommodate latencies up to 100 s in duration. As in Experiment 3.2, an interaction between IAC and Category was found, F (3.94, 65.07) = 4.33, p <.01, MSE = 0.01, indicating that the inter-move latency categories were affected differentially by IAC (see Figure 3.8).

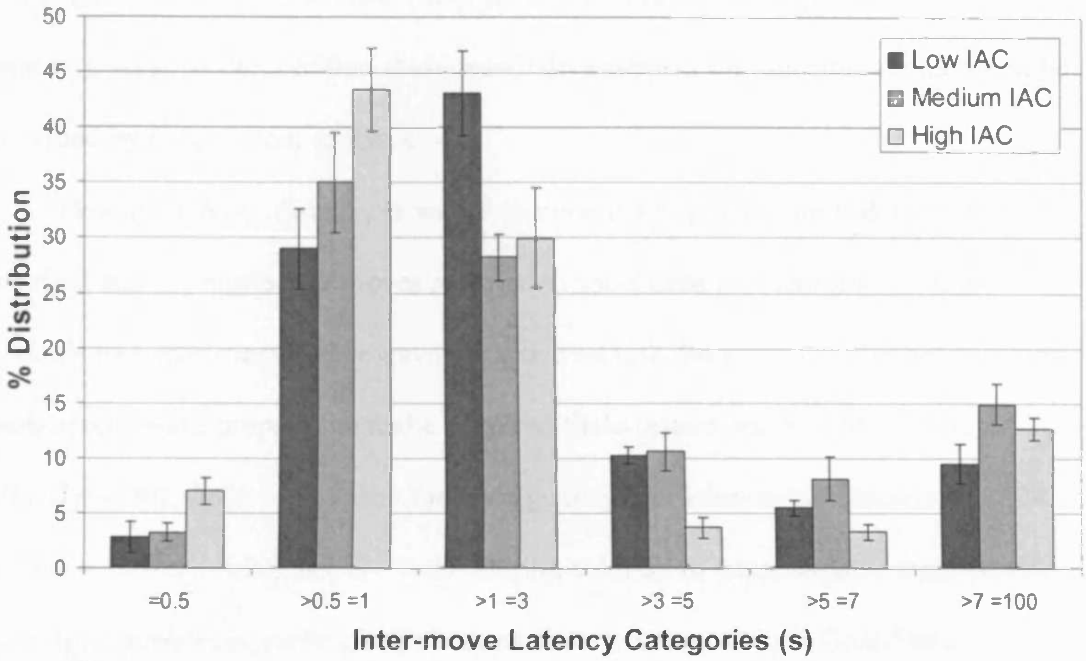


Figure 3.8: The distribution of inter-move latencies during solution to the first eight-puzzle-like BPST as a function of IAC (Experiment 3.3).

Note. First-move latencies were excluded from this analysis. Error bars represent ± 1 standard error.

Simple main effects analyses exploring the IAC x Category interaction revealed a similar pattern of data to that reported in Experiment 3.2. A higher proportion of inter-move latencies were found to come from Low than High IAC participants in the >1 =3 s and >3 =5 s categories (*ps* <.05), and significantly more =0.5 s inter-move latencies were found to come from High than Low IAC participants ($p < .05$). Although no significant effect of IAC was observed at the longest (>7 =100 s) inter-move latency category, F (2, 33) = 2.99, $p > .05$ $MSE = 0.01$, this non-significant effect was close to approaching the .05 alpha level ($p < .06$). Following from Davies's (2003) analysis of the initial and concurrent planning inter-move latency data, the current analyses again provide quantitative support for the assertion that concurrent planning is more commonplace in the Low IAC condition, whereas initial planning is promoted by higher access costs (throughout the duration of the problem solving process). As is the goal of this study, qualitative support for this interpretation will be provided by the protocol analysis.

Practice effects. Finally, as with Experiment 3.1, practice on task reduced both the time and the number of moves required to solve each problem (*ps* <.01). In contrast to Experiment 3.1, however, practice on task did affect the number of moves participants were prepared to make per Goal-State inspection, F (3.66, 120.62) = 10.84, p <.001, MSE = 0.01, and the average length of inter-move latencies, F (2.24, 49.34) = 15.02, p <.001, MSE = 0.20. During solution to six successive eight-puzzle-like BPST problems, participants chose to make more moves per Goal-State inspection, F (1, 33) = 23.32, p <.001, MSE = 0.01, and shorter inter-move latencies, F (1, 33) = 26.95, p <.001, MSE = 0.21. This may suggest the development of a generic encoding scheme for Goal-State information over time, and/or the development of superior problem solving schemata as a function of practice (see Chapter 4 for further investigation of this topic with respect to IAC). First-move latencies were not affected by trial, F (3.00, 98.87) = 0.58, p >.05, MSE = 211.54, suggesting that any differences in the nature of planning (contingent upon IAC) prevailed across the six trials.

Effect of IAC on types of verbalisation Given that some participants verbalised little, and together with some recording failures, it was decided to sample four participants protocols during the first and final trial from each of the three IAC conditions (cf. Davies, 2003). An independent blind coder also second coded half of these transcriptions, and an inter-coder reliability score of 93.77% was obtained as a simple percentage agreement between the two coders. The proportion of agreement after chance had been excluded was 93%, kappa (N = 289) = .93, p <.001, (Cohen, 1960). The two coders discussed each of the resultant discrepancies and subsequently came to an agreement concerning each.

Although a MANOVA statistical test may have provided particular insight into the inter-relationships between each of the verbal protocol categories, a number of moderate positive correlations between categories and relatively small sample sizes within categories suggested that MANOVA analysis would lack considerable power (Ramsey, 1982; Tabachnick & Fidell, 2001). Therefore, a series of separate ANOVAs were computed upon log transformed frequency data (required to meet assumptions of homogeneity of variance), and correlation analysis was used to assess the interrelationships between categories. A 3 (IAC Low/Medium/High) x 2 (Trial First/Final) ANOVA was computed upon the frequency data for each of the categories contained within Table 3.5, with the first factor manipulated between-subjects and the second factor manipulated within-subjects.

Table 3.5

The Effect of IAC upon Types of Verbalisation (Experiment 3.3).

	Low IAC		Medium IAC		High IAC	
Measure	Mean	SD	Mean	SD	Mean	SD
Initial plan	**0.25**	0.46	**1.25**	1.04	**1.63**	1.85
	(0.45)	(0.85)	(4.79)	(3.86)	(11.90)	(7.96)
Concurrent plan	**3.63**	2.45	**2.00**	1.69	**0.63**	1.06
	(15.05)	(11.20)	(9.14)	(9.16)	(5.74)	(8.78)
Initial intention	**1.63**	1.77	**3.88**	2.47	**3.25**	1.39
	(5.69)	(5.50)	(16.24)	(7.71)	(28.17)	(10.28)
Concurrent intention	**5.75**	3.92	**3.75**	3.54	**1.13**	0.99
	(22.07)	(8.85)	(14.34)	(9.49)	(9.76)	(10.71)
Perceptual description	**1.25**	1.28	**2.00**	1.31	**0.50**	0.53
	(4.43)	(4.07)	(8.71)	(5.14)	(5.50)	(7.59)
Action evaluation	**5.88**	4.42	**4.88**	2.95	**3.00**	3.21
	(20.56)	(7.07)	(19.94)	(7.30)	(19.13)	(16.22)
Performance evaluation	**1.25**	2.55	**1.00**	1.69	**0.38**	0.52
	(2.11)	(4.07)	(3.83)	(6.59)	(2.64)	(4.55)
Move description	**7.88**	6.29	**5.50**	5.01	**2.75**	5.47
	(28.66)	(13.47)	(20.34)	(12.08)	(10.25)	(13.76)
Other	**0.50**	1.07	**0.63**	0.74	**1.00**	1.41
	(0.97)	(2.20)	(2.68)	(4.07)	(6.92)	(11.33)

Note. Values represent mean frequency per trial (with proportional data provided within parentheses).

Initial and concurrent planning. As hypothesised, IAC significantly affected the frequency with which verbalisations were coded as belonging to the initial planning category, F (2, 9) = 5.19, p <.05, MSE = 0.03, with Bonferroni-corrected post hoc analyses revealing significantly more initial planning in the High than the Low IAC condition (p <.05). In addition, initial planning was significantly more prevalent during solution to the first trial (*Mean* = 1.58, *SD* = 1.62), than the final, (*Mean* = 0.50, *SD* = 0.67), F (1, 9) = 9.25, p <.05, MSE = 0.03. However, there was no interaction between IAC and trial, F (2, 9) = 0.56, p >.05, MSE = 0.02. An example of an initial plan formed by a participant in the High IAC condition is provided below. After three minutes of largely unsuccessful problem solving during the first trial, this particular participant decided to stop making moves in order to assess the situation and view both the Goal- and the Current-State simultaneously for the sixth time. The following verbalisation represents the subsequent attempt to rectify the situation via an episode of initial planning (during which no moves were made):

> *"Ok, so if I move the blue across, the purple has to go down, and get the red in the right position, black and pink are wrong, so if I try getting the red first."*

IAC also affected the frequency with which verbalisations were coded as representative of concurrent planning, F (2, 9) = 4.77, p <.05, MSE = 0.08. Again as hypothesised, Bonferroni-corrected post hoc analyses revealed concurrent planning to be higher in the Low than the High IAC condition (p <.05). No effect of trial was observed, F (1, 9) = 0.44, p >.05, MSE = 0.07, nor was an interaction found between IAC and trial, F (2, 9) = 0.34, p >.05, MSE = 0.07. An example of a concurrent plan formed by a participant in the Low IAC condition is provided below. This section of protocol is extracted following just thirty seconds of problem solving on the first trial. Fast and frequent saccadic eye movements between the Goal- and the Current-State have dominated problem solving thus far, and during the protocol below five moves were made representing concurrent planning.

> "*Move the yellow and the green to put the brown in its place, and the pink over and the purple up to swap the black and the pink around.*"

Initial and concurrent intentions. No significant effect of IAC was observed upon the frequency with which verbalisations were coded as representing initial intentions, F (2, 9) = 3.81, p >.05, MSE = 0.06. Similarly, no effect of trial was witnessed, F (1, 9) = 3.12, p >.05, MSE = 0.04, nor was an interaction found between IAC and trial, F (2, 9) = 0.07, p >.05, MSE = 0.04. It is important to note that any intentions contained within a segment of protocol coded as representing a plan were not also counted as an intention. Only intentions verbalised without an associated plan were coded as intentions. Therefore, it is possible that different results would have been observed if examples of intentions imbedded within a plan had also been coded as an intention. An example of an initial intention not attached to a plan formed by a participant in the High IAC condition is provided below. This section of protocol follows two minutes of careful problem solving, whereby the participant inspects the Goal-State on average once every ten moves (three times thus far in total). Upon cessation of a set of related moves, the participant simultaneously views the Goal- and Current-State, evaluates the previous set of moves, and forms the following initial intention:

> "*I'll try and get the black and the purple to where they need to be.*"

IAC did, however, influence the frequency with which verbalisations were coded as reflecting concurrent intentions, F (2, 9) = 5.66, p <.05, MSE = 0.08. Bonferroni-corrected post hoc comparisons revealed concurrent intentions to be more common in the Low than the High IAC condition (p <.05). Again, no effect of trial was observed, F (1, 9) = 1.27, p >.05, MSE = 0.06, nor was an interaction found between IAC and trial, F (2, 9) = 1.26, p >.05, MSE = 0.06. An example of a concurrent intention formed by a participant in the Low IAC condition is provided below. Joining the participant twenty seconds into the problem solving process for the final trial, following frequent saccadic eye movements between the Goal- and the Current-State and ten moves thus far, the participant makes the following statement (whilst making a further five moves in tandem):

"I need to, ummm, swap the black round and get the pink down to the bottom."

Perceptual descriptions and action evaluations. No support was found for the prediction that perceptual descriptions would be more commonplace when access costs were Low, F (2, 9) = 3.42, p >.05, MSE = 0.05. Again, no effect of trial was witnessed, F (1, 9) = 0.22, p >.05, MSE = 0.06, nor was an interaction found between IAC and trial, F (2, 9) = 0.28, p >.05, MSE = 0.06. An example of a perceptual description formed by a participant in the Medium IAC condition is provided below. Following seventeen and a half minutes of problem solving on the first trial, the participant acknowledged that they had reached an impasse. They chose to cease making moves for ten seconds and simultaneously view the Goal- and the Current-

State upon uncovering the Goal-State for the two hundredth time. The following perceptual description was made:

"So, the order goes black, green, yellow, pink."

Support was found for the hypothesis that frequent interleaving of action evaluations would be reported in the Low IAC condition (indicative of concurrent planning), but not the High. A main effect of IAC was observed, $F\ (2, 9) = 6.18, p < .05, MSE = 0.03$, with Bonferroni-corrected post hoc analyses revealing significantly more action evaluations in the Low than the High IAC condition ($p < .05$). No effect of trial was observed, $F\ (1, 9) = 3.86, p > .05, MSE = 0.11$, nor was an interaction found between IAC and trial, $F\ (2, 9) = 0.40, p > .05, MSE = 0.11$. Although the frequency data provides support for the assertion that verbalisations coded as action evaluations were more commonplace when problem solving with Low compared to High IAC, the relatively similar proportional means obtained for each IAC conditions (see values in parenthesis in Table 3.5) does suggest that this interpretation should be taken with caution. An example of an action evaluation made by a participant in the Medium IAC condition is provided below. Having made a series of six moves, one minute and twenty seconds into the problem solving process of the final trial, the participant uncovered the Goal-State for the fifteenth time and evaluated the outcome of the previous set of moves:

"So that's that bit done."

Performance evaluations, move descriptions and other. No difference was observed between IAC conditions with regards to the frequency with which participants verbalised segments of protocol were coded as performance evaluations, F (2, 9) = 0.25, p >.05, MSE = 0.08, although a reduction was observed as a function of trial, F (1, 9) = 7.83, p <.05, MSE = 0.04, (First Trial: *Mean* = 1.58, SD = 2.27; Final Trial: *Mean* = 0.17, SD = 0.39). No interaction was observed between IAC and trial, F (2, 9) = 1.05, p >.05, MSE = 0.04.

IAC significantly affected frequency with which verbalisations were coded as belonging to the move description category, F (2, 9) = 9.36, p <.01, MSE = 0.06, with Bonferroni-corrected post hoc analyses revealing a lower number of move descriptions recorded in the High than both the Low and Medium IAC conditions (*ps* <.05). Move descriptions were also found to reduce as a function of trial, F (1, 9) = 20.20, p <.001, MSE = 0.07, (First Trial: *Mean* = 3.75, SD = 1.95; Final Trial: *Mean* = 2.39, SD = 3.14), but no interaction was observed between IAC and trial, F (2, 9) = 0.90, p >.05, MSE = 0.07. A number of alternative explanations could be put forward for the fact that move descriptions were more commonplace in the Low than the High IAC condition. Firstly, the increased number of move descriptions observed in the Low IAC condition may reflect a tendency for participants in this condition to engage in more display-based problem solving. Alternatively, participants in the Low IAC condition may have found the eight-puzzle-like BPST harder to solve, and thus exhibited more trial and error behaviour, and more move descriptions. However, the lack of an effect of IAC upon moves-to-solution data (see above) suggests that differences in trial and error behaviour would be unlikely to account for this effect. The observation that IAC also affected the average number of moves verbalised within each move description, F (2, 9) = 6.18, p <.05, MSE = 0.05, and that these values were significantly higher in the Low than the High IAC condition ($p < .05$), gives weight to the possibility that participants in the Low IAC condition were more comfortable describing the moves they made (as determined by Bonferroni-corrected post hoc analyses). No effects of IAC, $F(2, 9) = 0.43, p > .05, MSE = 0.05$, nor trial, $F(1, 9) = 3.85, p > .05, MSE = 0.06$, were observed for the number of semantically related segments of protocol coded as belonging to the 'other' category.

Correlations. Pearson's *r* correlation analysis was also conducted upon the transformed frequency data in order to examine the presence of any interrelationships between each of the verbal protocol categories. Significant interrelationships observed indicated that the frequency with which concurrent intentions were verbalised significantly correlated with both move description and action evaluation frequency during problem solution to the first problem (+.59 and + .65, *ps* <.05). Correlations between action evaluation frequency and move description and action evaluation frequency and performance evaluation were also found during problem solving on the first trial (+.90, and +.58, *ps* <.05). During solution to the final trial, a significant correlation was also found between the number of moves contained within a move description and the frequency with which move descriptions were verbalised (+.83, $p < .001$). Non-significant correlations approaching the .05 alpha level were also found between the number of moves contained within a move description and both the frequency of concurrent and initial planning (+.55 and -.57 respectively, *ps* <.06), and between initial and concurrent planning themselves (-.56, $p < .06$).

High & Low IAC protocol examples. Before concluding and summarising the effects observed in Experiment 3.3, two excerpts of protocol are provided to represent characteristic differences between Low and High IAC participants over an extended period of time.

Table 3.6

An Extract of Low IAC Verbal Protocol (Experiment 3.3).

Time (s)	Statement	Moves	Fixation	Category
0.00-0.25	Right then, so erm, right, it's messy.	0	Both	Performance evaluation
0.27-0.44	If I bring the red one down, and then I need to get, if I get the green out the way, get the yellow one up.	3	Both	Move description
0.45-0.48	Erm, then get the green into its space.	2	Current-State	Concurrent intention
0.50-0.55	Ah, oh that's not going to work.	0	Both	Action evaluation
1.00-1.03	move the green one back and move the red one	2	Both	Move description
1.06-1.08	Try and get the red one in its space.	1	Current-State	Concurrent intention
1.09-1.26	Start with the red one. Then if I move the green one down into its space, and if I try and get the pink one, ah, er, into its space as well.	3	Both	Concurrent planning
1.29-1.30	Erm, brown.	1	Current-State	Move description
1.33-1.35	And then the blue one's going too far, er.	0	Current-State	Action evaluation
1.40-1.43	I need to swap the blue and the yellow.	0	Both	Initial intention
1.44-1.48	Erm, problem, ah.	0	Both	Action evaluation
1.57-2.03	Erm, ok having a blank, erm.	0	Both	Other
2.03-2.15	Ok, so I need to get the yellow one down and put that one next to the brown one.	5	Both	Concurrent intention
2.18-2.21	Erm, and then try and get the red one down as well.	2	Current-State	Concurrent intention
2.26-2.37	Right, erm, right, erm, problem.	2	Both	Action evaluation
2.38-2.42	Yellow one, right lets try another one	0	Both	Initial intention
2.43-2.45	So brown's in the right place, purple is in the right place.	0	Both	Perceptual description
2.45-2.56	Get the pink, erm, get the pink in the right place.	2	Both	Concurrent intention
3.00-3.04	Right, that's not going to work either.	0	Current-State	Action evaluation

Table 3.6 contains an extract of verbal protocol data from a Low IAC participant solving their first problem. Typically, the Low IAC participant begins to move blocks without having first instantiated a plan or intention. The solution process then quickly begins to involve the formulation of plans, intentions and regular action evaluations. In particular, the action evaluations illustrated in Table 3.6 appear typical of opportunistic planning (Hayes-Roth & Hayes-Roth, 1979), in the sense that they often stop a particular action upon realisation that it is not going to achieve the desired outcome. Although the particular excerpt provided in Table 3.6 does not include long episodes of move descriptions (often typical of Low IAC problem solving), the predominance of concurrent planning and intentions (as opposed to the initial variety), and the regular interleaving of action evaluations is argued to act as a good representation of problem solving with Low IAC. This particular participant recorded one episode of initial planning, six counts of concurrent planning, five initial intentions and fourteen concurrent intentions. In addition, three perceptual descriptions, fifteen action evaluations, and fourteen move descriptions were recorded.

Table 3.7 provides an example of protocol extracted from the verbalisations of a representative participant in the High IAC condition. In contrast to the Low IAC extract in Table 3.6, the High IAC protocol includes the formulation of an initial plan prior to making any moves. The corresponding moves are then made, and subsequently evaluated with regard to the prior intention. The participant then repeats this pattern of problem solving several times, and it is argued that this extract provides the reader with a good example of what is meant by the inclusion of several 'episodes' of initial planning. This particular participant went on to record six episodes of initial planning, no instances of concurrent planning, four initial intentions, and one concurrent intention. In addition, one perceptual description was recorded, along with ten action evaluations, and sixteen move descriptions.

Table 3.7

An Extract of High IAC Verbal Protocol (Experiment 3.3).

Time (s)	Semantically-related Statement	Moves	Fixation	Category
0.00-0.29	So, I'm going to move the black across, the brown down and the yellow across to start off with.	0	Both	Initial intention
0.35-0.54	Hang on a second, oh this is so confusing. Ok so I'm going to get, ok, ok, brown, I'll get brown into the top corner.	0	Both	Initial intention
0.54-1.02	Brown goes up, green across, yellow down, brown across.	4	Current-State	Move description
1.05-1.07	That's got that in the right place.	0	Both	Action evaluation
1.07-1.17	Now I need to get the yellow and the green out of there, so I'm going to move the yellow.	0	Both	Initial plan
1.18-1.31	Green up, yellow across, black up, purple right, blue right, pink down, red down, green across, yellow up.	9	Current-State	Move description
1.32-1.41	Is that the right way round? That's why, the bottom is messed up.	0	Both	Action evaluation
1.41-1.49	Need to get the pink to the other side, move the whole middle ones around.	0	Both	Initial plan
1.50-1.59	Black across, purple up, blue across, pink across, down black, across, purple across again.	7	Current-State	Move description
2.00-2.02	So pink's in the right place now.	1	Current-State	Action evaluation

Overall, the verbal protocol analyses supported the main hypothesis that small increments in IAC would encourage initial planning, yet discourage concurrent planning. Additional support was also provided for the contention that concurrent planning would be more common when access costs were Low. It is argued that soft constraints, not hard constraints, explain these findings (Gray & Fu, 2004); that is, the small changes to information accessibility within each of the IAC conditions did not force participants to select certain problem solving search strategies, but rather, encouraged the selection of some above others. Although it is thought that the distribution of memory based upon the conservation of time (Gray et al., 2006) remains largely responsible for determining the outcome of such least-effort tradeoffs, Experiment 3.3 demonstrated that different forms of planning can also be promoted by small changes to the cost associated with accessing task-relevant information during problem solving.

3.6. General Discussion

The three experiments contained within Chapter 3 set out to explore the manner by which small changes to the cost associated with accessing information from an interface could affect the selection between candidate problem solving search strategies. Firstly, based upon work investigating the adaptive nature of memory as a function of IAC during routine copying (Fu & Gray, 2000; Gray et al., 2006), it was anticipated that small changes to IAC would affect the use of memory during problem solving. Secondly, based upon previous work detailing the adaptive nature of planning during problem solving (O'Hara & Payne, 1998; 1999), it was hypothesised that small changes to IAC may also have consequences for planning behaviour. Support for both assertions was found.

Chapter 3 demonstrated that relatively small increments to the cost associated with accessing task-relevant information from the interface led to a stronger reliance upon memory during problem solving, and influenced the selection between initial and concurrent planning styles (Davies, 2003; 2004). The main results from each experiment will be summarised, and conclusions will be set in context with respect to the information access cost, memory and planning literatures. The applied consequences for display design will also be discussed.

3.6.1. The adaptive use of memory and planning as a function of IAC during problem solving

The primary aim of Chapter 3 was to assess the extent to which the selection between interaction- and memory-intensive strategies based upon changes in IAC previously observed during routine copying (Fu & Gray, 2000; Gray et al., 2006) would extend to a problem solving task. To this end, the BPST developed for Experiment 3.1 was designed so that the task parameters were similar to the BWT, but required more cognitively demanding interaction. Emphasis was placed upon finding the most efficient solution through the problem space, rather than simply copying a pattern. Thus, participants solving the BPST were required to evaluate move sequences for their costs and benefits, and consider the relations between several different components of the task space.

The results of all three experiments (3.1 - 3.3) suggest that during solution to a series of BPST problems, the selection between interaction- and memory-intensive strategies contingent upon IAC was governed in much the same way as previously observed during routine copying (Fu & Gray, 2000; Gray et al., 2006). Regardless of the complexity of planning required to complete each of the three versions of the BPST employed, small increments in IAC uniformly led to an increase in reliance upon memory during problem solving. When access costs were Low, participants were reluctant to make more than two moves per Goal-State inspection. However, when an additional small mouse movement and 2.5 s delay was associated with inspection of the Goal-State, participants chose to make in excess of eight moves per Goal-State inspection. Thus, further support was derived for Anderson and colleague's (Anderson & Milson, 1989; Anderson & Schooler, 1991; 2000) perspective of the adaptive nature of human memory. Seemingly, small changes to the design of the interface (the kind likely to be affected by integrated display design and information fusion) can have surprisingly large consequences for the deployment of human memory during interactive problem solving behaviour.

Despite the substantial influence of IAC upon the use of memory during problem solving, small changes to the accessibility of information provided within the interface generally had little effect upon problem solving proficiency (with the exception of a small but significant rise in the number of moves required to solve each BPST in Experiment 3.1 when IAC was High). Provided participants had the option to simultaneously view the Goal- and the Current-State when not making moves (Experiments 3.2, 3.3), small manipulations to the cost of accessing the Goal-State made no difference to the number of moves required to solve each BPST problem (although time-to-solution did increase as a function of IAC).

These discrepancies between the first and latter experiments in terms of the effect of IAC upon problem solving proficiency can be accounted for by previous work from two domains. Firstly, problem solving performance has been observed to degrade if the demands placed upon working memory are too high (Kotovsky et al., 1985; Phillips et al., 1999). Secondly, research reviewing the importance of spatial and temporal congruency within interface design has observed that, if externally represented information is to be integrated from multiple sources, performance will improve if the user is able to view all information to-be-integrated with high temporal and spatial congruency (Wickens & Hollands, 2000).

The current investigation, however, indicates that as long as problem solvers are provided with the facility to view all information to-be-integrated when they wish, small costs associated with performing this action can direct the adaptive nature of human cognition to encourage the selection of certain microstrategies, yet discourage the selection of others, without affecting overall problem solving performance. The application of this finding (and associated limitations) will be considered in more detail in Section 3.6.2. Firstly, however, the influence small changes to access cost can have upon the selection between initial and concurrent planning styles (Davies, 2003; 2004) will be summarised.

In stark contrast to the small but significant rise in the number of moves required to solve each BPST problem observed as IAC increased in Experiment 3.1, increasing the cost of making each move (IC) from pressing an arrow key to typing a string of letters significantly reduced the number of moves required to solve each BPST problem. Intuitively, it is argued that the positioning of the cost within the task is likely to explain the opposing effect of IAC and IC upon moves-to-solution. IAC will, by definition, be experienced only when the problem solver chooses to inspect the Goal-State. IC, on the other hand, will be experienced every time a participant chooses to make a move, thus magnifying its potential influence.

When comparing the effects of IAC and IC upon problem solving efficiency, it is important to note that increments in IAC led to an increase in reliance upon memory, whereas increments in IC led to a reduction in reliance upon memory during problem solving. According to Davies (2003; 2004), initial planning is largely memory-dependent, whereas concurrent planning is largely display-based. One interpretation of the results of Experiment 3.1, therefore, could be that increments to IAC promoted an initial planning style, whereas increments to IC encouraged a concurrent planning style. Indeed, the observation that inter-move latencies increased significantly as a function of IC, yet were not affected by IAC, further suggests that the two costs may induce different forms of planning.

In an attempt to explore the hypothesis that increasing IAC may promote initial, yet discourage concurrent planning, Experiment 3.2 provided an in-depth quantitative analysis of participants' inter-move latencies during solution to a modified BPST, designed to reduce the perceived cost of planning. Davies (2003) argued that relatively long and short inter-move latencies represent initial planning, with long latencies representing mental evaluation of the task space and short latencies representing pre-planned moves. In contrast, Davies (2003) proposed regular 'medium-sized' inter-move latencies to reflect periods of concurrent planning.

In support of the hypothesis that small increments to IAC would promote initial planning, first-move latencies were found to increase significantly as a function of IAC (Davies, 2003). Analysis of the distribution of the remainder of inter-move latencies during solution to the first BPST problem also indicated that the longest and shortest inter-move latencies (=0.5 s and >7 =30 s respectively) were significantly more commonplace in the High than the Low IAC condition, whereas medium-sized inter-move latencies (>1 =5 s) were more common in the Low than the High IAC condition. It is argued that initial planning is not limited to planning prior to solution, but that an extension is required to Davies's (2003) original conceptualisation of initial planning, acknowledging that initial planning may also take place in 'episodes' during solution to a problem.

Further qualitative support was gathered in Experiment 3.3 for the hypothesis that higher access costs would promote initial planning, whereas lower access costs would encourage concurrent planning. Using an eight-puzzle-like BPST, concurrent verbal protocol (Ericsson & Simon, 1993) was collected and coded according to modified versions of Ericsson's (1975) categories. Protocol coded as representing planning was further subdivided into either an initial category if no moves were made during the verbalisation, or a concurrent category if a minimum of one move was made during the verbalisation. Corroborating the analysis of inter-move latencies in Experiment 3.2, analyses of verbal protocols also revealed initial planning to be more commonplace when access costs were High, and concurrent planning to be more commonplace when access costs were Low. The novel coding scheme developed in Experiment 3.3 for distinguishing between initial and concurrent planning is argued to be an effective tool. Despite Davies's (2004) concern that the processes underlying opportunistic concurrent problem solving may not be as amenable to verbalisation as more goal-directed behaviour, the novel coding scheme appeared able to capture periods of initial and concurrent planning.

Overall, the three experiments contained within Chapter 3 have extended work conducted by Gray and colleagues (Fu & Gray, 2000; Gray et al., 2006, Gray & Fu, 2004) by demonstrating that the effect of IAC upon the selection between candidate microstrategies is not limited to low-level routine interactive behaviour, but can also influence adaptive interactive behaviour during problem solving. With regards to the planning literature, the results from Chapter 3 suggest that Davies's (2003; 2004) recent distinction between initial and concurrent planning is a useful one that can be used effectively to distinguish between two different, yet equally important forms of planning. It is argued that the novel categories developed for the distinction between initial and concurrent planning during the coding of verbal protocol collected in Experiment 3.3 were effective, and could be applied to an array of problem solving paradigms situated within the context of an external display. It also is argued that manipulating IAC in a number of problem solving contexts may prove a useful methodology for further uncovering the intertwined relationship between planning and memory.

3.6.2. Application to cognitive engineering

The application of these results to the field of cognitive engineering may provide display designers with the facility to build soft 'IAC-based' constraints into the interface in order to encourage operators to select certain memory- or display-based problem solving strategies when required. Compared to principles of simply increasing the accessibility of information provided within an interface (see Chapter 1), it is argued that the novel approach proposed herein may prove more fruitful when attempting to circumvent the limited capacity of the cognitive system (Simon, 1981) in order to reduce information overload (Kirsh, 2000). It is argued that a full appreciation of the effects of the constraints of display design upon the psychological processes selected to deal with particular design details is required in order to exploit the link between interface design and operator behaviour.

Different applied situations will require different methods of interaction with the interface, and it is proposed here that the intelligent use of IAC may go some way to orienting behaviour towards the selection of appropriate microstrategies. As demonstrated in Experiments 3.2 and 3.3, the selection between concurrent and initial planning styles can be heavily influenced by small changes to IAC. Although very few differences in terms of problem solving proficiency were observed contingent upon the selection between these two forms of planning within the current experiments, other research has begun to assess their costs and benefits in different situations. For example, Davies (2004) noted that concurrent planning is generally seen to reduce load upon working memory, and thus may be useful during times of high operator workload and task complexity. However, manipulation of the external display as a means of evaluating the efficiency of various alternatives (Kirsh & Maglio, 1994; Neth & Payne, 2002) is common during concurrent planning, and thus mistakes may be more frequent and backtracking through the problem space may be difficult. Although initial planning clearly imposes high demands upon working memory, the use of internal memory to evaluate different alternatives (without the implementation of moves) may lead to more accurate problem solving, particularly during the solution of easier problems (Davies, 2003).

There are a number of related issues that may impact upon the results reported herein. For example, the general observation made within Chapter 3 that higher access costs promote an initial planning style, whereas lower access costs promote a concurrent planning style, may be affected further by interactions with problem complexity (Naylor & Briggs, 1963) and particular task demands (Davies, 2003). Further work is needed to investigate these potentially important moderating factors. Although the effect of IAC upon the adaptive use of memory appears fairly robust across routine copying and more complex problem solving tasks, more research is required to assess under what conditions this may change.

Participant compliance and technical factors severely limited the number of verbal protocols available for coding, and thus, some caution must be taken when generalising the results from this relatively small sample to the general population. Furthermore, it is difficult to predict how the effects observed using students would generalise to highly trained operators of complex interface display units, such as those found in the modern aircraft cockpit. How would domain-specific knowledge, procedural instructions and motivational elements affect the results reported herein?

Although it would be theoretically interesting to extend the reported investigation to examine such questions, they are beyond the scope of the current thesis. As would be a more detailed investigation into the effect of IAC upon more general approaches to problem solving (Ericsson, 1975; Simon, 1975). Indeed, exploration into the precise nature of the IAC least-effort tradeoffs (see Gray et al., 2006 for a time-based account) is also a restriction on scope, but remains an avenue for future research. Instead, the focus of the next Chapter will be upon the effect of IAC on facilitating learning during repeated solution of an eight-puzzle-like BPST problem.

Chapter 4

EMPIRICAL SERIES 3

Learning solution to the eight-puzzle with varying goal-state accessibility

---

4.1. Introduction

The manner by which differences in the accessibility of task-relevant information influence the use of memory and the nature of planning during problem solving (see Chapter 3) poses the possibility that changes in Information Access Cost (IAC) may also lead to differential rates of learning under certain conditions. In support of a number of theories of learning by doing (see Anzai & Simon, 1979; Glaser & Bassok, 1989; Matthews, Davies, Westerman, & Stammers, 2000 for useful reviews), problem solving efficiency in Experiments 3.1 and 3.3 was seen to improve as a function of practice. However, Chapter 3 found no interaction between IAC and practice when participants were required to solve a series of *different* eight-puzzle-like Blocks Problem Solving Task (BPST) problems. Chapter 4 will therefore attempt to investigate the contribution, if any, of IAC to learning during *repeated* solution of an eight-puzzle-like BPST problem. As with Chapter 3, participants will not be instructed to learn during task completion. Therefore, any effects of learning observed will be incidental by nature (Postman, 1964), and not necessarily intentional.

4.1.1. Accumulative learning of the same problem

Firstly, it is important to formulate reasons why the effect of IAC on learning may differ between conditions where participants are required to solve a series of different problems, and conditions where participants are required to solve the same problem repeatedly. By definition, the number of identical elements between trials will be higher (Thorndike, 1903), and the cognitive processes required to solve the same problem will be more similar (Anderson, 1987), when the same problem is to be solved repeatedly than when a series of different problems are to be solved. Therefore, differing learning opportunities would be predicted when a participant is required to solve the same eight-puzzle-like BPST problem repeatedly, than when a participant is required to solve a series of different eight-puzzle-like BPST problems.

Importantly, when solving a problem repeatedly there will be greater opportunity to encode and use context-specific memory representations of earlier problem solving episodes to inform future problem solving (Atwood & Polson, 1976; Jeffries, Polson, Razran, & Atwood, 1977; Ross, 1984; 1987; 1989). Participants will be able to 'hone in' upon specific problem solving solutions, and will not need to develop higher-level general heuristics that would be required when solving a variety of problems that are similar in their structure, but different in terms of their solution sequences. Although the use of memory to inform repeated searches of the same problem space is not well established, familiarity of previous problem solving states has been found to guide problem solving under such conditions (Payne, Richardson, & Howes, 2000).

4.1.2. Display-based problem solving & learning

The development of cognitive skill has been characterised in many domains as a progression from problem solving search to memory-based routine skill (e.g., Card, Moran, & Newell, 1983; Logan, 1988; Newell, 1990). Indeed, Anderson's (1982; 1987) conceptualisation of 'compilation', argued to be integral to skill acquisition, is defined as movement away from the interpretive application of declarative knowledge towards specific procedures that directly apply such knowledge from memory. Such accounts of learning during problem solving are consistent with theories of situated cognition (e.g., Brown, Collins, & Duguid, 1989) that emphasise the use of contextual cues to learning. Importantly, the extent to which move sequences are evaluated externally (Kirsh & Maglio, 1994) versus internally (O'Hara & Payne, 1998) has been found to partially determine the efficiency with which such learning mechanisms may operate during solution to an eight-puzzle problem (see Sweller, 1988 for a review within the 'cognitive load' framework).

Display-based problem solving can often be used as an effective substitute for internally derived problem solving (see Chapter 3), and can offer certain advantages over memory-based methods that are typically more error-prone (Larkin, 1989). However, some work in the field of human-computer interaction has indicated that even expert users of display-based tasks cannot necessarily recall how to perform routine activities when assessed without consultation of the interface (Mayes, Draper, McGregor, & Oatley, 1988; Payne, 1991). Display-based strategies often rely heavily upon context-dependent recognition during interaction with the interface, and this under certain situations, can limit the encoding and learning observed during problem solving (see Chapter 2 for a similar discussion with regard to 'transfer-appropriate processing' – Morris, Bransford, & Franks, 1977).

Here, the suggestion is made that lower access costs may limit learning during repeated solution to an eight-puzzle-like BPST problem due to over-reliance on the display as an external memory source (O'Regan, 1992). Higher access costs experienced during repeated solution of an eight-puzzle-like BPST problem, on the other hand, may facilitate learning due to a stronger reliance upon memory during problem solving.

4.1.3. IAC as 'germane cognitive load'?

Three variants of cognitive load have been argued to influence general learning during problem solving (for recent reviews see Ginns, 2006; van Merriënboer, Schuurman, de Croock, & Paas, 2002). Firstly, the term 'intrinsic cognitive load' has been used to describe the inherent complexity or element interactivity of a task, given an individual's prior knowledge relating to that task (Sweller & Chandler, 1994). Secondly, 'extrinsic cognitive load' has been used to describe the resources spent completing a task that are not related to schema development. Finally, and of most relevance to the current discussion, is the recent conceptualisation of 'germane cognitive load' (van Merriënboer et al, 2002) – a term used to represent mental effort beneficial to learning that can be affected by the task design and structure. Perhaps small increments in IAC could be used as a method for introducing germane cognitive load for learning to problem solve, due to their effects upon promoting the regular use of memory during problem solving?

As discussed previously, a number of studies have emphasised the benefits of using memory to learning during problem solving (e.g., Chase & Simon, 1973; De Groot, 1966; Delaney, Ericsson, & Knowles, 2004; Payne et al., 2000; Ross, 1987), and concern has been raised in the literature that over-reliance upon display-based strategies may mask the mechanisms required for learning during problem solving (Sweller, 1988).

A good example of how over-reliance upon display-based methods during problem solving can impair learning is provided by Sweller & Levine (1982). Over a series of experiments, they demonstrated that providing problem solvers with more specific information pertaining to the goal-state paradoxically reduced the amount of exploratory learning exhibited during solution of both maze- and numericaltransformation problems (to be solved only once). The provision of highly specific goal-state information, Sweller and Levine (1982) argued, focused participants' attention on minimising differences between the current- and the goal-state (a problem solving strategy commonly referred to as 'means-end analysis'). This in turn, reduced awareness of other relations between problem states and the use of memory, both of which are required for effective learning. Strikingly, when specific goal-state information was not provided (e.g., participants were simply given a start-state and instructed to find the exit of a maze, rather than given a start- and goal-state and instructed to find an appropriate route), the use of what Sweller & Levine (1982) called 'history-cued strategies' promoted learning. History-cued strategies, by definition, require the problem solver to consider memory for previous experiences when attempting to solve problems. Thus, a stronger reliance upon internal processes, such as memory, was found to improve the exploratory learning of maze problems. This effect was exacerbated when problem solving was *visually* supported, and has subsequently received endorsement within other problem solving domains (Sweller, Mawer, & Ward, 1983; Vollmeyer, Burns, & Holyoak, 1996).

Parallels are proposed between Sweller & Levine's (1982) consideration of goal-state *specificity*, and the current investigation of goal-state *accessibility*. In Chapter 3, participants solving both simple and complex BPST problems were found to rely more upon interaction-intensive display-based strategies during problem solving when access costs were Low. In contrast, when access costs were High, participants were found to rely less upon the display as an external memory source (O'Regan, 1992), and more upon internal memory during problem solving. Perhaps such additional emphasis upon memory when access costs are High will act as germane cognitive load, and improve learning of specific actions taken (Payne et al.,

2000) and problem solving mechanisms and schemata used (Sweller, 1988; Sweller & Levine, 1982). Additional reliance upon memory during problem solving is likely to expedite the internalisation (and therefore automation) of problem solving activities, and facilitate learning of the problem (Cooper & Sweller, 1987).

Finally, the manner by which changes in IAC affect the selection between display- and memory-based planning (see Chapter 3) may also have consequences for what has been termed 'reflection' during problem solving. By imposing a keystroke limit on interactions during a visually presented programming task, Trudel & Payne (1995) successfully improved exploratory learning, when compared to learning with no key-stroke limit. Participants were given twenty minutes to explore the functions of a simulated digital watch, and when interaction with the digital watch was limited, participants were found to learn more, as evidenced by better performance at test (upon which they were required to perform a number of programming functions). Trudel and Payne (1995) argued that by imposing a limit on the number of keystrokes that could be used to explore the watch functions, each interaction with the device became more valuable, and thereby encouraged users to pay more attention to and think more carefully about each operation, thus enhancing learning.

A similar argument is put forward here with regard to the potential effects of IAC upon learning during repeated solution to an eight-puzzle-like BPST. As the cost of accessing the goal-state increases, it is predicted that participants will again choose to inspect the goal-state less frequently and make more problem solving moves per goal-state inspection. It is predicted that such additional emphasis upon internal problem solving strategies both during the process of mentally evaluating plans and during move execution will improve learning of the specific methods/schemata required to solve an eight-puzzle-like BPST problem.

As with Chapter 3, all studies contained within Chapter 4 will also assess the effect of IAC upon strategy selection, and any effects of IAC observed upon learning will be discussed within the context of Davies's (2003; 2004) conceptualisation of memory-based and display-based problem solving and planning; during which there has been no mention of the affordances provided by each in terms of learning. Experiment 4.1a initially examined the potential contribution of IAC to rate of learning during repeated solution of an eight-puzzle-like BPST problem. Experiment 4.1b extended this investigation to assess the effect of IAC upon transfer to a previously unseen eight-puzzle-like BPST problem. Finally, Experiment 4.2 concludes with an examination of the effects of small access costs experienced during problem solving on retrospective and prospective memory for task-critical information.

4.2. Eight-puzzle-like BPST

For all experiments, the eight-puzzle-like BPST used in Experiment 3.3 was adapted for the purpose of Chapter 4 (see Figure 4.1 for a screenshot & Section 3.2 for a description of this task). In order to assess the possibility that higher access costs are more conducive than lower access costs to learning during repeated solution to an eight-puzzle-like BPST problem, IAC was manipulated in the same way as in Experiments 3.2 and 3.3. Both windows were permanently uncovered when IAC was Low. In contrast, the Goal-State Window was covered by a grey mask when IAC was Medium or High, and could only be uncovered by placing the mouse cursor over the window. The grey mask then reappeared the moment the mouse cursor left the Goal-State window. There was also an additional 2.5 second lockout associated with uncovering the Target Window when IAC was High.

Figure 4.1: An example of an eight-puzzle-like BPST start-state in the Low IAC condition. Goal-State Window is left, Current-State Window is right (Experiment 4.1a).

All moves were made by selecting the block to be moved in the Current-State via a left mouse click and then pressing the corresponding arrow key depending on the direction of the desired movement. Rather than solving six different BPST problems one after the other (Chapter 3), all participants in Chapter 4 were required to solve the same eight-puzzle-like BPST problem repeatedly ten times. If IAC is found to operate in the same way as observed in Chapter 3, memory requirements of the task will again rise in conjunction with an increase in IAC. Learning on this task using working memory (Baddeley, 1986) will require participants to encode the Goal-State configuration of coloured blocks and/or a sequence of moves-to-solution and/or an appreciation of problem relations (schema) in order to reach solution. Two different problems were also used in all experiments in order to minimise the possibility of observing idiosyncratic effects due to particular surface characteristics.

4.3. Experiment 4.1a

The aim of Experiment 4.1a was to assess the contribution (if any) of the cost of accessing Goal-State information to the rate of learning during repeated solution of an eight-puzzle-like BPST problem. Based upon the findings of Chapter 3, no differences were expected in terms of solution efficiency at trial one. However, it was anticipated that due to mechanisms outlined in the introduction to Chapter 4, learning of the solution may be facilitated by higher access costs. Higher access costs were predicted to facilitate learning via stronger reliance on memory (e.g., Sweller, 1988; Sweller & Levine, 1982), and thus, additional benefits of an increase in IAC are expected to manifest as a function of practice.

Two main questions are of interest here. Firstly, how will the number of moves and time required to solve each trial be affected by IAC? Any benefits in problem solving proficiency in either of the IAC conditions would be reflected by reduced moves and/or time required for each solution. Secondly, how will Goal-State inspection frequency and time change within each IAC condition as participants become more practiced at the task? On the one hand, it could be argued that having successfully solved the same problem nine times previously, participants in all IAC conditions would have learned the Goal-State, and thus would no longer need to spend time inspecting such information on the final trial. However, the presence of more display-based problem solving apparent when access costs are Low (Chapter 3) may mean that learning of the Goal-State will take longer, when compared to problem solving with higher access costs, and thus regular inspections may still be necessary in the Low IAC condition during the final trial.

In order to explore the latter hypothesis, an eye-tracker was used throughout Experiment 4.1a in order to assess the extent to which memory- and display-based strategies were adopted during repeated solution to an eight-puzzle-like BPST problem. This will allow novel conclusions to be drawn with respect to the effect of IAC upon access strategies within a paradigm where considerable learning is expected.

4.3.1. Method

Participants Thirty-six Cardiff University undergraduate Psychology students participated in the study for course credit, and were randomly assigned to one of six conditions.

Apparatus/Materials The experiment was written in *Microsoft Visual Basic 6* and was conducted using a 2 Ghz Pentium IV PC connected to a *Tobii 1750* 34 x 27 cm eye-tracker monitor, extended keyboard and mouse. All eye movements were recorded at a rate of 15 frames per second, with time-stamp accuracy of +/-3 ms. Due to rapid fixation switches observed, no limit was set on the time a fixation must last in order to be classified as a fixation. Gaze estimation was within 1 degree of accuracy, even across large head movements. Mouse movements and key presses were also recorded and saved. The Goal- and Current-State Windows were the same size and each contained nine blocks residing within a 3 x 3 grid. Eight coloured blocks and one empty space resided within each window. No colours were used twice, and the empty space was white. The rules of the task determined that only one adjacent (horizontally or vertically) coloured block could be moved into the empty space at a time.

Design IAC and problem type were manipulated between-subjects in order to negate possible contamination via asymmetric transfer (Poulton, 1982). Independent manipulation of IAC and problem type resulted in six between-subject treatment conditions: Low IAC/problem A; Low IAC/problem B; Medium IAC/problem A; Medium IAC/ problem B; High IAC/problem A; High IAC/problem B. Problems A and B could both be solved in a minimum of 17 moves and shared the same Goal-State (see Appendix C). However, different solution paths were required in order to solve the two problems (due to different Start-States). The use of two different problems was purely a design issue to ensure that idiosyncratic effects were not reported associated only with a particular problem.

A number of dependent measures were taken throughout the experiment. Firstly, the frequency with which participants' chose to inspect the Goal-State was measured. Necessarily, data from the Low IAC condition came from the eye-tracker (consecutive fixations within the Goal-State window were collapsed and counted as one). In accordance with Gray & Fu (2001; 2004), data from the Medium and High IAC conditions came from the *Visual Basic* program (which measured the frequency with which the Goal-State Window was uncovered).

The eye-tracker also provided data pertaining to the time participants spent viewing the Goal-State when the IAC was Low or Medium. However, when IAC was High, eye-tracker data were not used to measure Goal-State viewing time due to the fact that participants were able to view the area defined within the eye-tracker analysis software as representing the Goal-State whilst enduring the 2.5 s lockout[5]. The *Visual Basic* program was therefore used to provide a measure of the time participants spent uncovering the Goal-State Window in the Medium and High IAC conditions.

---

[5] Despite the fact that the contents of the Goal-State were masked during the 2.5 s lockout, participants still chose to view the Goal-State area whilst waiting for the lockout to time out because there was nothing else of use to look at during this time. As it was not possible to instruct the eye-tracker analysis software not to record eye fixations during these lockout times, data contamination would likely occur in the High IAC condition if this measure was used.

In addition, the *Visual Basic* program recorded the number of moves made (total and palindromic), the time taken to complete each trial and first-move latencies. As in Chapter 3 (see Section 3.2), use of the term 'move' herein refers to single bock movements only. By dividing the number of moves by the number of Goal-State visits for each trial, an estimate of the number of moves made per Goal-State inspection was obtained.

Procedure Participants were seated approximately 50 cm away from the eye-tracker and handed an instruction sheet (see Appendix L). Following successful eye-tracking calibration, participants were required to complete a short block movement task that required the movement of each of the blocks in a sequence. This allowed participants in each of the IAC conditions to familiarise themselves with the method of moving blocks and inspecting the Goal-State. Each participant was then required to solve the same problem (A or B) ten times.

4.3.2. Results

A 3 (IAC Low/Medium/High) x 10 (Trial) ANOVA was computed for each of the dependent variables, with the first factor manipulated between-subjects, and the final factor manipulated within-subjects. Four participants' data were removed from all analyses due to obtaining scores in excess of $\pm$ 3 standard deviations from the mean. The eye-tracker failed to record eye fixations for three participants. On such occasions, missing data were excluded during analyses (Tabachnick & Fidell, 2001). Log transformations were required in order to obtain homogeneity of variance for the following variables: Time-to-solution, Goal-State inspection frequency and viewing/uncovering time. Where assumptions of sphericity were not met, Greenhouse-Geisser corrected degrees of freedom are reported. The effects of IAC and trial (along with any interactions) will be presented for each dependent variable in turn. Only minimal effects of problem type were observed, and thus this variable was collapsed during all analyses.

Moves-to-solution A main effect of trial was found, F (5.01, 145.39) = 6.41, p <.001, MSE = 5014.99, and a trial x IAC interaction was observed, F (10.03, 145.36) = 1.92, p <.05, MSE = 5014.99, suggesting that learning may have been differentially affected by IAC (see Figure 4.2).

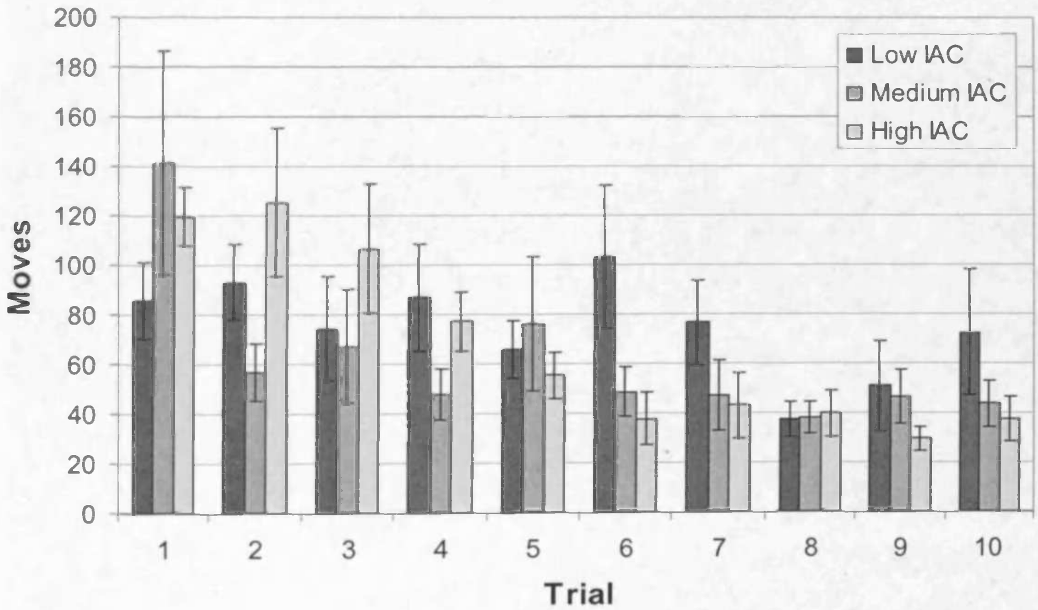


Figure 4.2: The interaction between trial and IAC for moves-to-solution (Experiment 4.1a).

Note. Error bars represent ± 1 standard error.

Initial simple main effect analyses revealed the only effects of trial to occur in the Low, F (9, 21) = 3.76, p <.01, and High IAC conditions, F (9, 21) = 3.14, p <.05, and the only simple main effect of IAC was observed to occur at trial six, F (2, 29) =

3.44, $p < .05$, $MSE = 3923.28$. However, of more importance to the current investigation was the potential effect of IAC upon learning across trials. In order to explore this, trend analyses were performed on the simple effects of trial in each of the IAC conditions. This revealed significant linear effects of trial in the Medium and High IAC conditions (*ps* <.01), but not in the Low ($p > .05$), suggesting better learning as IAC increased. No main effects of IAC, $F\ (2, 28) = 0.42$, $p > .05$, $MSE = 1109.60$, were observed. Although the average moves-to-solution data seem rather high at trial ten (given that the problem can be solved in a minimum of 17 moves), the Low IAC data are very comparable to that reported by O'Hara & Payne (1998). Large variation in the number of moves required to solve each problem meant that although a small number of participants did manage to solve the eight-puzzle-like BPST in the minimum number of moves, those that didn't (and these were sometimes way off) brought the average up considerably.

In terms of the effects of IAC and trial upon the number of palindromic move sequences, a similar pattern of results was observed to those presented above for the total number of moves. The only main effect was that of trial, $F\ (4.70, 136.25) = 5.15$, $p < .001$, $MSE = 147.05$, and the only interaction found was for trial x IAC, $F\ (9.40, 136.25) = 2.08$, $p < .05$, $MSE = 147.05$. Simple main effects found a non-significant effect of IAC at trial six that began to approach the .05 alpha level, $F\ (2, 29) = 3.13$, $p < .06$, $MSE = 57.06$, and simple main effects of trial were found only in the High and Low IAC conditions (*ps* <.05). Again, however, when assessing learning within each IAC condition, significant linear effects of trial were observed in the Medium and High IAC conditions, (*ps* <.05), but not the Low ($p > .05$).

Time-to-solution Solution times were reliably affected by trial, $F(4.88, 141.36) = 36.41$, $p < .001$, $MSE = 0.13$, and a significant trial x IAC interaction (see Figure 4.3) was again observed, $F(9.75, 141.36) = 2.59$, $p < .01$, $MSE = 0.13$. Simple main effects revealed an effect of IAC at trial one only, $F(2, 29) = 5.69$, $p < .01$, and Bonferroni-corrected comparisons indicated longer times in the High than the Low IAC condition ($p < .01$). Simple main effect analyses also revealed a significant effect of trial within each IAC condition (*ps* <.01), and further polynomial inspection revealed significant linear effects of trial within all IAC conditions (*ps* <.001). Removing data pertaining to time spent uncovering the Goal-State window in the High IAC condition made virtually no difference to the analyses reported above, and no further main effects were observed.

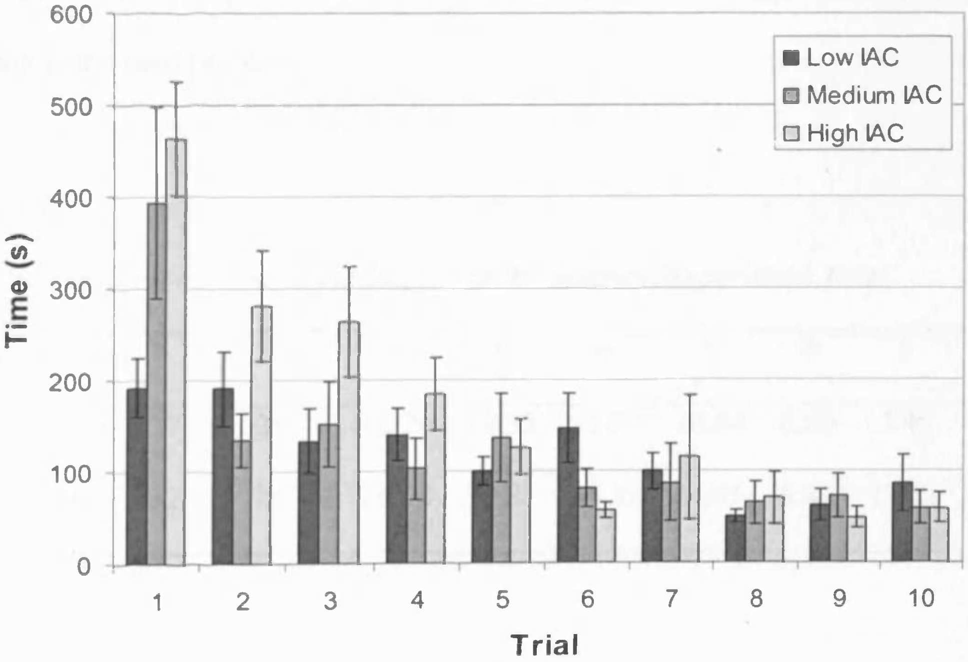


Figure 4.3: The interaction between trial and IAC for time-to-solution (Experiment 4.1a).

Note. Error bars represent ± 1 standard error.

Goal-State inspection frequency The number of Goal-State inspections for all IAC conditions was affected by both trial, $F(4.12, 111.15) = 56.21, p < .001, MSE = 0.17$, and IAC, $F(2, 27) = 21.84, p < .001, MSE = 0.76$. Inspection frequency reduced in a linear fashion as a function of trial (see Table 4.1), $F(1, 27) = 133.85, p < .001, MSE = 0.15$, and Bonferroni-corrected pairwise comparisons revealed significant differences between all IAC conditions (*ps* <.05), with Goal-State inspection frequencies reducing as a function of an increase in IAC (Low IAC: *Mean* = 37.63, *SD* = 37.76; Medium IAC: *Mean* = 12.88, *SD* = 17.38; High IAC: *Mean* = 4.97, *SD* = 7.45). No IAC x trial interaction was observed, $F(8.23, 111.15) = 0.62, p > .05, MSE = 0.10$, and the effect of IAC was apparent at every trial (*ps* <.05). It is interesting to note that by trial ten, participants in the High IAC condition chose to inspect the Goal-State very rarely (1.18 times on average), whereas participants in the Low IAC condition still felt it necessary to inspect the Goal-State on average 21.44 times during solution to the final problem.

Table 4.1

The Effect of Trial on Goal-State Inspection Frequency (Experiment 4.1a).

Trial	1	2	3	4	5	6	7	8	9	10
Mean	45.40	27.30	21.30	16.20	14.53	16.07	10.90	6.90	7.40	8.07
SD	39.73	32.11	24.16	20.54	17.81	36.32	14.81	8.42	13.22	23.13

Goal-State inspection time The time participants spent viewing the Goal-State in the Low and Medium IAC conditions (as measured by the eye-tracker) was also significantly affected by trial, $F(3.16, 53.65) = 36.34, p < .001, MSE = 0.17$, with values reducing in a linear fashion as a function of practice, $F(1, 17) = 75.47, p$

$<.001$, $MSE = 0.24$. No effect of IAC, $F(1, 17) = 0.13$, $p > .05$, $MSE = 0.84$, or an IAC x trial interaction was observed, $F(3.16, 53.65) = 0.33$, $p > .05$, $MSE = 0.17$.

The time spent uncovering the Goal-State in the Medium and High IAC conditions (as measured by the *Visual Basic* program) was affected in much the same manner by trial, $F(2.55, 48.45) = 57.31$, $p < .001$, $MSE = 0.30$, with values again reducing in a linear fashion, $F(1, 19) = 111.17$, $p < .001$, $MSE = 0.37$. As with Goal-State viewing time, no main effect of IAC, $F(1, 19) = 0.29$, $p > .05$, $MSE = 2.21$, nor any reliable interaction between IAC and trial was observed, $F(2.55, 48.45) = 1.47$, $p > .05$, $MSE = 0.30$.

First-move latency The time participants chose to spend considering their first move was significantly affected by trial, $F(3.29, 95.37) = 19.48$, $p < .001$, $MSE = 86.69$, and these values reduced in a linear fashion, $F(1, 29) = 45.90$, $p < .001$, $MSE = 109.77$. Confirming results reported in Chapter 3, IAC also affected the time participants spent considering their first move, $F(2, 29) = 4.20$, $p < .05$, $MSE = 391.71$, with Bonferroni-corrected post hoc analyses revealing significantly longer first-move latencies in the High, compared to the Low IAC condition ($p < .05$).

Simple main effects exploring a significant trial x IAC interaction (see Figure 4.4), $F(6.58, 95.37) = 3.64$, $p < .001$, $MSE = 86.69$, found reliable effects of IAC during the first four trials only ($ps < .05$), with Bonferroni-corrected pairwise comparisons revealing significantly longer latencies in the High than the Low IAC condition at each of these trials ($ps < .05$). Furthermore, a simple main effect of trial upon first-move latencies was observed in the High IAC condition, $F(9, 21) = 7.38$, $p < .001$, but not the Low or Medium ($ps > .05$). Additional trend examination of the simple effects of trial upon each IAC condition revealed linear effects of trial within the Medium and High IAC conditions (*ps* <.001), but not the Low (*p* >.05). Removing delays incurred via uncovering the Goal-State when the IAC was High (during which time no information could be viewed) made only minimal difference to the results reported above.

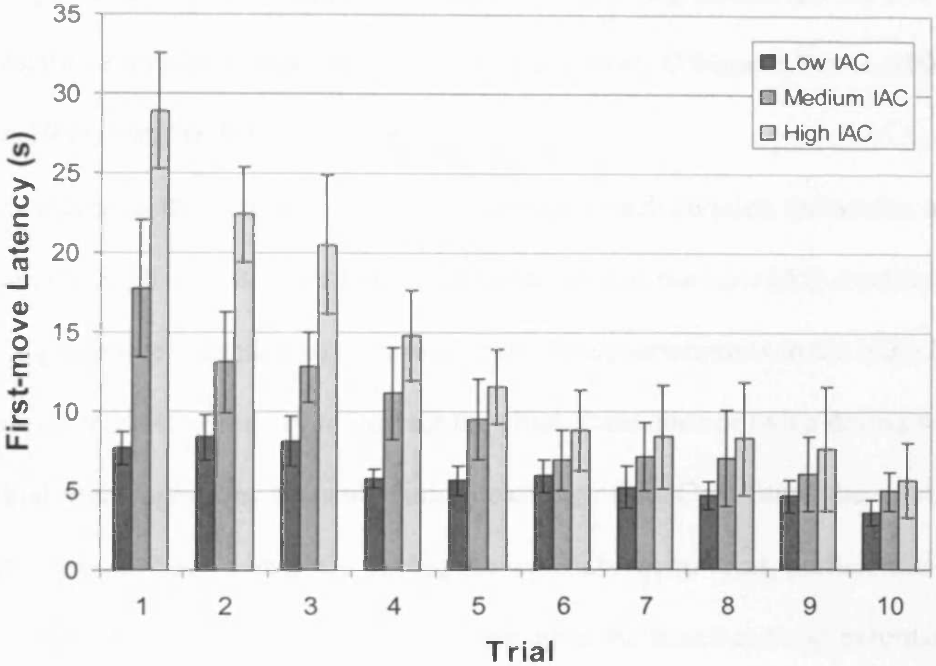


Figure 4.4: The interaction between IAC and trial for first-move latency data (Experiment 4.1a).

Note. Error bars represent ± 1 standard error.

4.3.3. Discussion

Corroborating the results observed in Chapter 3, and extending previous work examining information access strategies during routine behaviour (Gray & Fu, 2004; Gray et al., 2006) to a learning domain, small rises in IAC were found to increase reliance upon memory during problem solving. A significant reduction in the frequency with which participants chose to inspect the Goal-State was observed as a consequence of an increase in IAC, and this was evident at every trial. Measures of the time spent considering the first move suggested that changes in IAC affected the selection between initial memory-based and concurrent display-based planning styles (Davies, 2003; 2004) during early trials in much the same way as observed in Experiments 3.2 and 3.3. Experiment 4.1a also found some support for the assertion that stronger reliance upon memory during problem solving (moderated by IAC) would facilitate incidental learning (Delaney et al., 2004; O'Hara & Payne, 1998; Sweller, 1988; Sweller & Levine, 1982).

For example, the number of moves required to reach solution reduced in a linear fashion in the High and Medium IAC conditions, but not the Low IAC condition; suggesting expedited learning as IAC increased. Also, participants in the High IAC condition only felt it necessary to inspect the Goal-State once or twice during solution to the final trial, suggesting that substantial learning of the Goal-State, the solution sequence, or both, had taken place during the ten trials. In contrast, participants in the Low IAC condition still appeared to be relying upon the interface as an external memory source (O'Regan, 1992) during solution to the final trial (exhibiting twenty plus inspections on average), despite having successfully solved the same problem nine times previously. This may suggest a persistent reliance upon display-based strategies (Sweller, 1988) during all ten trials when access costs are low, with the transition to more memory-based problem solving largely masked (Card, Moran, & Newell, 1983; Logan, 1988; Newell, 1990).

Improved learning observed as a function of small increments to IAC is not overly surprising, given that an increase in IAC encouraged participants to use memory more during problem solving. However, limitations to the results observed in Experiment 4.1a include the observation that only at trial six were benefits of higher access costs observed for moves-to-solution, and at no point did higher access costs lead to a reduced time-to-solution. A further experiment was therefore deemed necessary in order to seek more tangible evidence for the role of access costs in promoting learning during repeated solution of an eight-puzzle-like BPST problem.

4.4. Experiment 4.1b

A more explicit attempt will feature in Experiment 4.1b to examine the proposal made above that Low access costs mask the mechanisms required for effective learning during repeated problem solving by encouraging persistent reliance upon display-based strategies, whereas small increments to the cost of accessing the Goal-State facilitate learning by promoting more memory-dependent problem solving strategies. In order to test this assertion further, a novel measure of transfer will be appended to the previous experiment in order to ascertain the extent to which small increments in IAC can be used to facilitate transfer to a previously unseen, but similar eight-puzzle-like BPST problem (different Goal- and Start-State).

Transfer of learning between problems Previous work examining transfer of learning between problems has often found poor performance, even when problems are highly related (Cooper & Sweller, 1987). It has been recognised that some solution methods may be effective for supporting solution to a specific problem, but may not support 'abstraction of knowledge' pertaining to the structure of the problem - which is required for effective transfer to novel but related problems (Sweller, 1988). Predictions are made here that higher access costs may facilitate transfer to a novel but similar problem, due to improved internalisation of the operators and strategy required to solve an eight-puzzle-like BPST. As will become evident, the effect of IAC upon the use of memory during problem solving is central to this hypothesis.

When access costs were Low in the previous experiment, regular inspections to the Goal-State were still made during solution to the final trial (in excess of twenty inspections). One interpretation of this finding is that problem solving remained largely display-based, with little internalisation of the Goal-State, the move sequences required to solve the problem, or both. Another related interpretation of this finding is that problem solving remained governed to some extent by comparisons between the Goal- and Current-State during the entire practice phase, when compared to the High IAC condition (in which only 1.18 inspections to the Goal-State were observed during solution to the final trial – Experiment 4.1a).

Relevant research conducted by Vollmeyer et al (1996) suggested that although reliance upon the Goal-State during problem solving can provide sufficient support for solution to the goal in question, negative consequences may arise when the problem solver is required to transfer knowledge to a similar but different problem from the same family. Using a complex dynamic 'biology-lab' task, Vollmeyer et al (1996, Experiment 2) were able to demonstrate similar rates of learning between two conditions on an initial task, one where participants were instructed to work towards the goal to be tested during an exploratory period, and the other where participants were not told to work towards a particular goal during the exploratory period. However, when faced with a novel but similar biology-lab problem, performance was significantly less error-prone when exploratory learning had not been largely dictated by minimising the differences between the Current- and the Goal-State.

The current experiment aimed to assess the extent to which IAC may play a similar role in determining the transfer of knowledge from the repeated solution of one eight-puzzle-like BPST problem to another. Based upon the finding reported in Experiment 4.1a that problem solving remained dictated by regular inspections of the Goal-State when IAC was Low, it is predicted that poor transfer to a novel problem may be observed in this condition. In contrast, better transfer is predicted in the High IAC condition, because of the more regular use of memory during problem solving and fewer inspections of the Goal-State observed during the practice phase.

4.4.1. Method

Participants & Materials Thirty-six Cardiff University undergraduate students participated in the study in return for course credit, and were randomly assigned to one of six conditions. All materials and apparatus were identical to those used in Experiment 4.1a, with the exception that no eye-tracker was employed as the focus was upon problem solving proficiency and transfer. The novel BPST could be solved in a minimum of 17 moves, and to ensure that any transfer of knowledge from the 'practice problem' to the 'novel problem' was not specific to the surface characteristics of the practice problem, previously unseen Start- and Goal-States were used (see Appendix C).

Design & Procedure All manipulations and experimental procedure during the practice phase were identical to those employed in Experiment 4.1a (see Appendix L for task instructions). Upon successful completion of trial ten, a text box informed participants that they were now to solve another, previously unseen problem. Solution to this problem commenced when the 'Start button' was pressed. All participants in all practice IAC conditions received the same novel problem with a Low IAC interface in order to improve comparability of transfer between conditions (O'Hara & Payne, 1998).

4.4.2. Results

Firstly, analyses examining the effect of IAC on strategy selection and learning during the practice phase will be presented. Secondly, an evaluation of transfer during solution to the novel eight-puzzle-like BPST problem will be reported as a function of IAC. Two participants' data were removed from all analyses due to obtaining scores in excess of ± 3 standard deviations from the mean. Again, problem type had only minimal effects upon the results, and thus this variable was collapsed during all analyses.

Practice phase The results concerning strategy selection largely corresponded with the data reported in the previous experiment. However, some important differences between the two experiments were observed with regard to the effect of IAC upon learning during the practice phase. A 3 (IAC Low/Medium/High) x 10 (Trial) ANOVA was computed for each of the dependent variables with the first factor manipulated between-subjects, and the final factor manipulated within-subjects. Log transformations were required in order to obtain homogeneity of variance for the following variables: Time and Moves to solution, and Goal-State uncover frequency and time. Where assumptions of sphericity were not met, Greenhouse-Geisser corrected degrees of freedom are reported. Data regarding strategy selection will be presented first, followed by analyses examining problem solving proficiency. Because no eye-tracker was used in Experiment 4.1b, some of the strategy selection analyses were necessarily conducted without Low IAC data.

Strategy selection. As with Experiment 4.1a, Goal-State uncover frequencies were found to reduce in a linear fashion as a function of trial, F (1, 20) = 294.99, p <.001, MSE = 0.65, and as IAC increased from Medium to High, F (1, 20) = 12.29, p <.001, MSE = 0.83. Again, participants in the High IAC condition chose to inspect the Goal-State very infrequently on the final trial (Medium IAC: *Mean* = 4.70, SD = 4.55; High IAC: *Mean* = 0.42, SD = 0.93), suggesting considerable learning of the Goal-State in this condition. Goal-State uncover time also reduced in a linear fashion as a function of trial, F (1, 20) = 280.43, p <.001, MSE = 0.17, but was again unaffected by IAC, F (1, 20) = 3.90, p >.05, MSE = 1.84.

The time taken considering the first move again reduced in a linear fashion as a function of trial, F (1, 31) = 39.18, p <.001, MSE = 93.48. A significant interaction between trial and IAC was also found, F (8.34, 129.36) = 2.91, p <.001, MSE = 79.17, with simple main effect analyses revealing an effect of IAC at trial one, F (2, 31) = 4.95, p <.05, MSE = 77.63, and Bonferroni-corrected post hoc analyses pointing to significantly longer first-move latencies in the High compared to both the Low and Medium IAC conditions (*ps* <.05), supporting the previous finding that initial planning was more extensive when IAC was High during the first trials. Some evidence of improved learning in the higher IAC conditions came from the replicated finding that first-move latencies were again found to reduce in a linear fashion as a function of trial in the Medium, F (1, 31) = 11.07, p <.001, MSE = 93.31, and High IAC conditions, F (1, 31) = 41.28, p <.001, MSE = 93.31 but not in the Low, F (1, 31) = 1.28, p >.05, MSE = 93.31. Excluding delays incurred upon uncovering the Goal-State in the High IAC condition (during which no information was visible) made no discernable difference to the results reported above.

Learning. Although IAC appeared to affect strategy selection in much the same manner as the previous experiment, and some evidence of improved learning as IAC increased was found with respect to the first-move latency data, the problem solving proficiency analyses do not fully support some of the conclusions made with regard to moves- and time-to-solution.

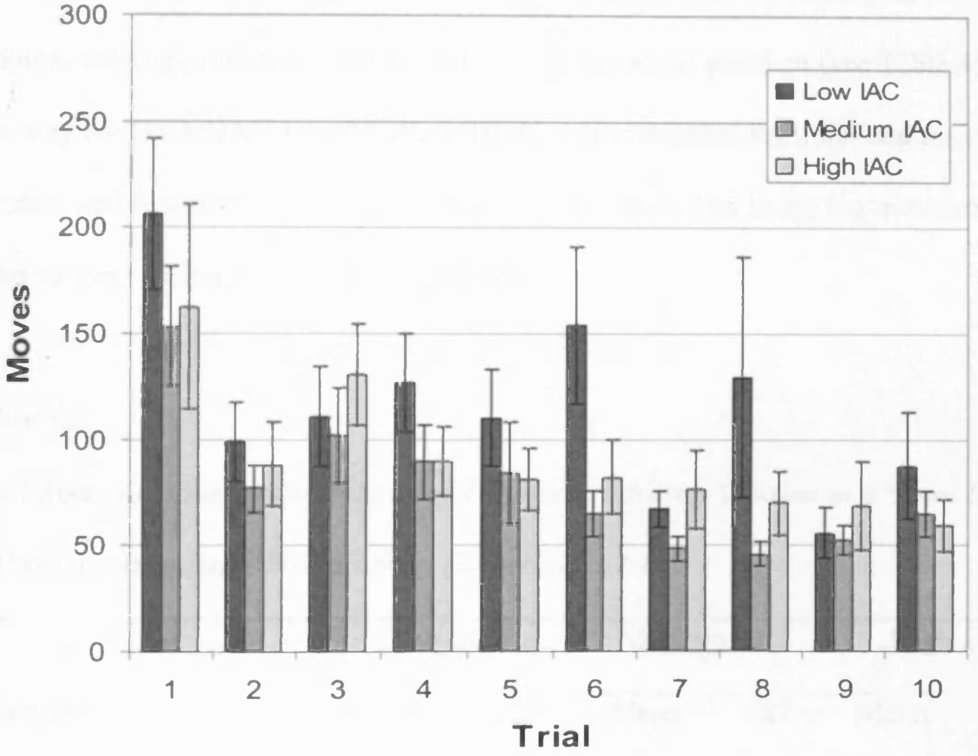


Figure 4.5: The effect of trial and IAC on moves-to-solution (Experiment 4.1b).

Note. Error bars represent ± 1 standard error.

Despite linear reductions in the number of moves (total and palindromic) and the time required to solve each problem as a function of trial (*ps* <.001), no trial x IAC interactions were found for total/palindromic moves or time-to-solution (*ps* >.05). This suggests that the effects of IAC upon learning reported in Experiment 4.1a need to be taken with some caution. It is worth noting, however, that the Low IAC condition again exhibited considerable variance with regard to moves and time-to-solution at trials six and eight (see Figure 4.5), indicating that some participants in this condition were still encountering greater problems during the solution process than participants in the higher IAC conditions.

Transfer effects All participants solved the novel problem with a Low IAC interface, but the effect of IAC in place during the practice phase was explored with respect to problem solving proficiency during solution to the novel problem (see Table 4.2). A one-way ANOVA (IAC Low/Medium/High) was computed for time- and moves-to-solution and first-move latencies, with time-to-solution data being log transformed in order to meet assumptions of homogeneity.

Table 4.2

The Effect of IAC on Problem Solving Proficiency during Solution to a Novel Low IAC eight-puzzle-like BPST problem (Experiment 4.1b).

	Low IAC		Medium IAC		High IAC	
Measure	Mean	SD	Mean	SD	Mean	SD
Time-to-completion (s)	214.71	169.56	139.19	92.13	109.42	61.44
Moves-to-completion	152.17	102.13	92.60	63.24	61.83	29.83
Palindromic move sequences	25.83	20.91	13.60	11.91	9.42	4.83
First-move latencies (s)	13.51	9.41	10.91	6.66	14.89	8.01

Although no main effect of IAC was observed for the time required to solve the novel problem (despite a 100% reduction in time-to-solution), F (2, 28) = 1.74, p >.05, MSE = 0.09, evidence of improved transfer when access costs were High was found when examining moves-to-solution data, $F(2, 28) = 4.46$, $p < .05$, $MSE = 5650.66$. Post hoc analyses (Bonferroni-corrected) revealed significantly fewer moves were required to solve the novel problem in the High IAC condition, when compared to the Low ($p < .05$). In addition, a main effect of IAC upon palindromic moves was observed, $F(1, 28) = 3.90$, $p < .05$, $MSE = 222.12$, with Bonferroni-corrected post hoc analyses revealing significantly fewer reflected move sequences in the High, compared to the Low IAC condition ($p < .05$). No effect of IAC was found for first-move latencies obtained during solution to the novel problem, $F(2, 28) = 0.62$, $p > .05$, $MSE = 70.60$.

4.4.3. Discussion

First and foremost, it is theoretically interesting to note that despite the observation that access cost manipulations in place during the practice phase manifested relatively small effects upon problem solving proficiency during repeated solution to the same eight-puzzle-like BPST problem, higher access costs in place during practice were found to significantly facilitate the transfer of problem solving knowledge from one eight-puzzle-like BPST problem to another. Secondly, although the effects of IAC on problem solving during the practice phase revealed a largely similar pattern of results to those observed in Experiment 4.1a, the fact that very few interactions between IAC and trial were observed suggests that some of the conclusions drawn from Experiment 4.1a with regard to the influence of IAC upon learning should be taken with caution and require further examination.

It is proposed that over-reliance upon display-based strategies may be responsible for the problematic transfer of knowledge from one eight-puzzle-like BPST problem to another when IAC was Low (Sweller, 1988; Sweller & Levine, 1982), and regular use of memory during problem solving was responsible for the improved transfer effect when IAC was High. In particular, it is argued that by persisting to interleave regular inspections of the Goal-State between moves made in the Current-State, problem solvers in the Low IAC condition failed to learn the structure of the problem as a whole, which would be required for effective transfer to novel but similar problems (Vollmeyer et al., 1996).

In both Experiments 4.1a and 4.1b, the very low values exhibited by participants in the High IAC condition for Goal-State inspection frequency during solution to the final trial suggested that much of the information required to solve the eight-puzzle-like BPST problem had been internalised by participants in these conditions, and would thus have been available for transfer. In particular, it is argued that higher access costs led to the abstraction and internalisation of general problem solving methods used to solve the eight-puzzle (Gick & Holyoak, 1987). Had higher access costs simply resulted in better internalisation of the specific methods required to solve the particular problem experienced during the practice phase, no general transfer would have been observed when faced with a similar but different problem (Singley & Anderson, 1989). The transfer test could therefore be seen as a means of differentiating between specific and general transfer.

The lack of differences between IAC conditions in terms of first-move latencies during solution to the novel problem suggested that the advantages of higher access costs observed at transfer were likely due to knowledge learned during the practice period, rather than differences in the selection between initial and concurrent problem solving strategies during solution to the novel problem. However, more attention needs to be paid to exploring the precise strategies used to solve the novel problem. Nevertheless, it is important to note the novelty and applied appeal of the current finding that small increments to IAC can be seeded within interface design to improve transfer between similar, but different problems. This is in contrast to limited and rather unreliable benefits to transfer observed when problem solvers are given explicit hints to encourage transfer (see Patrick, 1992 for a review).

A simplistic reading of the transfer-appropriate processing literature (e.g., Morris et al., 1977) could be argued to posit that practice in the Low IAC condition should have led to better transfer of knowledge to the novel problem, because the test conditions were more similar to practice. However, rather than emphasising the importance of congruency between the two test conditions, Experiment 4.1b found support for the importance of congruency between the type of processing encouraged by the structure of the interface and the type of processing required by the testing situation. Such a finding can be accommodated by various extensions of the transfer-appropriate processing framework, such as the 'materials-appropriate processing' model (McDaniel & Einstein, 1989), and Anderson's (1987) well documented perspective on transfer:

> "*There will be positive transfer between skills to the extent that the two skills involve the same productions.*" (Anderson, 1987, p. 197)

The failure of Experiment 4.1b to replicate the differential rates of learning during the practice phase as a function of IAC questions to what degree IAC affects problem solving performance during repeated solution of an eight-puzzle-like BPST problem. In order to explore this further, a final experiment sought to investigate the extent to which IAC affects criteria other than problem solving efficiency. Following from recent work conducted examining memory for goals (see Altmann & Trafton, 2002 for a review), memory for problem solving information was assessed at various points during the repeated solution process. It was anticipated that persistent reliance upon display-based strategies in the Low IAC condition may result in inferior memory for problem solving information (e.g., goals), when compared to the memory-intensive microstrategies employed by participants in the High IAC condition.

4.5. Experiment 4.2

The ability to decompose a complex problem into subgoals has long been seen as a cornerstone of human cognition (e.g., Miller, Galanter, & Pribram, 1960; Newell & Simon, 1972). The mental representation of a goal to achieve some specific state of the world must often be suspended and resumed at a later date (particularly when hierarchical problem solving is required). Traditionally, a last-in, first-out 'popping' mechanism has been instantiated, whereby memory for goals is seen as perfect (e.g., Anderson & Lebiére, 1998). Recent research, however, has eschewed the assumption made by prominent theories of problem solving that memories for goals have a specialised status (see Altmann & Trafton, 2002 for a more detailed discussion of this debate).

Strengthening memory for goals Converging evidence now strongly suggests that memories for goals are equally vulnerable to the cognitive constraints that shape declarative memory in general (Altmann & Trafton, 1999; 2002; Anderson & Douglas, 2001; Hodgetts & Jones, 2006; Morgan, 2005). The activation of goals has been noted as an important determining factor during the solution of tasks such as the Tower of Hanoi (Altmann & Trafton, 1999), and in their 'goal activation model', Altmann & Trafton (2002) predict that memory for goals is subject to interference, and the amount of strengthening a goal receives determines whether or not it will be sampled during a retrieval cycle, or whether it will decay in memory. Of particular relevance is the following statement provided by Altmann & Trafton (2002):

> "*For planning to succeed, each intermediate state (goal) must be immediately and reliably available to the cognitive system, which has to form that state into its successor by means of task-related cognitive operations. The strengthening process is what makes the state immediately and reliably available, by making that state more active than its competitors*". (p. 47)

Although Altmann & Trafton (2002) limit their discussion on the strengthening of goals to the frequency with which a goal has been sampled (throughout a lifetime) and the duration of the goal's lifetime since it was first processed, there is reason to predict that changes in the cost associated with accessing Goal-State information may have consequences for the activation of such information. For example, higher costs associated with accessing Goal-State information have been shown to increase reliance upon memory for said information during routine copying (e.g., Gray, Simms, Fu, & Schoelles, 2006) and problem solving (see Chapter 3). Higher access costs may prompt rehearsal effects in goal memory (Anderson & Douglass, 2001), and theories of 'transfer-appropriate processing' (Morris et al., 1977) would predict that the use of memory during completion of a task will improve memory for that information at test. Indeed, Anderson & Milson (1989) argued that the organisation of human memory reflects the statistical properties of the environment we find ourselves in so as to make more available memories of information used more often by the memory system.

In light of evidence gathered here that problem solvers working with High access costs internalise information more often (Experiments 3.1 - 3.3, 4.1a & 4.1b), the primary aim of Experiment 4.2 was to explore the possibility that memory for goals can be improved by strategy changes induced by small increments in IAC. To this end, participants were again required to solve an eight-puzzle-like BPST problem repeatedly, but were interrupted at various points, and in a similar vein to Delaney et al (2004, Experiment 3), retrospective memory for Goal-State information was then assessed.

Based upon differences observed between IAC conditions in Experiments 4.1a and 4.1b for Goal-State inspection frequency, it was anticipated that differences would be observed between IAC conditions with regard to memory for Goal-State information. Following from the transfer-appropriate processing framework (Morris et al., 1977), it was predicted that the use of memory-intensive strategies during problem solving when access costs are High will lead to superior memory for Goal-State information, compared to display-based problem solving when access costs are Low.

As acutely demonstrated by Payne (1991), it is surprising how little is actually committed to memory when interactive behaviour is largely display-based. For example, when asked questions relating to the effects of several commands used every day in word processing tasks, even experienced users were unable to recall much information without the context of the display in front of them. Thus, it is quite plausible that largely display-based problem solving witnessed in Low IAC conditions in Experiments 4.1a and 4.1b may leave knowledge of Goal-State information fragmented, even after several successful solutions to the same problem. In addition to memory for Goal-State information, Experiment 4.2 also aimed to measure retrospective memory for Current-State information following interruption. Predicting the effect of IAC upon memory for Current-State information, however, is less straight forward. Although higher access costs have been shown to lead to stronger reliance upon memory for Goal-State information during solution to the current task, this does not necessarily hold true for Current-State information. At no point during the problem solving process are participants in any IAC condition required to rely upon memory for Current-State information for more than the duration of a small saccadic eye movement. Based upon this argument, it is difficult to predict differences in memory for Current-State information as a consequence of small manipulations to IAC. However, it is possible that by virtue of manipulating the external display in order to evaluate move sequences (Kirsh & Maglio, 1994; Neth & Payne, 2002), memory for the Current-State configuration may be found to be problematic when access costs are Low and concurrent planning is more commonplace (Davies, 2003; 2004). In contrast, the mental evaluation of different move sequences during initial planning (more evident in High IAC conditions) may improve memory for Current-State information (O'Hara & Payne, 1998).

The final measurement taken following interruption was that of prospective memory (e.g., Brandimonte, Einstein, & McDaniel, 1996; Einstein & McDaniel, 2005) for plans instantiated during problem solving. Participants were asked to recall any information they could pertaining to suspended plans. Again, if participants in the Low IAC condition persisted with display-based problem solving during solution to all ten trials, it would be predicted that memory for plans would be inferior, when compared to that exhibited by participants in the High IAC condition, where planning has previously been found to be largely maintained mentally (Experiment 3.3).

An important subsidiary goal of this study was to use an interference paradigm in an attempt to identify and differentiate the nature of encoding between IAC conditions (see Delaney et al., 2004, Experiment 4 for similar methodology). In order to do this, two different interference tasks were experienced upon interruption (prior to the aforementioned memory tests) in an attempt to provide insight into the relative importance of memory for Goal-State and planning information (see Section 4.5.1 for more details). By definition, the 'pattern recollection task' required participants to use their memory to complete the task, but no planning was required. Although there were common features between the two tasks, the 'move sequence task' did not necessitate the use of memory, but did require participants to engage in planful behaviour.

Based upon the work of Delaney et al (2004), it was predicted that both interference tasks would result in retroactive interference with regard to memory for problem solving information. However, if greater internalisation of Goal-State and/or planning information was responsible for the improved learning in previous experiments when access costs were higher, one would expect higher recall values in the High than the Low IAC condition. Moreover, it was proposed that any interactions between interference task and IAC would reflect differences in the organisation of memory between conditions.

4.5.1. Method

Participants & Materials Seventy-two Cardiff University undergraduate students participated in the study for either course credit or £5, and were randomly assigned to one of twelve conditions. Identical materials were used as described in Experiments 4.1a and 4.1b.

Design & Procedure IAC was manipulated in the same manner as previous experiments. Problem type was again manipulated between-subjects in order to avoid reporting idiosyncratic effects concerned with one particular problem. Different Goal-and Start-States were set for each problem (see Appendix D), yet problems A and B could both be solved in a minimum of 17 moves. The scheduling of interruptions to problem solving activity was designed to minimise anticipation on behalf of the participant. All participants were interrupted during solution to trials two, five and nine, and the onset of each interruption occurred as soon as four coloured blocks were placed in the correct positions. This performance criterion was designed to be approximately equivalent for all participants (see Section 4.5.2), and was deemed more suitable than interrupting participants after a pre-defined number of moves made, or time on task. Interfering task content was manipulated between-subjects, with participants receiving either a different pattern recollection task at each of the three interruptions or a different move sequence task at each of the three interruptions. The order in which participants received these three tasks (pattern recognition or move sequence) was counterbalanced using a Latin Square design.

When interrupted with one of the three 'pattern recollection' tasks, a 3 x 3 configuration of eight coloured blocks and one empty space was presented for five seconds with the instruction to "memorise this pattern". The block configuration subsequently disappeared and was replaced by a 3 x 3 grid of empty blocks and the instruction "recreate the pattern from memory using the coloured blocks below". This was achieved by clicking and dragging coloured blocks from a resources pile below within a time limit of twenty-five seconds. The configuration of coloured blocks making up the pattern differed for each of the three pattern recollection tasks (see Appendix E). When interrupted with one of the 'move-sequence' tasks, a 3 x 3 configuration of eight coloured blocks and one empty space appeared, and again the configuration of these colours differed for each of the three tasks (see Appendix F). On-screen instructions differed according to which of the three move sequence tasks was active. As with the pattern recollection task, total time on the planning task was limited to thirty seconds.

Upon completion of each interference task, participants were assessed on their memory for the Goal-State, the Current-State and any plans they intended to implement prior to the interruption. The order in which these three memory tests were administered was also counterbalanced using a Latin Square design. Assessment of Goal- and Current-State information required participants to use a resource of coloured blocks located at the bottom of the screen to fill an empty Goal- or Current-State grid. Memory for plans was evaluated by prompting participants to enter a text description of any plan(s) they intended to implement prior to the interruption. For each of these assessments, a time limit of one minute was imposed. Having been assessed on each of the three memory tests, the experimental program then returned to the problem solving task. However, participants were not required to complete the trial they were interrupted on. Instead, the next trial was presented afresh. (See Appendix M for task instructions.)

4.5.2. Results

The results for Experiment 4.2 will be presented in two sections. Firstly, the effects of IAC and interference task will be assessed with regard to post-interruption memory for Goal-State, Current-State and planning information. Contributions of IAC to the number of moves and time taken to reach the criterion of interruption onset, and performance on each interference task will also be evaluated. Secondly, the problem solving search strategies used by participants during non-interrupted trials for each of the IAC conditions will be summarised. The manipulation of problem type again had minimal effect, and this variable was thus collapsed during all analyses. Three participants' data were removed from all analyses due to obtaining scores ± 3 standard deviations from the mean, and where the assumption of sphericity was not met, Greenhouse-Geisser corrected degrees of freedom are reported. For seven participants, the *Visual Basic* program failed to record appropriately the time spent uncovering the Goal-State on interrupted trials. On such occasions, missing data were excluded during all analyses.

Effect of IAC on recall A 3 (IAC) x 2 (interfering task) x 3 (interrupted trial 2, 5, 9) ANOVA was computed on memory for Goal-State, Current-State and planning data, with the first two factors manipulated between-subjects, and the final factor manipulated within-subjects. In terms of memory for the Goal-State configuration, IAC significantly affected recall, F (2, 59) = 3.86, p <.05, MSE = 7.10, (see Table 4.3) with planned comparisons revealing significantly higher rates of recall in the Medium and High IAC conditions than the Low (ps <.05).

Table 4.3

The Effect of IAC on Memory for Goal-State Information (Experiment 4.2).

	Low IAC		Medium IAC		High IAC	
	Mean	SD	Mean	SD	Mean	SD
(Max. = 9)	6.33	2.95	7.50	2.33	7.37	2.44

Goal-State memory was also affected by trial, F (1.73, 102.22) = 22.29, p <.001, MSE = 5.93, with linear improvements revealed as a function of practice, F (1, 59) = 32.24, p <.001, MSE = 5.87. However, no effects of interference task were found, F (1, 59) = 1.31, p >.05, MSE = 7.10, and no variables interacted. Memory for Current-State information also improved in a linear fashion as a function of trial, F (1, 62) =

12.95, $p < .001$, $MSE = 3.48$, and an additional significant interaction between IAC and trial was observed, $F(4, 124) = 2.79$, $p < .05$, $MSE = 3.01$, (see Figure 4.6).

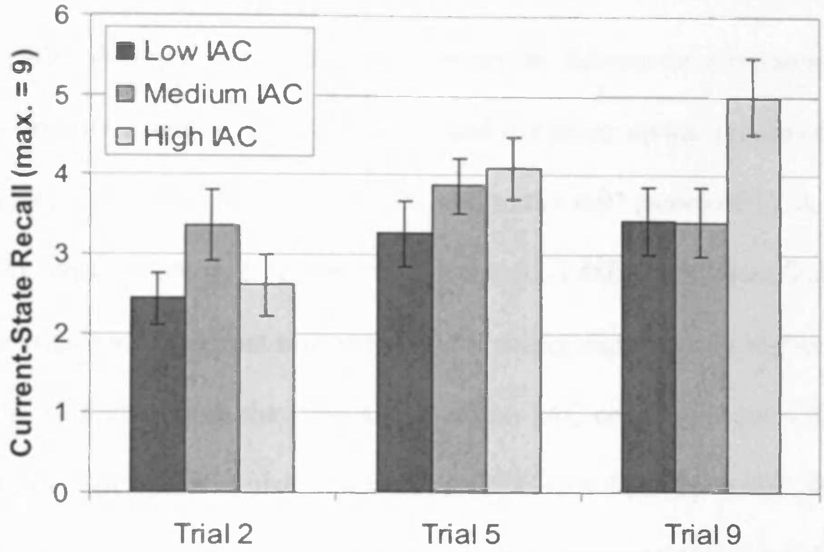


Figure 4.6: The interaction between IAC and trial for Current-State recall (Experiment 4.2).

Note. Error bars represent ± 1 standard error.

Simple main effects revealed the only effect of IAC to occur at trial nine, $F(2, 62) = 3.92$, $p < .05$, $MSE = 4.78$, with Bonferroni-corrected post hoc analyses indicating a significant benefit of High IAC over Medium ($p < .05$), and a non-significant benefit of High over Low that began to approach the .05 alpha level ($p < .06$). In addition, only in the High IAC condition did Current-State memory improve as a function of trial, $F(2, 61) = 9.44$, $p < .001$. No effects of interfering task, $F(1, 62) = 0.31$, $p > .05$, $MSE = 6.07$, nor any further interactions were observed.

With regard to memory for planning data, ANOVA was computed on the number of plans recalled by participants at each interruption. Responses were given a score of zero when participants stated they either 'had no plans', 'could not remember their plans', or the information provided was 'too vague' to determine what blocks were to be moved (e.g., "I intended to move each block into its correct position"). Each accurately recollected plan was given a score of one, and the maximum number of plans recalled during a single response was seven. Responses were sometimes as detailed as 'move the red up, the yellow left and the green down' (score of 3), and sometimes as general as 'rotate the black round to the left' (score of 1). A significant effect of IAC was observed, F (2, 63) = 5.95, p <.01, MSE = 3.57, (see Table 4.4) with Bonferroni-corrected post hoc analyses revealing significantly higher recall values in the High than both the Low and Medium IAC conditions (ps <.05). A significant effect of trial was also observed, F (2, 126) = 12.27, p <.001, MSE = 1.19, with trend analyses revealing significant linear improvements over time, F (1, 63) = 24.42, p <.001, MSE = 1.19. No effect of interference task was found, F (1, 63) = 0.98, p >.05, MSE = 3.57, nor were any interactions observed (ps >.05).

Table 4.4

The Effect of IAC on the Number of Plans Recalled (Experiment 4.2).

Low IAC		Medium IAC		High IAC	
Mean	SD	Mean	SD	Mean	SD
1.18	1.26	1.97	1.55	2.23	1.46

Effect of IAC on interruption onset & performance In order to ensure that the memory results reported above were not contaminated by differential effects of IAC on the number of moves required for participants to reach the interruption onset criteria, time on task, or time spent viewing the Goal-State prior to interruption onset, several 3 (IAC Low/Medium/High) x 3 (interrupted trial 2, 5, 9) ANOVAs were computed for these dependent variables. The first factor was manipulated between-subjects, and the final factor was manipulated within-subjects. For the purpose of parsimony, it is simply reported that no effects of IAC were observed for any of the above measures (*ps* >.05). The number of moves and time required to reach interruption onset generally reduced as a function of practice (*ps* <.05).

A number of 3 (IAC Low/Medium/High) x 3 (interrupted trial 2, 5, 9) ANOVAs were also computed on scores obtained by participants during the interference tasks in order to ensure that participants in each IAC condition were equally capable of completing the tasks. A score was recorded for each pattern recollection task depending upon the number of blocks placed in the correct position (maximum of nine). A score was also recorded for each move sequence task depending upon whether none, one or two of the objectives were achieved (maximum of two). No effects of IAC or trial were observed with regard to the number of blocks correctly placed during the pattern recollection task or the number of objectives achieved during the move sequence task (*ps* >.05).

Effect of IAC on problem solving strategy A number of 3 (IAC Low/Medium/High) x 7 (non-interrupted trial 1, 3, 4, 6, 7, 8, 10) ANOVAs were computed on problem solving strategy data in order to ensure that the effects of IAC upon recall were a consequence of the influence of IAC upon the selection between memory- and display-based strategies (as detailed in Experiments - 3.1 - 3.3, 4.1a & 4.1b). The first factor was manipulated between-subjects, and the second factor was manipulated within-subjects. Goal-State inspection frequency and uncover time data were log transformed in order to control for differences in variance between conditions.

As in previous experiments, IAC significantly affected the frequency with which participants chose to inspect the Goal-State, $F\ (2, 66) = 77.47, p < .001, MSE = 0.40$, with Bonferroni-corrected post hoc analyses revealing fewer inspections in the High than the Medium and Low IAC conditions, and in the Medium than the Low IAC condition (*ps* <.001). The effect of IAC was apparent at every non-interrupted trial (*ps* <.001), and Goal-State inspection frequency reduced in a linear fashion as a function of trial, $F\ (1, 66) = 242.14, p < .001, MSE = 0.17$.

Goal-State viewing (as measured by the eye-tracker) and uncover time (as measured by the *Visual Basic* program) were also affected in much the same way as reported in Experiments 4.1a and 4.1b, as was the first-move latency data, suggesting that the selection between initial and concurrent forms of planning was affected by IAC in much the same way as detailed previously. Both time and moves-to-solution reduced in a linear fashion as a function of trial (*ps* <.001), but neither was affected by IAC (*ps* >.05), either via main or interaction effects.

4.5.3. Discussion

Support was found for the primary hypothesis that an increase in IAC would improve memory for Goal-State information. In addition, small increments to IAC were found to improve memory for Current-State information when assessed at trial nine, and the number of plans that could be recalled. Importantly, these effects were not complicated by differences between IAC conditions with regard to time, moves or Goal-State viewing time prior to interruption onset. No effect of IAC was observed for performance during either of the interference tasks, suggesting that all participants were applying similar effort within each thirty second interference task.

These results are argued to represent a further extension to Gray and colleagues' work on the role played by IAC during interactive behaviour. Not only has the current thesis demonstrated that the effects of IAC upon the selection between microstrategies originally observed during simple routine copying and programming tasks (Fu & Gray, 2000; Gray & Fu, 2001; 2004; Gray et al., 2006) extend to more complex problem solving behaviour (Experiments 3.1 - 3.3, 4.1a & 4.1b of the current thesis), but small changes to IAC have now also been found to impact upon memory for task-relevant information. It is proposed that regular use of memory during problem solving when access costs were High accounted for the strengthening of Goal-State memory observed in the High IAC conditions in Experiment 4.2 (Anderson, Fincham, & Douglass, 1999; Anderson & Milson, 1989), and can be seen as a form of transfer-appropriate processing (Morris et al., 1977) or germane cognitive load (van Merriënboer et al., 2002) when such information is later to be recalled. It is important to acknowledge that the effect of IAC upon the recollection of problem solving information is unlikely to be a result of hard constraints imposed by design characteristics, but rather, soft constraints biasing stronger reliance upon memory during problem solving in the High IAC condition. Although participants were informed that they would be interrupted at various points and memory for the primary task would be assessed, it is not known whether motivating participants to remember more could have achieved the same benefits as the High IAC condition.

The replicated observation that participants in the Low IAC condition persisted to interleave regular inspections of the Goal-State with problem solving moves during solution to trial ten may explain the inferior memory for Current-State information observed at trial nine when access costs were Low. Participants in the High IAC condition appeared content solving the final trial with attention focussed firmly on making moves in the Current-State window (largely without reference to the Goal-State). Not having to divide attention between the Goal- and Current-State may account for the superior Current-State memory observed in the High IAC condition at trial nine (see Baddeley, Lewis, Eldridge, & Thomson, 1984; Craik, Govoni, Naveh-Benjamin, & Anderson, 1996 for useful reviews on the effects of divided attention on recall).

Higher access costs also improved participant recollection of plans. This result, therefore, suggests that IAC may also be used as a design characteristic to improve prospective memory for plans and goals (Altmann & Trafton, 2002; Brandimonte et al., 1996; Einstein & McDaniel, 2005), as well as retrospective memory for problem states (Delaney et al., 2004).

The observation that no effects of interfering task were observed on Goal- or Current-State memory means that few inferences can be made with regard to the particular encoding strategies for the different IAC conditions. The pattern-recollection and move-sequence task appeared to have equal affect on IAC conditions. However, the lack of a control interruption condition meant that it was not possible to determine *whether*, and if so, *how* disruptive the interference tasks were to memory for problem solving information. Inclusion of a control interruption condition where participants were interrupted using the same performance criteria, but assessed on their memory for Goal- and Current-State information immediately (without interference task), may have provided a comparable baseline measure for memory for Goal- and Current-State information. This would have allowed evaluation of how disruptive the interference tasks were, and would have also provided insight into immediate working memory for visually presented information.

Overall, the problem solving strategies adopted by participants in response to different access costs were in line with previous experiments reported within Chapter 4 and work conducted by Gray and colleagues (Gray & Fu, 2004; Gray et al., 2006). Despite the fact that problem solving was interrupted mid-way through trials two, five and nine, participants in the current study again demonstrated stronger reliance upon memory as a function of increased IAC, and more display-based problem solving when access costs were Low. This may represent tentative evidence that the effects of IAC extend to interruption-prone environments. Further discussion of this prospect as an area of future research can be found in Chapter 5.

4.6. General Discussion

Experiments 4.1a, 4.1b and 4.2 provide support for the hypothesis that small increments in IAC can be used to facilitate learning during repeated problem solving. As observed in Chapter 3, participants working with lower access costs (operationalised by a saccadic eye movement) persisted with the use of display-based strategies, whereas higher costs associated with accessing Goal-State information (additional mouse movement and 2.5 second delay) promoted stronger reliance upon memory during problem solving. With regard to learning during repeated solution of an eight-puzzle-like BPST problem, higher access costs led to partial improvements in problem solving efficiency (Experiments 4.1a), better transfer between similar problems (Experiment 4.1b), and improved memory for Goal-State, Current-State and planning information when assessed at various points during the problem solving process (Experiment 4.2). Each of these novel findings will be discussed in turn and then evaluated with respect to the hypotheses developed within the introduction.

4.6.1. IAC as a marker of expedited learning

Generally speaking, improvements in problem solving efficiency during repeated solution of the same eight-puzzle-like BPST problem were seen in each of the IAC conditions, supporting Anzai & Simon's (1979) theory of learning by doing. However, in Experiment 4.1a learning across the ten trials was observed to be more reliable with higher access costs, thus providing support for the assertion that learning during problem solving may be improved when memory is used more often (Sweller, 1988; Sweller & Levine, 1982). This conclusion, however, must be taken with some caution as this result was not statistically replicated in Experiment 4.1b. Further experimentation is required to assess the validity of the effect of IAC upon learning observed in Experiment 4.1a.

Higher access costs were also found to facilitate transfer from one eight-puzzle-like BPST problem to another. Parallels can be drawn between this finding, and that of Vollmeyer et al (1996), whereby reductions in the extent to which problem solving was directed by regular comparisons between the Goal- and the Current-State improved transfer to a novel problem. An increase in the cost associated with accessing Goal-State information (Experiment 4.1b) increased the extent to which participants mentally evaluated moves to be made (rather than using the display to evaluate moves to be made), and thus improved learning during problem solving. Unlike the problem encountered by O'Hara & Payne (1998) when interpreting the effects of increased implementation cost on improved transfer, the effect of IAC on transfer was not complicated by differences in the time spent attempting to solve the eight-puzzle-like BPST problem during practice.

Unfortunately, no measures were taken of the strategies used to solve the novel problem in Experiment 4.1b. All participants solved the novel problem with a Low

IAC interface and no eye-tracker data was collected. Therefore, no inferences can be made with respect to the influence of IAC upon the selection between memory- and display-based strategies during solution to the novel problem. For example, it is not known to what extent the memory-intensive microstrategies adopted by participants working with High access costs during the practice phase transferred to solution of the novel Low IAC problem. However, first-move latency data did suggest that no differences during solution to the novel problem between IAC conditions were apparent in terms of Davies's (2003) original conceptualisation of initial planning.

In terms of the transfer literature, the results of Experiment 4.1b appear generally in line with Sweller and colleague's assertion that internal processes that make use of memory during problem solving are required for effective transfer between problems (Sweller, 1988; Sweller & Levine, 1982). Although at first glance the frequency with which comparisons are made between the Goal- and Current-State in the Low IAC condition appear suggestive of weak problem solving methods such as means-end analysis (Anderson, 1982; 1987), it is difficult to make inferences with regard to the use of such strategies without further experimentation. However, it can be concluded that frequently interleaving inspections of the Goal-State between moves during problem solving can inhibit learning in situations such as the one used here.

Increasing the cost of inspecting the Goal-State also improved memory for Goal-State, Current-State and planning information. When assessed following interruption, Goal-State memory was found to be superior, and the number of plans recalled improved significantly. In addition, Current-State information was found to be superior when access costs were High at trial nine. Such results are in line with theories of transfer- and materials-appropriate processing (McDaniel & Einstein, 1989; Morris et al., 1977), because the use of memory during problem solving was more commonplace when access costs were High. In a general sense, Chapter 4 also provides support for the recent acceptance that memory for goals is not perfect (Anderson & Lebiére, 1998), and is instead vulnerable to the general constraints of human memory (Anderson & Douglas, 2001; Hodgetts & Jones, 2006). Some support is provided for Altmann & Trafton's (2002) goal activation model, and in particular the assertion that memory for goals can be strengthened by increased processing.

4.6.2. Further issues

More attention needs to be paid to uncovering the reasons why small increments in IAC facilitated transfer of problem solving to a novel eight-puzzle-like BPST in Experiment 4.1b (following repeated solution of a similar problem from the same family), but did not facilitate transfer between similar, but different eight-puzzle-like BPST problems in Experiment 3.3. Most likely, it is envisaged that superior transfer was observed in Experiment 4.1b because of the increased opportunity to learn the common methods used to solve the eight-puzzle (Ericsson, 1974a) when solving the same problem repeatedly. Solving a series of similar eight-puzzle-like BPST problems with different surface characteristics may have prevented effective learning of the general method used to solve the eight-puzzle. These studies should be taken as an initial investigation into the potential use of small access costs to improve learning and transfer during problem solving. A more detailed analysis of the problem solving strategies used to solve the eight-puzzle-like BPST would provide useful insight into the potential impact of IAC upon the use of weak problem solving methods such as means-end analysis (Anderson, 1987). Further research is required in this area.

Likewise, it seems important that future research considers the potential moderating factor of motivation. Since memory has, on occasion, been found to depend upon motivational factors (see Nilsson, 1987), a useful avenue of research may seek to explore the possibility that the High IAC effects could have been achieved by simply instructing participants that by committing the Goal-State to memory they would not have to keep referring to it, and would thus find the task easier. Although it is possible that such instructions would have prompted participants in the Low IAC condition to use memory more often, and thus negated the need for a High IAC manipulation, it is important to note that a major benefit of using IAC to promote the use of memory is that this non-deliberate selection (see Gray & Fu, 2004) is induced by interface design, and is not dependent upon extrinsic factors such as instructions (that can be forgotten). With regard to the effect of IAC upon recollection of problem solving information (Experiment 4.2), it is also important to note that, as in Chapter 2, participants were instructed that their memory for elements of the primary task would be assessed at various stages (see Appendix M). Finally, the results reported in Chapter 4 regarding the effect of IAC upon memory for problem solving information are limited in the sense that no measure of memory for past solutions was taken. Previous work has found memory for past solutions to be poor when assessed in tasks such as the Tower of Hanoi (Karat, 1982), Missionaries and Cannibals (Reed, Ernst, & Banerji, 1974) and Water Jars (Delaney et al., 2004). It would be theoretically interesting to explore the possibility that small increments to IAC may also be used to improve memory for past solutions of a problem.

In summary, increasing the cost associated with accessing information from the interface may provide a means of improving learning and memory of problem solving information (see also the findings reported in Chapter 2), and transfer from one problem solving task to another. Indeed, the use of memory to chunk large problem states is often cited as responsible for expert performance in a number of domains, including chess (Chase & Simon, 1973; De Groot, 1966). A new line of research may wish to address the possibility that access costs seeded appropriately into interface design could be used as a means of ameliorating the often negative effects of task interruption (e.g., Edwards & Gronlund 1998; Einstein, McDaniel, Williford, Pagan, & Dismukes, 2003). The development of a more elaborate internal representation of the problem space due to higher access costs may go some way to facilitating recovery from interruption. A discussion of this concept as a line of further research can be found in Chapter 5.

Chapter 5

General Discussion

5.1. Aims of the thesis

In response to the deluge of information currently available to operators of complex systems, such as those found in the modern aircraft cockpit, much applied attention has been focussed upon developing methods of increasing the accessibility of information provided within the interface (see Bennett & Flach, 1992; Vicente, 2002; Wickens & Hollands, 2000 for useful reviews). However, the topic has not received commensurate attention within relevant fields such as cognitive and psychological science, where recent work suggests that increasing the accessibility of information provided within the interface may lead to excessive reliance upon the display as an external memory source (Gray & Fu, 2004; Gray, Simms, Fu, & Schoelles, 2006). The present thesis, therefore, endeavoured to redress this issue, demonstrating that unwanted side effects can occur with design methods used to increase the accessibility of information provided with the interface.

5.2. Summary of key findings

First and foremost, the current thesis demonstrates that externally supported human behaviour is highly sensitive to small changes associated with the accessibility of information provided within the interface. Secondly, a number of unwanted side effects associated with increasing the accessibility of information provided within an interface have been observed (concerning memory, planning and learning). Prior to discussion of the relevant contributions to psychological theory and applied practice, each of the key findings will be summarised in turn.

5.2.1. Improved memory for fused information with reduced availability of onscreen information

One of the major concerns outlined in Chapter 1 associated with the use of integrated display design and information fusion more generally, was that human operator behaviour may become overly characterised by reliance upon the display as an external memory source (Gray & Fu, 2004; Gray et al., 2006; O'Regan, 1992; Parasuraman & Riley, 1997). One potential negative consequence of this might be problematic retention of visual-spatial information. The thrree experiments contained within Chapter 2 provided empirical support for this assertion.

Experiments 2.1 and 2.2 demonstrated that fusion of information improved pilots' ability to use onscreen information to estimate the position of a number of locations during simulated flight missions, relative to the provision of unfused information. Thus, much work demonstrating the benefits of integrated display design was supported (e.g., Vicente, Moray, Lee, Rasmussen, Jones, Brock, & Djemil, 1996; Woods, 1991). More interesting, in terms of the goals of the current thesis however, was the finding that the fused display also yielded a disproportionate rate of forgetting, when compared to the unfused display.

In an attempt to uncover the processes responsible for this effect and improve the retention of visual-spatial information when working with a fused display, Experiment 2.3 manipulated the temporal availability of fused information provided onscreen during flight missions. Perhaps counter intuitively, reducing the availability of onscreen information improved memory for fused information tested ninety seconds following each flight mission. Importantly, this was not at the expense of disadvantages experienced during flight missions and is likely to be due to stronger reliance upon internal processes (such as memory and inference making) during flight missions. In fact, reduced availability also had the unexpected benefit of improving location accuracy during flight performance.

5.2.2. The deployment of memory & planning during problem solving as effected by the cost of accessing information

A more detailed analysis of the effects of small changes to the cost of accessing information from an interface during problem solving is reported in Chapter 3, where three experiments examined the selection between microstrategies during solution to several different Blocks Problem Solving Task (BPST) problems. In particular, the use of memory and planning as a function of small changes to the cost of accessing goal-state information was examined using both quantitative (e.g., inter-move latencies, time and moves to solution) and qualitative measures (eye-tracking and verbal protocol data).

Experiments 3.1 and 3.2 found small changes to information access cost (IAC) to have large consequences for the deployment of memory and planning during problem solving. Essentially, small increments in IAC led to stronger reliance upon internal, as opposed to external, representations during problem solving, extending previous work conducted with routine copying (Fu & Gray, 2000; Gray et al., 2006) and programming tasks (Gray & Fu, 2001; 2004). The results obtained using a relatively simple BPST were also found to generalise to a more complex eight-puzzle-like BPST used in Experiment 3.3, and the nature of planful behaviour was observed to change according to small differences in IAC (Davies, 2003).

5.2.3. The effect of goal-state accessibility on learning during solution to an eight-puzzle-like BPST

Based upon work positing that the extent to which problem solving is display- or memory-based will affect learning (e.g., Sweller, 1988; Sweller & Levine, 1982), three experiments contained within Chapter 4 found evidence for the hypothesis that small increments in IAC (in line with those studied in Chapter 3) could facilitate learning during problem solving. During repeated solution of an eight-puzzle-like BPST problem, Experiments 4.1a and 4.1b found that participants working with lower access costs persisted with the use of display-based strategies, whereas higher costs associated with accessing goal-state information prompted stronger reliance upon memory during problem solving. Although only partial improvements in problem solving proficiency were observed during the ten trials (Experiment 4.1a), Experiment 4.1b provided strong support for the hypothesis that higher access costs would facilitate transfer of learning to a novel but similar problem. Finally, Experiment 4.2 found that higher access costs improved memory for goal-state, current-state and planning information at various points during repeated solution to an eight-puzzle-like BPST problem. This effect, it is argued, is a direct consequence of the regular use of memory during problem solving, as induced by higher access costs.

5.3. Contributions to psychological theory

5.3.1. Soft constraints hypothesis

The experimental chapters contained within this thesis provide much additional support for the 'soft constraints hypothesis' of interactive behaviour (see Gray & Fu, 2004). Gray & Boehm-Davis (2000) demonstrated that changes in the realm of milliseconds to the design of an interface can manifest substantial consequences for the selection between different microstrategies. Following from this, Gray and colleagues (Gray et al., 2006) have begun to establish the role of soft constraints in bounding the use of memory during a routine copying task, the Blocks World Task (BWT), originally developed by Ballard, Hayhoe, & Pelz, (1995). Focus has been upon changes to information accessibility, and essentially, as the cost associated with accessing information from the BWT increased, a shift in strategy was observed from an interaction-intensive reliance upon external representations, to memory-intensive reliance upon internal representations. Despite this reliable finding during routine low-level tasks (Ballard et al., 1995; Fu & Gray, 2000; Gray & Fu, 2001; 2004; Gray et al., 2006), no notable attempt had previously been made to extend this work to the problem solving domain (cf. Fu & Gray, 2006).

Using an externally supported problem solving task, the BPST, Chapter 3 found substantial evidence that least-effort tradeoffs concerning the use of memory observed during routine copying were also evident during problem solving. Small changes to the cost associated with accessing goal-state information reliably led to substantial differences between the selection of interaction- and memory-intensive microstrategies, with higher access costs inducing more memory-dependent problem solving, and lower access costs encouraging display-based problem solving. The current research thus adds weight to the adaptive view of human memory (Anderson & Milson, 1989; Anderson & Schooler, 1991; 2000), and the growing perspective that studies of human memory need to consider its properties as a reflection upon the structure of the environment it is used within.

Furthermore, the soft constraints imposed by changes to IAC reported within Chapter 3 were not found only to bound the use of memory, but also influenced the selection between planning styles (Davies, 2003). This is discussed at length in

Section 5.3.2. Chapter 4 also extended the analysis of soft constraints to a learning environment, whereby participants were required to repeatedly solve the same BPST problem ten times. The effects of IAC upon the selection between memory- and display-based problem solving search strategies were not found to be limited to scenarios within which participants were required to solve a series of different problems. Rather, very similar consequences were observed within situations where considerable learning would be expected (see Section 5.3.3).

5.3.2. Initial & concurrent planning

Recent advances within the rational analysis framework (Anderson, 1990) have also demonstrated that the extent to which individuals engage in planful behaviour can be based upon cost-benefit considerations of the task structure (e.g., O'Hara & Payne, 1998). However, no attempts previous to the current thesis had been made to elucidate the role of the task structure in influencing selection between different *forms* of planning (with the exception of task complexity – Davies, 2003). Nor had any attempt been made to evaluate the effect of information accessibility upon planning behaviour. In a recent synthesis of the planning literature, Davies (2003) made an important distinction between what he termed 'initial' and 'concurrent' planning. Both are seen as important forms of planning, but each make different demands upon working memory. According to Davies (2004), initial planning occurs prior to the solution of a problem, and is thus evaluated mentally, whereas concurrent planning occurs 'on-line', with external representations of the task space playing a substantial role during evaluation of plans. Parallels drawn between the demands of these two types of planning upon memory and the effect of IAC upon memory during problem solving (see Chapter 3) led to the prediction that small changes to information accessibility may affect the nature of planning deployed.

Quantitative analysis of participants' inter-move latencies in Experiments 3.2 and 3.3 in line with Davies (2003), and qualitative analysis of concurrent verbal protocol collected in Experiment 3.3 in line with Ericsson (1975), suggested that lower access costs encouraged more display-based concurrent planning, whereas higher access costs promoted more memory-based initial planning. Furthermore, as detailed in Chapter 3, an extension of Davies's (2003) original conceptualisation of initial planning was deemed necessary in order to capture the 'episodes' of memory-based planning observed when access costs were high.

The results reported within Chapter 4 also advance our understanding of initial and concurrent planning, by providing the first evaluation of the effect of each upon learning. Support was found for the assumption that learning during problem solving is, in part, determined by the extent to which problem solving is display/memory-based (O'Hara & Payne, 1998; Sweller, 1988; Sweller & Levine, 1982). Given that initial planning is largely memory-based, whereas concurrent planning is predominantly display-based, it was perhaps not surprising to find that the selection of an initial planning style (prompted by High access costs) led to more reliable learning during repeated solution of an eight-puzzle-like BPST problem (Experiment 4.1a), better transfer of problem solving knowledge from one problem to another (Experiment 4.1b), and superior memory for problem solving information (Experiment 4.2). Although a concurrent display-based planning strategy appeared capable of supporting problem solving when participants were required to solve a series of different BPST problems (Chapter 3), it provided fewer affordances for learning during repeated solution of the same eight-puzzle-like BPST problem in

Chapter 4. This finding is a novel contribution to the planning literature, and may be of great interest to the applied community where the same probem often needs to be solved repeatedly (Delaney, Ericsson, & Knowles, 2004).

5.3.3. Performance versus learning

One common theme residing within this thesis is that individuals may attempt to maximise short-term performance at the expense of longer-term learning. This potential tradeoff is not unique to the current discussion, and has been explicitly entrenched within psychological theories such as 'transfer-appropriate processing' (Morris, Bransford, & Franks, 1977), 'cognitive load theory' (Sweller, 1988), the transfer literature from Woodworth onwards (e.g., Thorndike & Woodworth, 1901), and 'levels of processing' since the seminal work of Craik & Lockhart (1972).

With regard to transfer-appropriate processing, it has been argued that matching the cognitive demands experienced during learning with those observed at retrieval will give rise to better retention of skill than mismatched learning and retrieval conditions. Thus, if the cognitive demands experienced as a result of maximising short-term task performance do not match those required for long-term retention, poor learning will be observed. Relatedly, recent perspectives on cognitive load theory have stipulated that so-called 'germane cognitive load' can facilitate learning (van Merriënboer, Schuurman, de Croock, & Paas, 2002), provided that the mental effort directed towards learning does not exceed the available processing capacity (which may also be affected by 'intrinsic' and 'extraneous' cognitive load).

Support for the assertion that optimising behaviour to facilitate short-term performance can, under certain circumstances, decrease long-term learning can be found in Chapters 2 and 4 of the current thesis. In particular, emphasis is placed upon the role of information accessibility as a moderator of this phenomenon. For example, Experiment 4.1b demonstrated that when information was highly accessible during externally supported problem solving, participants chose to rely heavily upon external (rather than internal) representations of the task space (presumably in an attempt to reduce overall cognitive load and the time required to solve the first few problems). This persistent reliance upon display-based problem solving was found to mask the development of generic problem solving schemata necessary for the transfer of knowledge from one eight-puzzle-like BPST problem to another. Similar over-reliance upon display-based problem solving also led to decrements in the long-term retention of problem solving information in Experiment 4.2. The general argument put forward in this thesis that making information harder to access can improve memory and learning is more closely aligned to incidental, as opposed to intentional, theories of learning and memory (see Postman, 1964).

These experiments add to the growing body of knowledge derived from a number of domains suggesting that relevant internal processing can improve subsequent task performance relative to passive reliance upon equivalent information provided within the environment (e.g., Duggan & Payne 2001; Glover, 1989; Lintern, 1980; McNamara, Kintsch, Songer, & Kintsch, 1996; Palmiter, Elkerton, & Bagget, 1991; Schmidt & Wulf, 1997). Indeed, the results summarised above are much aligned with Schmidt & Bjork's (1992) important review of the training literature. In this review, Schmidt & Bjork noted that in both the fields of motor and cognitive skill training, strategies that facilitated immediate performance would often lead to poor long-term retention or transfer. Effective training methods were identified as ones that introduced difficulties to the learner, and allowed for the practice of problems likely to be faced in transfer situations.

It is argued that the accessibility of information provided within the interface can be seen as a moderator of transfer-appropriate processing (Morris et al., 1977) and germane cognitive load (van Merriënboer et al., 2002). Although previous work has shown that concurrent feedback can act as a temporary crutch to performance and lead to subsequent decrement to retention of skill over time (e.g., Lintern, 1980; Patrick & Mutlusoy, 1982; Schmidt & Wulf, 1997), the experiments contained within Chapters 2 and 4 of the current thesis are the first to demonstrate the sensitivity that exists with respect to small changes to the accessibility of information provided within the interface. Furthermore, they allude to a novel approach in encouraging transfer of learning, which has historically been dogged by failed attempts and inconsistent findings (Stammers & Patrick, 1975). Rather than explicitly providing the problem solver with hints in an attempt to promote transfer of learning, the novel approach developed herein is able to utilise display design, and information access costs in particular, to encourage the problem solver to select certain behavioural microstrategies that are known to facilitate learning and transfer.

5.4. Implications for information fusion & integrated display design

The findings reported herein strongly suggest that more work is required from a Human Factors perspective when evaluating the effectiveness of integrated display design and information fusion technologies alike. Indeed, support is derived for the importance of paying close attention to the affordances provided by different cognitive artefacts within an operator's task space (Garbis, 2002), and how these may orient, and can be used to, support operator behaviour in the best possible way.

At no point during the current thesis is it suggested that we would be better off without integrated/fused approaches to display design. Indeed, the current findings do not take away from the many benefits observed when principles of information integration are put into practice (Bennett & Flach, 1992; Vicente, 2002; Wickens & Hollands, 2000). However, when considering the primary goal of increasing the accessibility of information provided within an interface, the current thesis does demonstrate that boundary conditions may apply, determining tradeoffs and situations within which increased accessibility to externalised information can have unwanted side effects for operator behaviour.

5.4.1. Consequences of 'mundane' changes to information accessibility

Complex system interface designers should be aware of the behavioural repercussions caused by very small, or in Gray & Boehm-Davis's (2000, p. 322) words "mundane", changes to the accessibility of information provided within an external display. Although most of the effects observed within the current thesis concern changes in information accessibility within the realms of two (Chapters 3 & 4) and eight seconds (Chapter 2), it is important to note that smaller manipulations to information accessibility have also been shown to have significant consequences for participant behaviour. For example, the frequency with which participants chose to inspect the goal-state during solution to a number of different BPST problems decreased significantly as a consequence of an increase in IAC from Low (saccadic eye movement to the goal-state) to Medium (saccadic eye movement plus mouse movement to the goal-state). This was true for all experiments contained within Chapters 3 and 4. Indeed, this had reliable consequences for the number of moves made per goal-state inspection, and thus can be said to have had a substantial effect upon the use of memory during problem solving. In addition, the same increase in IAC from Low to Medium significantly altered the distribution of inter-move latencies and the length of first-move latencies (Experiments 3.2 & 3.3) in a manner suggestive of initial planning strategies (Davies, 2003).

It is envisaged that a more detailed analysis of the consequences of small changes to information accessibility may further our understanding of the effective use of integrated display design and information fusion more generally. Although increasing the ease with which information can be extracted from an interface is often cited as the primary goal of integrated display design (Bennett & Flach, 1992; Vicente, 2002; Wickens & Carswell, 2000), it is anticipated that when developing methods for integrating and fusing information from multiple disparate sources, small changes to information availability and accessibility will often be overlooked. The current thesis suggests that complacency over such issues may lead to substantial and often counterintuitive repercussions for operator behaviour. The findings reported in the current thesis recommend that when information from different sources needs to be integrated in order to arrive at a decision, it is important that all information be accessible with minimum spatial or temporal displacement due to the limited capacity of working memory (Baddeley, 1986). However, if no boundaries are set upon the viewing of externally presented information, the current thesis also suggests that internal processing will be discouraged.

We know that *reducing* the availability of information provided within a fused display can actually *improve* operator performance during or after simulated flight missions (Experiment 2.3), and increasing the cost associated with accessing task-relevant information during problem solving can facilitate learning during problem solving (Experiments 4.1b and 4.2). Perhaps there are other paradoxical and under-researched issues surrounding information accessibility and display design? Indeed, it is worth bearing in mind the many caveats associated with the use of automation (see

Parasuraman & Riley, 1997 for a review). Recent research in this area has begun to explore the potential benefits of 'adaptive automation' (e.g., Hillburn, Jorna, Byrne, & Parasuraman, 1997; Kaber, Wright, Prinzel, & Clamann, 2005), and this may be a concept worth exploration in the field of information fusion and integrated display design (see Section 2.6.3 for a more in-depth discussion of this topic).

Further research needs to consider the possibility that, by encouraging stronger reliance upon memory during interactive behaviour, higher access costs may have a deleterious effect on cognitive resources such as concurrent attention and comprehension (Wickens & Hollands, 2000). There is a need to examine the effect of IAC on processes other than memory before implementing any delays to display access in contexts where multi-tasking is relevant. Indeed, it would also seem important to examine more closely how user experience may be affected by IAC. For example, one might predict that usability ratings would have been most favourable when information accessibility was highest in all experiments reported in this thesis (had participants been questioned in such a way). Before implementing access costs as a design feature for encouraging the use of memory during interactive behaviour, research would need to address how this may affect other aspects of user experience. One would certainly want to avoid potential frustration with the interface brought about by a perceived display-task mismatch. Figure 3.4 suggests that the use of memory varied more when IAC was High than Low in Experiment 3.1. Perhaps this is indicative of negative user experiences with a High access cost interface?

5.5. Limitations and caveats

Before proceeding to future directions and conclusions, a number of further limitations need to be raised that may impact upon the results reported within this thesis. For the purpose of brevity, specific issues concerned only with particular experiments discussed previously will not be reiterated here. Instead, a general discussion of important caveats can be found under the headings 'methodological issues' and 'restrictions of scope'.

5.5.1. Methodological issues

Firstly, although the majority of performance measures used within the current thesis appear relatively straightforward (e.g., location accuracy, time-on-task, moves-to-solution), one must always be cautious when generalising from behavioural data to cognitive processes. Perhaps the most straightforward measure used within the current thesis was that of location accuracy used in all experiments contained with Chapter 2. However, how far can one generalise from the measures of *recollection* reported in Chapter 2 (and in a similar fashion in Experiment 4.2) to processes of *recognition*? Before definitive recommendations can be made regarding the design of fused environments for the effective retention of visual-spatial information, research must address the potential influence of external cues within many applied tasks that may affect recognition memory (see Yonelinas, 2002 for an extensive review of the distinctions made between recollection and recognition memory).

Potentially more contentious, has been the extensive use of eye-tracking data to make inferences regarding the use of internal and external memory sources throughout Chapters 3 and 4, and the use of verbal protocol data to make inferences regarding planning style within Experiment 3.3. Many of the arguments concerning the use and misuse of verbal protocol have already been covered in Experiment 3.3, and thus will not be discussed again. However, it is worth noting that it is not possible to determine the precise validity of the protocols collected, the extent to which obtained given the materials chosen will generalise to other task domains. The first experimental chapter chose to base all experiments upon the Information Fusion Testbed (IFT) purely for ecological validity. However, inherent applied constraints meant that problems of representational equivalence (Larkin & Simon, 1987) surrounded the first experiment (2.1), and thus had to be removed in subsequent laboratory-based studies (Experiments 2.2 & 2.3). The BPST was developed for use in Chapters 3 and 4 in an attempt to effectively measure the consequences of manipulating IAC within a problem solving context. Without further experimentation, however, it is impossible to determine the extent to which the results observed using the BPST extend to more established problem solving tasks such as the Tower of Hanoi (e.g., Anderson & Douglass, 2001), or everyday problems such as compiling a manuscript using a typical word processing package with information located at a number of different sources (O'Hara, Taylor, Newman, & Sellen, 2002). It is currently not known to what degree, and by what manner, differences in the accessibility of information provided within the interface may affect other cognitive activities such as decision making (Payne, Bettman, & Johnson, 1988).

The experiments reported within this thesis have tended to adopt conservative between-subjects designs whenever possible (with random assignment to conditions) so as to minimise the influence of carry-over effects and asymmetric transfer (Poulton, 1982). Although all inferential statistics reported were scrutinised at the .05 alpha level using two-tailed hypotheses, there is always the possibility that pertinent results may have been overlooked (type II errors) due to a lack of experimental power (Cohen, 1988). Where planned comparisons were not fully justified, all paired comparisons were Bonferroni-corrected in order to reduce the chance of making type I errors.

5.5.2. Restrictions of scope

Additional restrictions of scope must also be noted. First and foremost, individual differences associated with completion of the IFT or BPST have not been addressed. It is quite probable that different participants will have approached these tasks in different ways, with certain individuals having preconceived preferences for specific strategies and methods (Davies, 2003; Roberts & Newton, 2001). Rather than examining individual differences, the current data have been used to develop average scores for each treatment condition. In addition to the limited analysis of behavioural data, it must also be noted that the observed effects of IAC have been couched at a rather general level. That is, contributions of time, physical and mental effort (all of which can be considered components of the heterogeneous variable IAC) were not separated. One previous study that has attempted to delineate the components of IAC (Gray et al., 2006) determined time to be the only important factor. Although not explicitly stated, time also features as an underlying theme within O'Hara & Payne's (1998; 1999) cost-benefit analysis of the implementation cost. The results reported within the current thesis also accommodate such a proposal, but do not provide definitive evidence either way. Furthermore, existing models of strategy selection based upon IAC (e.g., Gray et al., 2006) have thus far failed to acknowledge or control for the potential costs incurred via switches in attention between disparate sources of information (e.g., van der Heijden, 1992). Examination of such contributions may reveal consequences for access strategies based upon the nature of the material one is switching between.

Although examined to some extent, problem solving complexity was not manipulated successfully within task due to inherent difficulties in defining complexity within tasks presenting problem spaces as large as those found in the eight-puzzle-like BPST (Naylor & Briggs, 1963). For example, much 'element interactivity' (Ginns, 2006) and a lack of task constraints and structured subgoaling (Anderson, 1990) during solution to the BPST proved problematic in defining complexity. Davies (2003) used the more structured Tower of Hanoi (ToH) as a foundation for his exploration of the benefits of initial and concurrent planning under different levels of task complexity, and it is argued here that a more structured task, such as the ToH, would be required to effectively evaluate any interaction between the effects of IAC upon problem solving efficiency and task complexity.

One particular finding that carries much weight within Chapter 4, but requires further attention, is the improved transfer of knowledge from one eight-puzzle-like BPST problem to another when access costs were High following *repeated* solution to the same problem. Firstly, a fine-grained analysis of the use of memory and planning during transfer is required to provide much needed insight into the underlying mechanisms responsible for this transfer effect (discussed in Section 4.6). Secondly, further experimentation is needed to answer the question: would the observed effects of IAC upon transfer generalise to a different problem solving task? O'Hara & Payne (1998, Experiment 4) addressed this question with regard to manipulation of the implementation cost (IC). Despite finding that increasing IC during repeated solution to an eight-puzzle problem improved transfer to another eight-puzzle problem, no evidence of transfer from one problem solving task to another was obtained.

Finally, the current thesis is not able to comment upon Gray & Boehm-Davis's (2000) explicit proposal that individuals acquire rules that govern the selection between candidate microstrategies during previous experiences, which predicate a largely automatic process in the future. Assumptions of automaticity (at the neuronal or subsymbolic level) are also shared by Card, Moran, & Newell (1983), the soft constraints hypothesis (Gray & Fu, 2004) and cognitive theories of ACT-R (Anderson & Lebiére, 1998). Specific measurements of the selection between memory and planning strategies reported within the current thesis did not attempt to determine levels of deliberation or automaticity. Likewise, no comments were made on this topic by O'Hara & Payne (1998) when discussing the effects of the implementation cost on planning behaviour, or Davies's (2003; 2004) conceptualisation of the selection between initial and concurrent planning styles.

5.6. Future directions

In addition to the points of interest noted above, the current thesis also proposes two major research topics that require further attention; the externalisation of information to support human behaviour (Section 5.4.1) and the potential use of access costs to facilitate recovery from task interruption (Section 5.4.2).

5.4.1. Are external representations always helpful?

Research has been conducted in a number of domains detailing the benefits to human performance of providing an external representation of the task space (e.g., Larkin, 1989; Kotovsky, Hayes, & Simon, 1985; Zhang, 1997; Zhang & Norman, 1994; Valleé-Tourangeau & Penney, 2005). Indeed, much work has also focussed on developing methods for increasing the speed with which users can interact with, and extract information from, such devices (see Bennett & Flach, 1992; Hutchins, Hollan, & Norman, 1985; Wickens & Hollands, 2000 for useful reviews). However, very little work has begun to consider the potential caveats associated with the use of external representations as an aid to performance (cf. van Nimwegan, Oostendorp, & Tabachneck-Schijf, 2004; 2005). Indeed, within the context of graphical displays, Scaife & Rogers (1996) concluded that despite the "intuition" of the value of such external representations, "we have no well-articulated theory as to how such an advantage might work" (p. 200).

Coupled with previously reported decrements to performance associated with increasing the speed of interaction with an interface (O'Hara & Payne, 1998; 1999; Golightly, 1996; Svendsen, 1991), the current findings further suggest that the philosophy of designing an interface so as to speed interactions with, and extract information from, can be seen as an overly simple one (Payne, Howes, & Reader, 2001). The cognitive resources available for interacting with an interface are not always used most effectively (Svendsen, 1991). Instead, the allocation of cognitive resources can be seen as an optimisation to the particular task structure (Anderson, 1990). Therefore, in order to improve performance on a particular task, the solution will not always be as simple as externalising as much information as possible. Rather, effective design of human-interface interaction should encourage the use of desired behavioural strategies and discourage others. If learning of information provided within an interface is a requirement, behaviour should be directed towards the use of available mental resources, and less so towards the use of external representations of the task space. However, if fast interaction is of utmost importance, behaviour should be directed towards the use of perceptual-motor strategies, and therefore more externalisation of available information may be useful.

Put simply, there is a growing need for the Human Factors and Ergonomics communities to appreciate both sides of the coin: external representations of the task space can provide much support in the form of information, but can also detract from the development of internal representations of the task space. The application of this conclusion is not limited to the purposeful design of an interface to encourage certain behaviours, but may also prove a useful tool when evaluating the potential consequences of unavoidable costs to accessing task-relevant information. For example, when searching for raw data within a highly integrated interface, operators are often required to 'drill-down' into higher-order functional information to retrieve to physical data (Vicente, 2002). This action, ironically, has its own cost in terms of access time and effort and will thus require assessment.

5.6.2. The use of access costs to ameliorate the negative effects of task interruption

Reinstating intentions and memory following task interruption can be difficult (Einstein, McDaniel, Williford, Pagan, & Dismukes, 2003), and limited research has begun to examine methods by which task and interface design can support interruption tolerance (e.g., Oulasvirta & Saariluoma, 2006). Given the finding that small increments in IAC can facilitate the development of a robust internal representation of the task space (Experiments 4.1b & 4.2), one might argue that an obvious application of this effect would be to seed access costs appropriately within interface design in an attempt to temper the often negative effects of task interruption.

The use of access costs to facilitate the development of memory skills during interface interaction may provide some resilience to the negative effects of interruption (Edwards & Gronlund, 1998). Indeed, when interrupted at various points during repeated solution to an eight-puzzle-like BPST problem, Experiment 4.2 found participants working with higher access costs (within the primary task) to demonstrate superior memory for goal-state, current-state and planning information. Improved memory for goals and 'position-in-task' is likely to facilitate recovery from interruption in some form or another (e.g., faster resumption times – Trafton, Altmann, Brock, & Mintz, 2003; fewer procedural errors – Byrne, & Bovair, 1997).

The major advantage associated with the use of access costs to aid recovery from interruption would be that the structure of the interface would encourage the general development of a robust internal representation of the task space, and thus, the operator would be equipped to deal with interruptions more effectively when no time was available for preparation (Trafton et al., 2003) or negotiation of interruption onset (McFarlane, 2002), and the use of exogenous cues was not practical (McDaniel, Einstein, Graham, & Rall, 2004). In fact, there are currently mixed opinions regarding the effectiveness of each of these more established methods for combating task interruption (Morgan, 2005).

Contrary to the above suggestion that interruption tolerance may be supported by harnessing the interface so as to encourage the use of memory-based strategies, Larkin (1989) has posited the opposite: that one major advantage of display-based strategies may be supporting recovery from interruption. Although Larkin provided no empirical evidence for this assertion, she argued that the cues contained within an external display can direct the user towards the required behaviour following interruption. Task interruption is often likely to affect an individual's internal representation of a task, but will rarely affect the external representation of a task. It is clear that future work intending to disentangle the use of memory- and display-based strategies to ameliorate the negative effects of interruption will require careful consideration and intelligent experimental design. The first step, however, may be to replicate and extend the results reported in Experiment 4.2.

No control interruption condition was used in Experiment 4.2, and thus it is not possible to infer how 'immediate' memory would have compared to the obtained measures of 'delayed' memory. Interference type had no effect upon memory for goal- or current-state information. However, both interference tasks were highly similar to the primary BPST, and thus further investigation of similarity of interruption, complexity of interruption and length of interruption would likely provide further insight into the processes responsible for improved memory for task-relevant information in the High IAC condition (Hodgetts & Jones, 2006). In addition, future experiments should return participants to the primary task following interruption and gather measures of resumption lag, procedural errors and prospective memory (Altmann & Trafton, 2002; Brandimonte, Einstein, & McDaniel, 1996; Einstein & McDaniel, 2005).

An equally worthy future direction would be to investigate the extent to which the effects of IAC observed during routine copying (Fu & Gray, 2000; Gray et al., 2006), programming tasks (Gray & Fu, 2001; 2004) and problem solving (Chapters 3 & 4 of the current thesis) extend to analogous, yet 'interruption-prone' environments. Would the selection between microstrategies be affected, for example, by the frequency with which one is interrupted during task performance? As frequency increases, one might expect to see adaptation (Anderson, 1990). For example, regular interruptions may invoke more internalisation on behalf of operators working with lower access costs in the hope that more robust internal representations of the task space may facilitate recovery from interruption.

Frequent interruptions experienced by operators working with higher access costs may prompt the strategic decision to internalise less, if users felt relying upon information provided within the external display was a better method of tolerating frequent interruption. Although not designed to answer such questions, Experiment 4.2 provides preliminary tentative evidence that the effects of IAC observed during Experiment 4.1a (interruption-free) extend to an analogous, yet interruption prone, problem solving environment. However, it is anticipated that larger manipulations of interruption frequency (Monk, 2004) will be required in order to discover the extent to which the selection between memory- and display-based strategies contingent upon IAC is affected by task interruption and multi-tasking (Adams, Tenney, & Pew, 1995; Monsell, 2003).

It is envisaged that a variety of other methods may exist by which tweaks to the design of an interface could discourage over-reliance upon the display as an external memory source (O'Regan, 1992) and facilitate the development of a robust internal representation of the task space. Manipulations of information accessibility have been the current focus (Experiment 2.3, 4.2). However, intuition suggests that previously discussed manipulations to the implementation cost (O'Hara & Payne, 1998; 1999) may also lead to improved memory for task-relevant information. As of yet, this has not received empirical assessment.

5.7. Conclusions

The current thesis provides the first detailed analysis of the effects of increasing the accessibility of information provided within an interface upon memory, planning and learning. Much support is derived for the 'soft constraints hypothesis' of interactive behaviour (Gray & Fu, 2004; Gray et al., 2006), within which small changes to the cost of accessing information lead to surprisingly large consequences for the selection between different behavioural strategies. The current thesis found the use of memory and planning to be influenced by variations upon IAC within the realms of milliseconds (Gray & Boehm-Davis, 2000). Without contesting the many established benefits of integrated display design and information fusion more generally (see Bennett & Flach, 1992; Vicente, 2002; Wickens & Hollands, 2000 for useful reviews), a number of negative side effects are detailed associated with increasing the accessibility of information provided within an interface.

At first glance, many of the findings appear counterintuitive. However, theories of 'transfer-appropriate processing' (Morris et al., 1977) and 'germane cognitive load' (van Merriënboer et al., 2002) are proposed to account for the results. It is argued that when information is highly accessible within an interface, over-reliance upon the display as an external memory source (O'Regan, 1992) may mask the development of a robust internal representation of the task space. Advances are made to our understanding of the adaptive use of memory (Anderson, 1990; Anderson & Milson, 1989; Anderson & Schooler, 1991) and planning during problem solving (Davies, 2003; O'Hara & Payne, 1998). Least-effort tradeoffs determining reliance upon internal and external representations of the task space contingent upon small changes in IAC, previously reported with a routine copying task (Fu & Gray, 2000), have been extended to a similar but more complex problem solving task (Chapter 3).

Overall it is argued that designers of complex system displays must exercise more caution when manipulating the accessibility of information provided within the interface. In particular, there is a growing need for the Human Factors and Ergonomics communities to appreciate the perceived tradeoff between the use of external representations to provide the operator with highly accessible information, and the associated risk in terms of the problematic development of internal representations of the task space. Importantly, the reported findings have been observed during both static (Chapters 3 & 4) and dynamic presentation of visual information (Chapter 2), thus strengthening the case for careful consideration, regardless of medium (Chandler, 2004), of the cognitive and behavioural factors associated with manipulations of information accessibility.

REFERENCES

Adams, M. J., Tenney, Y. J., & Pew, R. W. (1995). Situation awareness and the cognitive management of complex systems. *Human Factors, 37*, 85-104.

Ahlum-Heath, M. E., & Di Vesta, F. J. (1986). The effect of conscious controlled verbalization of a cognitive strategy on transfer in problem solving. *Memory and Cognition, 14*, 281-285.

Altmann, E. M., & Gray, W. D. (2002). Forgetting to remember: The functional relationship of decay and interference. *Psychological Science, 13*, 27-33.

Altmann, E. M., & Trafton, G. J. (1999). Memory for goals: An architectural perspective. In M. Hahn & S. C. Stones (Eds.). *Proceedings of the 21st Annual Conference of the Cognitive Science Society*, 19-24. Hillsdale, NJ: Erlbaum.

Altmann, E. M., & Trafton, G. J. (2002). Memory for goals: An activation-based model. *Cognitive Science, 26*, 39-83.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*, 369-406.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge MA: Harvard University Press.

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

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.

Anderson, J. R., Bothell, D., & Douglass, S. (2004). Eye movements do not reflect retrieval processes. *Psychological Science, 15*, 225-231.

Anderson , J. R., & Douglass, S. (2001). Tower of Hanoi: Evidence for the cost of goal retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 1331-1346.

Anderson, J. R., Fincham, J. M., & Douglass, S. (1999). Practice and retention: A unifying analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1120-1136.

Anderson, J. R. & Lebiére, C. (1998). *Atomic components of thought.* Hillsdale, NJ: Erlbaum.

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review, 96*, 703-719.

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science, 2*, 396-408.

Anderson, J. R., & Schooler, L. J. (2000). The adaptive nature of memory. In E. Tulving & F. I. M. Craik (Eds.). *Handbook of human memory*, 557-570. New York: Oxford University Press.

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

Atwood, M. E., & Polson, P. G. (1976). A process model for water jug problems. *Cognitive Psychology, 8*, 191-216.

Baddeley, A. D. (1986). *Working memory.* Oxford, England: Clarendon Press.

Baddeley, A. D., Lewis, V., Eldridge, M., & Thomson, N. (1984). Attention and retrieval from long-term memory. *Journal of Experimental Psychology: General, 13*, 518-540.

Bainbridge, L. (1987). Ironies of automation. In J. Rasmussen, K. Duncan, & J. Leplat (Eds.). *New technology and human error* (pp. 271-283). New York: Wiley.

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience, 7*, 66-80.

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioural and Brain Sciences, 20*, 723-767.

Banbury, S., & Tremblay, S. (Eds.) (2004). *A cognitive approach to situation awareness: Theory and application*. Aldershot: Ashgate.

Bennett, K. B., & Flach, J. M. (1992). Graphical displays: Implications for divided attention, focussed attention, and problem solving. *Human Factors, 35*, 71-97.

Bennett, K. B., Payne, M., Calcaterra, J., & Nittoli, B. (2000). An empirical comparison of alternative methodologies for the evaluation of configural displays. *Human Factors, 42*, 287-298.

Berardi-Coletta, B., Buyer, L. S., Dominowski, R. L., & Rellinger, E. R. (1995). Metacognition and problem solving: A process-oriented approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 205-223.

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin, (Eds). *From learning processes to cognitive processes: Essays in honour of William K. Estes*, pp. 35-67. Hillsdale, NJ: Lawrence Erlbaum Associates.

Brandimonte, M., Einstein, G. O., & McDaniel, M. A. (eds.). (1996). *Prospective memory: Theory and applications*. Mahaw, NJ: Erlbaum.

Brinkman, J. A. (1993). Verbal protocol accuracy in fault diagnosis. *Ergonomics, 36*, 1381-1397.

Brown, J. S., Collins, A., & Duguid, S. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*, 32-42.

Buttigieg, M. A., & Sanderson, P. M. (1991). Emergent features in visual display design for two types of failure detection tasks. *Human Factors, 33*, 631-651.

Byrne, M. D., & Bovair, S. (1997). A working memory model of a common procedural error. *Cognitive Science, 21*, 31-62.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction*. London: Erlbaum.

Carswell, C. M., & Wickens, C. D. (1987). Information integration and the object display: An interaction of task demands and display superiority. *Ergonomics, 30*, 511-527.

Cary, M., & Carlson, R. A. (1999). External support and the development of problem solving routines. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1053-1070.

Cary, M., & Carlson, R. A. (2001). Distributing working memory resources during problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 836-848.

Chandler, P. (2004). Commentary: The crucial role of cognitive processes in the design of dynamic visualizations. *Learning and Instruction, 14*, 353-357.

Chandler, P. and Sweller, J. (1996). Cognitive load while learning to use a computer program. *Applied Cognitive Psychology, 10*, 151-170.

Chapanis, A. (1988). Some generalizations about generalization. *Human Factors, 30*, 253-267.

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

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37-46.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cohen, P. R. (1995). *Empirical methods for artificial intelligence*. Cambridge, MA: MIT Press.

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

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behaviour, 11*, 671-684.

Craik, F. I. M., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval in human memory. *Journal of Experimental Psychology: General, 125*, 159-180.

Curiel, J. M, & Radvansky, G. A. (1998). Mental organization of maps. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 202-214.

Dasarathy, B. V. (2001). Information Fusion – what, where, why, when, and how? *Information Fusion, 2*, 75-76.

Davies, S. P. (2003). Initial and concurrent planning in solutions to well-structured problems. *The Quarterly Journal of Experimental Psychology, 56A(7)*, 1147-1164.

Davies, S. P. (2004). Planning and problem solving in well-defined domains. In R. Morris, & G. Ward (Eds.). *The cognitive psychology of planning*. Hove: Psychology Press.

De Groot, A. (1966). Perception and memory versus thought: Some old ideas and recent findings. In B. Kleinmuntz (Ed.). *Problem solving*. New York: Wiley.

Delaney, P. F., Ericsson, A. K., & Knowles, M. A. (2004). Immediate and sustained effects of planning in a problem-solving task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 1219-1234.

Dinadis, N., & Vicente. K. J. (1999). Designing functional visualizations for aircraft systems status displays. *International Journal of Aviation Psychology, 9*, 241-269.

Duggan, G. B., Banbury, S., Howes, A., Patrick, J., & Waldron, S. M. (2004). Too much, too little or just right: Designing data fusion for situation awareness. In *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting*, 528–532.

Duggan, G. B., & Payne, S. J. (2001). Interleaving reading and acting while following procedural instructions. *Journal of Experimental Psychology: Applied, 7*, 297-307.

Edwards, M. B., & Gronlund, S. D. (1998). Task interruption and its effects on memory. *Memory, 6*, 665-687.

Einstein, G. O., & McDaniel, M. A. (2005). Prospective memory: Multiple retrieval processes. *Current Directions in Psychological Science, 14*, 286-290.

Einstein, G. O., McDaniel, M. A., Williford, C. L., Pagan, J. L., & Dismukes, R. K. (2003). Forgetting of intentions in demanding situations is rapid. *Journal of Experimental Psychology: Applied, 9*, 147-162.

Ericsson, K. A. (1974a). *Problem-solving behaviour with the 8-puzzle I: Time to solution*. Report No. 431. Department of Psychology, University of Stockholm.

Ericsson, K. A. (1974b). *Problem-solving behaviour with the 8-puzzle II: Distribution of latencies*. Report No. 432. Department of Psychology, University of Stockholm.

Ericsson, K. A. (1974c). *Problem-solving behaviour with the 8-puzzle III: Process in terms of latencies*. Report No. 433. Department of Psychology, University of Stockholm.

Ericsson, K. A. (1975). *Instruction to verbalise as a means to study problem-solving processes with the 8-puzzle: A preliminary study*. Report No. 458. Department of Psychology, University of Stockholm.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). Cambridge, MA: MIT Press.

Fleck, J. I., & Weisberg, R. W. (2004). The use of verbal protocols as data: An analysis of insight in the candle problem. *Memory & Cognition, 32*, 990-1106.

Fu, W.-T., & Gray, W. D. (2000). Memory versus perceptual-motor tradeoffs in a blocks world task. In L.R. Gleitman and A.K. Joshi (Eds.). *Proceedings of the 22nd Annual conference of the Cognitive Science Society*, 154-159. Hillsdale, NJ: Erlbaum.

Fu, W.-T., & Gray, W. D. (2006). Suboptimal tradeoffs in information seeking. *Cognitive Psychology, 52*, 195-242.

Gagne, R. M., & Smith, E. C. (1962). A study of the effects of verbalisation on problem solving. *Journal of Experimental Psychology, 63*, 12-18.

Garbis, C. (2002). Exploring the openness of cognitive artifacts in cooperative process management. *Cognition, Technology & Work, 4*, 9-21.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Gick, M., & Holyoak, J. (1987). The cognitive basis for knowledge transfer. In S. Cormier & J. Hagman (Eds.). *Transfer of learning* (pp. 9-46). San Diego, CA: Academic Press.

Gilhooly, K. J. (2004). Working memory and planning. In R. Morris, & G. Ward (Eds.). *The cognitive psychology of planning*. Psychology Press.

Gilhooly, K. J., Phillips, L. H., Wynne, V., Logie, R. H., & Della Sala, S. (1999). Planning and age in the five-disc tower of london. *Thinking and Reasoning, 5*, 339-361.

Ginns, P. (2006). Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction, 16*, 511-525.

Glaser, R., & Bassok, M. (1989). Learning theory and the study of instruction. *Annual Review of Psychology, 40*, 631-666.

Glover, J. A. (1989). Improving readers' estimates of learning from text: The role of inserted questions. *Reading Research Quarterly, 28*, 68-75.

Goldberg, J. H., Stimson, M. J., Lewenstein, M., Scott, N., & Wichansky, A. M. (2002). Eye tracking in web search tasks: Design implications. In *Proceedings of the 2002 Symposium on Eye Tracking Research & Applications*, 51-58. New York: ACM Press.

Golightly, D. (1996). Harnessing the interface for domain learning. *Proceedings of CHI*, 37-8.

Goodstein, L. P. (1981). Discriminative display support for process operators. In J. Rasmussen & W. B. Rouse (Eds.). *Human detection and diagnosis of system failures* (pp. 433-449). New York: Plenum.

Gray, W. D., & Boehm-Davis, D. A. (2000). Milliseconds matter: An introduction to microstrategies and their use in describing and predicting interactive behaviour. *Journal of Experimental Psychology: Applied, 6*, 322-335.

Gray, W. D., & Fu, W.-T. (2001). Ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head: Implications of rational analysis for interface design. In *CHI Letters, 3(1)*, 112-119.

Gray, W. D., & Fu, W.-T. (2004). Soft constraints in interactive behaviour: The case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. *Cognitive Science*, 28, 359-383.

Gray, W. D., Simms, C. R., Fu, W.-T., & Schoelles, M. J. (2006). The soft constraints hypothesis: A rational analysis approach to resource allocation for interactive behavior. *Psychological Review*, *113*, 461-482.

Gregoriades, A. & Sutcliffe, A. S. (2006). Automated assistance for human factors analysis in complex systems. *Ergonomics, 49*, 1265-1287.

Gronlund, S. D., Dougherty, M. R. P., Durso, F., Canning, J. M., & Mills, S. H. (2005). Planning in Air Traffic Control: Impact of problem type. *The International Journal of Aviation Psychology, 15*, 269-293.

Hayes-Roth, B., & Hayes-Roth, F. (1979). A cognitive model of planning. *Cognitive Science, 3*, 275-310.

Hillburn, B., Jorna, P., Byrne, E. A., & Parasuraman, R. (1997). The effect of adaptive air traffic control (ATC) decision aiding on controller mental workload. In M. Mouloua & J. M. Koonce (Eds.). *Human-automation interaction: Research and practice* (pp. 84-91). Mahwah, NJ: Erlbaum.

Hodgetts, H. M., & Jones, D. M. (2006). Interruption of the tower of london task: Support for a goal-activation approach. *Journal of Experimental Psychology: General, 135*, 103-115.

Hollingworth, A. (2005). The relationship between online visual representation of a scene and long-term scene memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 396-411.

Howes, A., Vera, A., & Lewis, R. L. (in press). Bounding rational analysis: Constraints on asymptotic performance. In W. D. Gray (Ed.). *Integrated models of cognitive systems*. New York: Oxford University Press.

Hutchins, E. (1995a). *Cognition in the Wild*. Cambridge, MA: MIT Press.

Hutchins, E. (1995b). How a cockpit remembers its speed. *Cognitive Science, 19*, 265-288.

Hutchins, E. L., Hollan, J. D., & Norman, D. A. (1985). Direct manipulation interfaces. *Human-Computer Interaction, 1*, 311-338.

Hwang, M. I., & Lin, J. W. (1999). Information dimension, information overload and decision quality. *Journal of Information Science, 25*, 213-218.

Jones, P. M., Wickens, C. D., & Deutsch, S. J. (1990). The display of multivariate information: An experimental study of an information integration task. *Human Performance, 3*, 1-17.

Jeffries, R. P., Polson, P. G., Razran, L., & Atwood, M. (1977). A process model for missionaries-cannibals and other river-crossing problems. *Cognitive Psychology, 9*, 412-440.

Kaber, D. B., Wright, M. C., Prinzel, L. J., & Clamann, M. P. (2005). Adaptive automation of human-machine system information-processing functions. *Human Factors, 47*, 730-742.

Karat, J. (1982). A model of problem solving with incomplete constraint knowledge. *Cognitive Psychology, 14*, 538-559.

Kirsh, D. (2000). A few thoughts on cognitive overload. *Intellectica, 30*, 19-52.

Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic actions. *Cognitive Science, 18*, 513-549.

Kirwan, B. B., & Ainsworth, L. K. (1992). *A guide to task analysis*. New York: John Wiley & Sons.

Knowles, M. E., & Delaney, P. F. (2005). Lasting reductions in illegal moves following an increase in their cost: Evidence from river-crossing problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 670-682.

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from tower of hanoi. *Cognitive Psychology, 17*, 248-294.

Larkin, J. H. (1989). Display-based problem solving. In D. Klahr & K. Kotovsky (Eds.). *Complex information processing: The impact of Herbert A. Simon.* Hillsdale, NJ: Erlbaum.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*, 65-99.

Lee, J. D., & Moray, N. (1994). Trust, self confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies, 40*, 153-184.

Lind, M. (1990). Representing goals and functions in complex systems: An introduction to multilevel flow modelling. *Technical report, 90-D-380, Institute of Automatic Control Systems*, Technical University of Denmark, Lyngby, Denmark.

Lintern, G. (1980). Transfer of landing skill after training with supplementary visual cues. *Human Factors, 22*, 81-88.

Lintern, G., Waite, T., & Talleur, D. A. (1999). Functional interface design for the modern aircraft cockpit. *International Journal of Aviation Psychology, 9*, 225-240.

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

Lohse, G. L., & Johnson, E. J. (1996). A comparison of two process tracing methods for choice tasks. *Organisational Behaviour and Human Decision Processes, 68*, 28-43.

Luck, S. J., & Vogal, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*, 279-281.

Marino, C. J., & Mahan, R. P. (2005). Configural displays can improve nutrition-related decisions: An application of the proximity compatibility principle. *Human Factors, 47*, 121-131.

Matthews, G., Davies, D. R., Westerman, S. J., & Stammers, R. B. (2000). *Human Performance: Cognition, Stress and Individual Differences*. Hove: Psychology Press.

Mayes, J. T., Draper, S. W., McGregor, M. A., & Oatley, K. (1988). Information flow in a user interface: The effect of experience and context on recall of MacWrite screens. In D. M. Jones & R. Winder (Eds.). *People and computers IV*. Cambridge, England: Cambridge University Press.

McDaniel, M. A., & Einstein, G. O. (1989). Material-appropriate processing: A contextualist approach to reading and studying strategies. *Educational Psychology Review, 1*, 113-145.

McDaniel, M. A., Einstein, G. O., Graham, T., & Rall, E. (2004). Delaying execution of intentions: Overcoming the costs of interruptions. *Applied Cognitive Psychology, 18*, 533-547.

McGeorge, P., & Burton, A. M. (1989). The effects of concurrent verbalisation on performance in a dynamic systems task. *British Journal of Psychology, 80*, 455-465.

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*, 1-43.

McFarlane, D. C. (2002). Comparison of four primary methods for coordinating the interruption of people in human-computer interaction. *Human-Computer Interaction, 17*, 63-139.

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

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior*. New York: Holt, Rinehart, & Winston.

Monk, C. A. (2004). The effect of frequent versus infrequent interruptions on primary task resumption. In *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting*, 295-299.

Monsell, S. (2003). Task Switching. *Trends in Cognitive Sciences, 7*, 134-140.

Morgan, P. (2005). *"Now where was I?" A cognitive experimental analysis of the influence of interruption on goal-directed behaviour*. Unpublished D. Phil. Thesis, Cardiff University.

Morris, C.D., Bransford, J.D. & Franks, J.J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behaviour*, 16, 519-533.

Morton, J. (1967). A singular lack of incidental learning. *Nature, 215*, 203-204.

Mosier, K., Skitka, L., Burdick, M., & Heers, S. (1996). Automation bias, accountability, and verification behaviour. In *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting*, 204-208.

Mosier, K., Skitka, L., Heers, S., & Burdick, M. (1997). Automation bias: Decision making and performance in high-tech cockpits. *International Journal of Aviation Psychology, 8*, 47-63.

Naylor, J. C., & Briggs, G. E. (1963). Effects of task complexity and task organization on the relative efficiency of part and whole training methods. *Journal of Experimental Psychology, 65*, 217-224.

Neisser, U. (1976). *Cognition and reality.* San Francisco: Freeman.

Neth, H., & Payne, S. J. (2002). Thinking by doing? Epistemic actions in the tower of hanoi. In *Proceedings of the Twenty-Fourth Annual Conference of Cognitive Science Society*, 691-696. Mahwah, NJ: Erlbaum.

Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice Hall.

Nickerson, R. S., & Adams, M. J. (1979). Long-term memory for a common object. *Cognitive Psychology, 11*, 287-307.

Nilsson, L. (1987). Motivated memory: Dissociation between performance data and subjective reports. *Psychological Research, 49*, 183-188.

Norman, D. A. (1990). The 'problem' with automation: Inappropriate feedback and interaction, not 'over-automation'. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 327*, 585-593.

Norman, D. A. (1993). Cognition in the head and in the world: An introduction to the special issue on situated action. *Cognitive Science, 17*, 1-6.

O'Hara, K. P., & Payne, S. J. (1998). The effects of operator implementation cost on planfulness of problem solving and learning. *Cognitive Psychology, 35*, 34-70.

O'Hara, K. P., & Payne, S. J. (1999). Planning and the user interface: The effects of lockout time and error recovery cost. *International Journal of Human-Computer Studies, 50*, 41-59.

O'Hara, K. P., Taylor, A., Newman, W., & Sellen, A. J. (2002). Understanding the materiality of writing from multiple sources. *International Journal of Human-Computer Studies, 56*, 269-305.

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology, 46*, 461-488.

Ormerod, T.C. (2005). Planning and ill-defined problems. In R. Morris and G. Ward (Eds.): *The cognitive psychology of planning*. London: Psychology Press.

Oulasvirta, A., & Saariluoma, P. (2006). Surviving task interruptions: Investigating the implications of long-term working memory theory. *International Journal of Human-Computer Studies, 64*, 941-961.

Paas, F., & Kester, L. (2006). Learner and information characteristics in the design of powerful learning environments. *Applied Cognitive Psychology, 20*, 281-285.

Palmiter, S., Elkerton, J., & Bagget, P. (1991). Animated demonstrations vs. written instructions for learning procedural tasks: A preliminary investigation. *International Journal of Man-Machine Studies, 34*, 687-701.

Parasuraman, R. (1993). The effects of adaptive function allocation on human performance. In D. J. Garland & J. A. Wise (Eds.). *Human factors and advanced aviation technologies* (pp. 147-157). Daytona Beach, FL: Embry-Riddle Aeronautical University Press.

Parasuraman, R. (2000). Designing automation for human use: Empirical studies and quantitative models. *Ergonomics, 43*, 931-951.

Parasuraman, R., Mouloua, M., & Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems. *Human Factors, 38*, 665-679.

Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation induced "complacency". *International Journal of Aviation Psychology, 3,* 1-23.

Parasuraman, R., & Riley. V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*, 230-253.

Patrick, J. (1992). *Training: Research and practice*. London: Academic Press.

Patrick, J., & James, N. (2004). Process tracing of complex cognitive work tasks. *Journal of Occupational and Organizational Psychology, 77*, 259-280.

Patrick, J., & Mutlusoy, F. (1982). The relationship between types of feedback, gain of a display and feedback precision in acquisition of a simple motor task. *Quarterly Journal of Experimental Psychology, 34A*, 171-184.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 534-552.

Payne, S. J. (1991). Display-based action at the user interface. *International Journal of Man-Machine Studies, 35*, 275-289.

Payne, S. J., Howes, A., & Reader, W. R. (2001). Adaptively distributing cognition: A decision-making perspective on human-computer interaction. *Behaviour & Information Technology, 20*, 339-346.

Payne, S. J., Richardson, J., & Howes, A. (2000). Strategic use of familiarity in display-based problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1685-1701.

Pawlak, W. S., & Vicente, K. J. (1996). Inducing operator control through ecological interface design. *International Journal of Human-Computer Studies, 44*, 653-688.

Pfeiffer, T. (2004). Problem solving with a simple transformation problem with and without continuous external support. *European Journal of Cognitive Psychology, 16*, 555-572.

Phillips, L. H., Wynn, V., Gilhooly, K. J., Della Sala, S., & Logie, R. H. (1999). The role of memory in the tower of london task. *Memory, 7*, 209-231.

Phillips, L. H., Wynn, V., McPherson, S., & Gilhooly, K. J. (2001). Mental planning and the tower of london task. *Quarterly Journal of Experimental Psychology, 54A*, 579-597.

Pirolli, P. & Card, S. (1999). Information foraging. *Psychological Review, 106*, 643-675.

Postman, L. (1964). Short-term memory and incidental learning. In A. W. Melton (Ed.). *Categories of human learning* (pp. 146-201). New York: Academic Press.

Poulton, E. C. (1982). Influential companions: Effects of one strategy on another in the within-subjects designs of cognitive psychology. *Psychological Bulletin, 9*, 673-690.

Powell, G., Marshall, D., Milliken, R., & Markham, K. (2004). A data fusion system for object recognition based on transferable belief models and kalman filters. In *Proceedings of the 7th International Conference on Information Fusion.*

Ramsey, P. H. (1982). Empirical power of procedures for comparing two groups on *p* variables. *Journal of Educational Statistics, 7*, 139-156.

Rasmussen, J. (1983). Skills, rules, and knowledge: Signals, signs, symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man and Cybernetics, SMC-13*, 257-266.

Rasmussen, J. (1985). The role of hierarchical representation in decision making and system management. *IEEE Transactions on Systems, Man and Cybernetics, SMC-15*, 234-243.

Rasmussen, J., & Vicente, K. J. (1989). Coping with human errors through system design: Implications for ecological interface design. *International Journal of Man-Machine Studies, 31*, 517-534.

Ratterman, M. J., Spector, L., Grafman, J., Levin, H., & Harward, H. (2001). Partial and total-order planning: Evidence from normal and prefontally damaged populations. *Cognitive Science, 25*, 941-975.

Reed, S. K., Ernst, G. W., & Banerji, R. (1974). The role of analogy in transfer between similar problem states. *Cognitive Psychology, 6*, 436-450.

Reising, D. V. C., & Sanderson, P. M. (2004). Minimal instrumentation may compromise failure diagnosis with an ecological interface. *Human Factors, 46*, 316-333.

Roberts, M. J., & Newton, E. J. (2001). Understanding strategy selection. *International Journal of Human-Computer Studies, 54*, 137-154.

Ross, B. H. (1984). Remindings and their effects in learning a cognitive skill. *Cognitive Psychology, 16*, 371-416.

Ross, B. H. (1987). This is like that: The use of earlier problems and the separation of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 629-639.

Ross, B. H. (1989). Distinguishing types of superficial similarities: Differing effects on the access and use of earlier problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 456-468.

Sarter, N. B., & Woods, D. D. (1995). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human Factors, 37*, 5-19.

Scaife, M., & Rogers, Y. (1996). External cognition: How do graphical representations work? *International Journal of Human-Computer Studies, 45*, 185-213.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207-217.

Schmidt, R. A., & Wulf, G. (1997). Continuous concurrent feedback degrades skill learning: Implications for training and simulation. *Human Factors, 39*, 509-526.

Schooler, J. W., Ohlsson, S., & Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology: General, 122*, 166-183.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63*, 129-138.

Simon, H. A. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology, 7*, 268-288.

Simon, H. A. (1981). *Sciences of the artificial* (2^{nd} Ed.). Cambridge, MA: MIT Press.

Simon, H. A., & Reed, S. K. (1976). Modelling strategy shifts in a problem-solving task. *Cognitive Psychology, 8*, 86-97.

Singley, M., & Anderson, J. R. (1989). *The transfer of cognitive skill.* Cambridge, MA: Erlbaum.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74*, (Whole No. 498).

Stammers, R. B., & Patrick, J. (1975). *The psychology of training.* London: Metheun.

Sutcliffe, A. G. (1995). *Human-computer interface design.* London: Macmillan Press Ltd.

Svendsen, G. B. (1991). The influence of interface style on problem solving. *International Journal of Man-Machine Studies, 35*, 379-397.

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

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*, 185-233.

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

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

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th edition). Boston: Allyn & Bacon.

Thorndike, E. L. (1903). *Educational psychology*. New York : Lemke and Buechner.

Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review, 8*, 247-261.

Trafton, G. J., Altmann, E. M., Brock, D. P., & Mintz, F. E. (2003). Preparing to resume and interrupted task: Effects of prospective goal encoding and retrospective rehearsal. *International Journal of Human-Computer Studies, 58*, 583-603.

Trudel, C., & Payne, S. J. (1995). Reflection and goal management in exploratory learning. *International Journal of Human-Computer Studies, 42*, 307-339.

Vallée-Tourangeau, F., & Penney, A. K. (2005). The impact of external representation in a rule discovery task. *European Journal of Cognitive Psychology, 17*, 820-834.

van der Heijden., A. L. C. (1992). *Selective attention in vision*. New York: Routledge.

van Gog, T., Pass, F., van Merriënboer, J. J. G., & Witte, P. (2005). Uncovering the problem-solving process: Cued retrospective reporting versus concurrent and retrospective reporting. *Journal of Experimental Psychology: Applied, 11*, 237-244.

van Merriënboer, J. J. G., Schuurman, J. G., de Croock, M. B. M., & Paas, F. G. W. C. (2002). Redirecting learner's attention during training: Effects on cognitive load, transfer test performance and training efficiency. *Learning and Instruction, 12*, 11-37.

van Nimwegen, C., van Oostendorp, H., & Tabachneck-Schijf, H. (2004). Externalization vs. Internalization: The influence on problem solving performance. In K. Chee-Kit Looi, E. Sutinen, D. Sampson, I. Aedo, L. Uden, & E. Kahkonen (Eds.). *Proceedings of the 4th IEEE International Conference on Advanced Learning Technologies* (pp. 311-315). Los Alamitos, CA: IEEE Computer Society.

van Nimwegen, C., van Oostendorp, H., & Tabachneck-Schijf, H. (2005). The role of interface style in planning during problem solving. In B. Bara, L. Barsalou, & M. Bucciarelli (Eds.). *Proceedings of the 27th Annual Cognitive Science Conference* (pp. 2271-2276). Mahwah, NJ: Lawrence Erlbaum.

Venturino, M. (1997). Interference and information organization in keeping track of continually changing information. *Human Factors, 39*, 532-538.

Vicente, K. J. (1992). Memory recall in a process control system: A measure of expertise and display effectiveness. *Memory & Cognition, 20*, 356-373.

Vicente, K. J. (2002). Ecological interface design: Progress and challenges. *Human Factors, 44*, 62-78.

Vicente, K. J., & Rasmussen, J. (1992). Ecological interface design: Theoretical foundations. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-22,* 589-606.

Vicente, K. J., Moray, N., Lee, J. D., Rasmussen, J., Jones, B. G., Brock, R., & Djemil, T. (1996). Evaluation of a rankine cycle display for nuclear power plant monitoring and diagnosis. *Human Factors, 38*, 506-521.

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,* 75-100.

Whitfield, D., Ball, R. G., & Ord, G. (1980). Some human factors aspects of computer-aiding concepts for air traffic controllers. *Human Factors, 22*, 569-580.

Wickens, C. D. (2002). Situation awareness and workload in aviation. *Current Directions in Psychological Science, 11*, 128-133.

Wickens, C. D., & Carswell. C. M. (1995). The proximity compatibility principle: Its psychological foundation and relevance to display design. *Human Factors, 37*, 473-494.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering Psychology and Human Performance* (3rd Ed.). New Jersey: Prentice-Hall Inc.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*, 625-636.

Woods, D. D. (1991). The cognitive engineering of problem representations. In J. Alty & G. Weir (Eds.). *Human-computer interaction in complex systems* (pp.169-188). London: Academic.

Woods, D. D., Patterson, E. S., & Roth, E. M. (2002). Can we ever escape from data overload? A cognitive systems diagnosis. *Cognition, Technology, & Work, 4*, 22-36.

Woods, D. D., Wise, J., & Hanes, L. (1981). An evaluation of nuclear power plant safety parameter display systems. In *Proceedings of the Human Factors 25th Annual Meeting*, 110-114. Santa Monica, CA: Human Factors Society.

Xiong, N., & Svensson, P. (2002). Multi-sensor management for information fusion: Issues and approaches. *Information Fusion, 3*, 163-186.

Yantis, S. (1993). Stimulus driven attentional capture. *Current Directions in Psychological Science, 2*, 156-161.

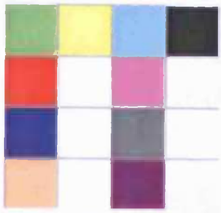
Yantis, S., & Jonides, J. (1996). Attentional capture by abrupt onsets: New perceptual objects or visual masking? *Journal of Experimental Psychology: Human Perception and Performance, 22*, 1505-1513.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*, 441-517.

Zhang, J. (1997). The nature of external representations in problem solving. *Cognitive Science, 21*, 179-217.

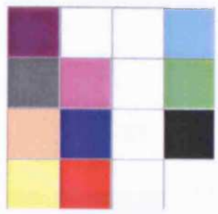
Zhang. J., & Norman, D. A. (1994). Representations in distributed cognitive tasks. *Cognitive Science, 18*, 87-122.

Appendix A: The twelve different BPST Goal- and Start-States used in Experiments 3.1 & 3.2.

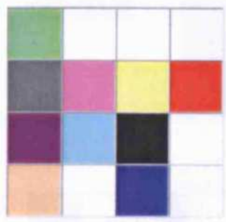

1 Goal

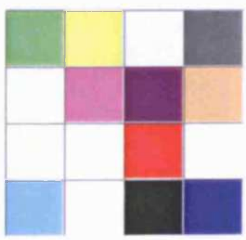

Start

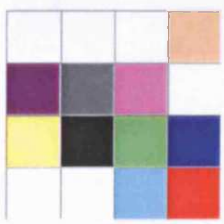

2 Goal


Start


3 Goal

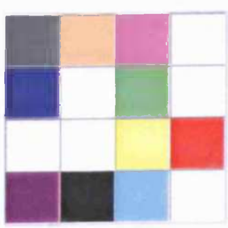

Start


4 Goal


Start


5 Goal


Start


6 Goal

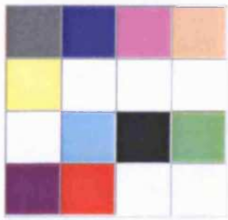

Start


7 Goal


Start


8 Goal


Start


9 Goal


Start


10 Goal


Start


11 Goal


Start


12 Goal


Start

Appendix B: The six different 17 move eight-puzzle-like BPST Goal- and Start-States used in Experiment 3.3.

1 Goal

Start

2 Goal

Start

3 Goal

Start

4 Goal

Start

5 Goal

Start

6 Goal

Start

Appendix C: The two different 17 move BPST Start-States used in Experiments 4.1a & 4.1b (same Goal-State), and the 17 move Goal- and Start-State for the novel transfer problem used in Experiment 4.1b.

Goal-State

Start-State A

Start-State B

Novel Goal-State

Novel Start-State

Appendix D: The two different 17 move eight-puzzle-like BPST Goal- and Start-States used in Experiment 4.2.

Goal-State

Start-State

Problem A

Goal-State

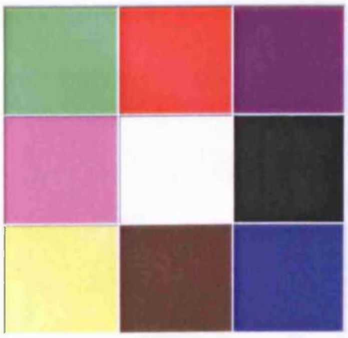
Start-State

Problem B

Appendix E: The three configurations used in the pattern-recollection interference task in Experiment 4.2.

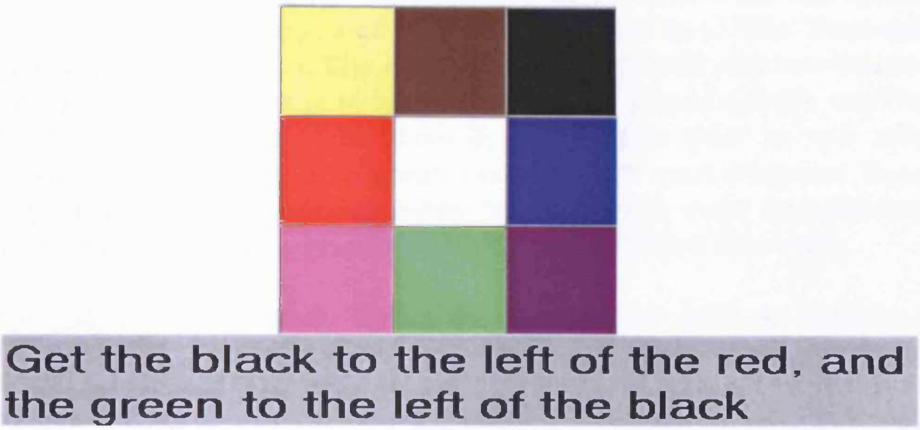
1

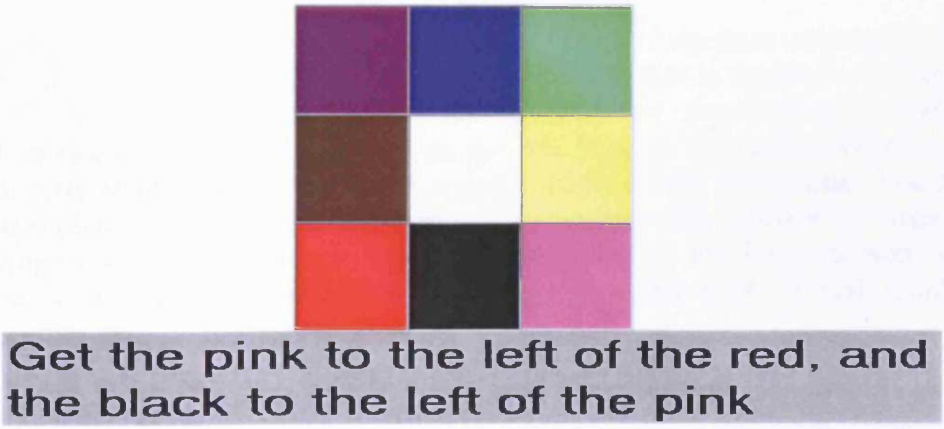
2

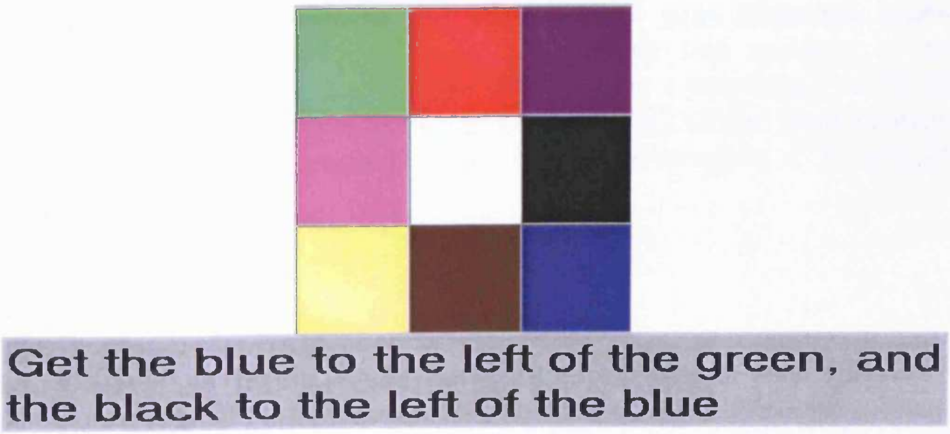


3

Appendix F: The three Start-States and the objectives to be achieved in the move-sequence interference task used in Experiment 4.2.







Appendix G: Experiment 2.1 Instructions.

You are required to complete a number of different scenarios flown using the Information Fusion Testbed (IFT). The scenarios are based on a mission against one, two or three sites within a designated Area of Responsibility (AOR). Your task is to estimate the location of the sites. The number and location will vary between one, two or three for each trial. The task is to be undertaken by a pair of aircraft, one flown by you as the lead pilot and the other flown by our resident 'pilot' as your wingman. During the trial please refrain from communicating with your wingman. During the scenarios, your task will change depending on the display under investigation. You will receive verbal instructions describing each scenario before they begin.

Your main task is to accurately locate the sites in the AOR as quickly and as accurately as you can. You will be advised of the AOR, as the terrain surrounding the AOR will appear greyed out. It is essential that you to stay within the designated AOR, as flying outside of this area will cause the trial to be restarted. You should try to fly the task at a constant speed of 600 Kts and at 15,000ft.

We advise you to fly a square pattern within the AOR to maximise coverage by the 40 nautical mile radius of the Radar. In addition, when fusion is enabled, see Display 2 below, the IFT tracking accuracy can be enhanced by maintaining a steady and continuous rate of change of sight-line to the site. It is, therefore, beneficial to avoid flying directly at any site, be that either a single entity or one of a cluster. You should try to complete the task with the aim of minimising the amount of time spent identifying sites. The time taken to identify sites will be recorded as a measure of your performance as will your accuracy in judgement of location to the 'actual' location of sites.

At regular intervals in distance from sites a 'RESPOND NOW' prompt will appear on screen. When this happens you are required to respond using the touch screen display to indicate where you think the sites are located. Please make all responses as quickly and accurately as possible. When you have entered your responses, please say 'responses entered' to acknowledge to the experimenter that you have entered the location(s). In some cases you may not be able to make a judgement on the location and therefore, please refrain from touching the screen and let the experimenter know that there are 'no responses entered' Once the trial is complete a 'TRIAL OVER' prompt will appear.

Formats

- Display 1

This display, known as the non-fused display, presents information from your Radar. The radar spokes indicate the direction, but not range, of the location(s). The spokes will remain on the screen for two seconds. In the non-fused display, there will be a facility to share information between platforms. However, you must make your judgement as to the location of the sites by using the information presented by the radar on the on-screen map. Please note that some spokes may overlap and appear as one. You should check that the spokes are not overlapping to help you judge the number of sites you are trying to locate.

- Display 2

In the fused display, the sharing of information between platforms is possible, as for Display 1 above but, in this case, an indication of the site location(s) will be depicted in a track or tracks generated by the IFT. The tracks will appear upon the on-screen map indicating the estimated location of each and every site. Please note that the system may not be entirely accurate in its estimates of site locations.

Throughout the trial, the map scale will be kept constant.

Questionnaires

To help us gain an understanding of your situational awareness whilst using the different displays, you will be asked to complete a Situational Awareness Rating Technique (SART) questionnaire at the end of each trial. Situational Awareness refers to your knowledge and understanding of the situation or context associated with your current task.

SART uses subjective estimates of personal and task dependant factors which affect task performance and understanding to give a measure of your situational awareness.

We also want to gain a subjective measure of your workload relating to the tasks. You will be asked to complete a DRA Workload Scale (DRAWS) questionnaire. The questions will ask you to rate the demand imposed by different aspects of the task you have just completed. The demands are related to the reception of information necessary to do the task, the performance of mental operations on the information, making a response and the time pressure associated with the task. You are required to enter your response by moving the bar underneath the statement to a position that you believe to be indicative of the workload associated with the trial you have just completed.

Trial running order

At the start of the trial you will be given the opportunity to familiarise yourself with the task and different data fusion methods under investigation through completing a set of practice trials. If you have any questions regarding the task, please raise them at this stage of the trial, as the experimenter will not be able to help you during the main trial.

Once the practice trials are complete you will start the main experiment. The experimenter will inform you at the beginning of each scenario which display you will be given. At the end of each trial, you will be asked to complete the SART and DRAWS questionnaires. Your responses should only relate to the trial that you have just completed.

Once you have completed all of the conditions for the trial, you will be given a debrief by the experimenter and will be asked to fill out a post-test questionnaire. This will provide you with an opportunity to raise any issues you have during the trials and provide us with feedback.

Appendix H: Experiment 2.2 Instructions.

This experiment requires you to complete a number of different trials where you are required to guide an aircraft over a map and locate where a number of sites are situated. You must only guide your aircraft within your Area of Responsibility. If you move out of your Area of Responsibility you will be alerted by a blue pop-up message in the bottom right hand corner of your display telling you to "Keep within your Area of Responsibility". If this happens, simply guide the aircraft back onto the map. In the real thing there will always be 1 or 3 sites on the map. However, in the 3 practice trials, examples using 2 sites will be used.

Your task is to estimate the location of the sites and click on the map where you think they are located using the left mouse button when prompted to do so. You will be asked to do this 4 times per trial and after the 4th estimation the trial will finish. Each of the 4 prompts in each trial refer to the location of the same site(s). However, the location of the site(s) changes from trial to trial. To guide the aircraft, press the left and right arrow keys on the keyboard. These will move the aircraft a certain degree left and right over the map each time you press them. There will be 3 practice trials for you to get familiar with guiding the aircraft.

Description of main display

There will always be 2 aircraft (represented by green circles) on the screen. You are guiding the darker green circle (the blue line indicates the direction in which you are travelling in). The lighter green circle is a confederate aircraft and will always fly within a close separation to your aircraft. You do not have to guide this aircraft. The reason for the second aircraft is so that the 2 aircraft can share information in relation to the whereabouts of the site(s). You can use information from both aircraft to make your judgements as to where you think the site(s) are located.

In all conditions information regarding the whereabouts of the site(s) from your own aircraft will be displayed on the Radar Display. This is shown as spokes representing the direction the site(s) are estimated to be in relation to your aircraft. The Radar display does not include information from the confederate aircraft.

Fused conditions

In the fused conditions red squares will appear on the map estimating where the 2 aircraft believe the site(s) to be. This information is only available when the 2 aircraft are within a certain distance to the site(s) and is not always 100% accurate. Although the red squares give you an estimation of where the 2 aircraft believe the site(s) to be, it is up to you to use this information how you wish to decide upon the locations of the site(s).

Non-Fused conditions

In the non-fused conditions spokes will fire from each aircraft on the map in the direction they each estimate the site(s) to be. This information is only available when the 2 aircraft are within a certain distance to the site(s) and is not always 100% accurate. Although the red spokes give you an estimation of where the 2 aircraft believe the site(s) to be, it is up to you to use this information how you wish to decide upon the locations of the site(s).

Response required

The crucial point to remember is that when a red box appears covering both the aircraft and prompting you to "Click where you think the site(s) are located" you must do so immediately and then click on the red box to the right of the map that says "Response Entered" as quickly as possible. Please only click on the map when prompted to do so and only left click as many times as you think there are site(s). For example, if you thought there were 3 sites on a given trial and you were prompted to "Click where you think the site(s) are located" you should then click three times on the map as close as possible to where you think they were situated at that moment in time and then click on the "Response Entered" button straight away. This will happen 4 times per trial and needs to be repeated for each trail as accurately and quickly as possible. It is important to note that you should always be aiming to guide the aircraft as close as possible to each site. Each trial will finish when you get close enough to the site(s) on that given trial. You should complete the task with the aim of minimising the amount of time spent identifying site locations. Each trial is different. Do not use any information from a previous trial to help locate site(s) in another trial.

Questionnaire and memory test

At the end of each trial a screen will appear telling you to fill out a questionnaire. The questionnaire is the same for each trial but you must answer each question with regard to your previous trial only. The questionnaires are provided in paper format. Once you have completed the questionnaire you must click on the button on the screen to proceed to the next trial. Before the next trial starts however, a copy of the map that you were guiding your aircraft over in the previous trial will appear on the screen. On this map you are required to click (left mouse button) where you think the site(s) were located in the previous trial. Please only click as many times as you think there where sites and be as accurate as possible. If you have no idea, just guess. Once you have made your responses and have clicked on the button to proceed to the next trial a screen will appear showing the opening display of the next trial. Simply click on the start button and repeat the same procedure for the following trials.

General points

At the beginning of each trial simply let your aircraft fly north (see compass situated top right of your display) until information starts to appear on the Radar Display telling you which direction the site(s) are likely to be in. You can then use this information to fly towards and locate the site(s). As mentioned earlier, the information on the Radar Display is taken from your aircraft only. However, once both of the aircraft get close enough to the site(s) shared information from both aircraft will be displayed on the map itself either as red squares or red spokes.

Appendix I: Experiment 2.3 Instructions.

This experiment requires you to complete a number of different trials in which you are required to guide an aircraft over a map and locate where a number of sites are situated (always either 1 or 3). You must only guide your aircraft within your Area of Responsibility. If you move out of your Area of Responsibility you will be alerted by a blue pop-up message in the bottom right hand corner of your display telling you to "Keep within your Area of Responsibility". If this happens, simply guide the aircraft back onto the map.

Your task is to estimate the location of the sites and click on the map where you think they are located using the left mouse button when prompted to do so. You will be asked to do this 4 times per trial and after the 4th estimation the trial will finish. Each of the 4 prompts in each trial refer to the location of the same site(s). However, **the location of the site(s) changes from trial to trial**. To guide the aircraft, press the left and right arrow keys on the keyboard. These will move the aircraft a certain degree left and right over the map each time you press them. There will be 2 practice trials for you to get familiar with guiding the aircraft.

Description of main display

There will always be 2 aircraft (represented by green circles) on the screen. You are guiding the darker green circle (the blue line indicates the direction in which you are travelling in). The lighter green circle is a confederate aircraft and will always fly within a close separation to your aircraft. You do not have to guide this aircraft. The reason for the second aircraft is so that the 2 aircraft can share information in relation to the whereabouts of the site(s). You can use information from both aircraft to make your judgements as to where you think the site(s) are located.

In all conditions information regarding the whereabouts of the site(s) from your own aircraft only will be displayed on the Radar Display. This is shown as spokes representing the direction the site(s) are estimated to be in relation to your aircraft. The Radar display does not include information from the confederate aircraft.

Shared Information Regarding Site Locations

Red squares will appear on the map estimating where the 2 aircraft believe the site(s) to be. This information is only available when the 2 aircraft are within a certain distance to the site(s) and is not always 100% accurate. Although the red squares give you an estimation of where the 2 aircraft believe the site(s) to be, it is up to you to use this information how you wish to decide upon the locations of the site(s).

Response Required

The crucial point to remember is that when a red box appears covering both the aircraft and prompting you to "Click where you think the site(s) are located" you must do so immediately and then click on the red box to the right of the map that says "Response Entered" as quickly as possible. Please only click on the map when prompted to do so and only left click as many times as you think there are sites. For example, if you thought there were 3 sites on a given trial and you were prompted to

"Click where you think the site(s) are located" you should then click three times on the map as close as possible to where you think they were situated at that moment in time and then click on the "Response Entered" button straight away. This will happen 4 times per trial and needs to be repeated for each trail as accurately and quickly as possible. It is important to note that you should always be aiming to guide the aircraft as close as possible to each site. Each trial will finish when you get close enough to the site(s) on that given trial. You should complete the task with the aim of minimising the amount of time spent identifying site locations. Each trial is different. Do not use any information from a previous trial to help locate site(s) in another trial.

Questionnaire and memory test

At the end of each trial a screen will appear telling you to fill out a questionnaire. The questionnaire is the same for each trial but you must answer each question with regard to your previous trial only. The questionnaires are provided in paper format. As soon as the button on the screen turns green you must click the button straight away to proceed to the next trial. If you have not finished the questionnaire by the time the button turns green do not worry, press the button anyway and put the unfinished questionnaire to one side. If you finish the questionnaire before the button turns green, please wait until the button turns green before pressing it! Before the next trial starts, a copy of the map that you were guiding your aircraft over in the previous trial will appear on the screen. On this map you are required to click (left mouse button) where you think the site(s) were located in the previous trial. Please only click as many times as you think there where sites and be as accurate as possible. If you have no idea, just guess. Once you have made your responses and have clicked on the button to proceed to the next trial a screen will appear showing the opening display of the next trial. Simply click on the start button and repeat the same procedure for the following trials.

General points

At the beginning of each trial simply let your aircraft fly north (see compass situated top right of your display) until information starts to appear on the Radar Display telling you which direction the site(s) are likely to be in. You can then use this information to fly towards and locate the site(s). As mentioned earlier, the information on the Radar Display is taken from your aircraft only. However, once both of the aircraft get close enough to the site(s) shared information from both aircraft will be displayed on the map itself as red squares.

Appendix J: Experiment 3.1 and 3.2 Instructions.

The object of the task is to move the blocks in the Current-State Window around until they match the configuration of the blocks in the Goal-State Window.

Rules
Only one move can be made at a time by moving a coloured block into one of the empty white spaces. The coloured block must be adjacent to an empty white space to be moved.

Complete trial
Once you think you have matched the two patterns, click on the 'stop trial' button at the bottom of the screen. If the two configurations match, you will be taken to the next trial. If they do not match, a message will appear telling you 'incorrect, please try again'. If this appears, just go back to the task until you get it right.

General points
You will be required to complete a number of different problems. An eye-tracker will be used for this study.

Appendix K: Experiment 3.3 Instructions.

The object of the main task is to move the blocks in the Current-State Window around until they match the configuration of the blocks in the Goal-State Window.

Rules
Only one move can be made at a time by moving a coloured block into the currently empty white space. The coloured block must be adjacent to the empty white space to be moved.

Complete trial
Once you think you have matched the two patterns, click on the 'stop trial' button at the bottom of the screen. If the two configurations match, you will be taken to the next trial. If they do not match, a message will appear telling you 'incorrect, please try again'. If this appears, just go back to the task until you get it right.

General points
You will be required to complete a number of different problems. An eye-tracker will be used for this study.

Verbal protocol

During this study, we are interested in what you think about during the exercises. In order to do this I would like you to THINK ALOUD as you work on each exercise. What I mean by think aloud is that I want you to tell me EVERYTHING you are thinking from the time the exercise starts. I would like you to think aloud CONSTANTLY. I don't want you to plan what you say or try to explain to me what you are saying. Just act as if you were alone in the room speaking to yourself. If you are silent for any long period of time I will prompt you to think aloud.

Do you understand what I want you to try and do?

Now, we will begin with some think-aloud practice problems. First, I want you to add these 2 numbers in your head and tell what you are thinking as you get an answer. Don't worry about the answer you give me. I just want you to practice telling me what you are thinking as you try to work out what the result is of adding 215 to 318.

Now I want you to think aloud as you try to work out how many rooms are there in your house.

Appendix L: Experiment 4.1a and 4.1b Instructions.

The object of the task is to move the blocks in the Current-State Window around until they match the configuration of the blocks in the Goal-State Window.

Rules
Only one move can be made at a time by moving a coloured block into the currently empty white space. The coloured block must be adjacent to the empty white space to be moved.

Complete trial
Once you think you have matched the two patterns, click on the 'stop trial' button at the bottom of the screen. If the two configurations match, you will be taken to the next trial. If they do not match, a message will appear telling you 'incorrect, please try again'. If this appears, just go back to the task until you get it right.

General points
You will be solving the same problem several times over. An eye-tracker will be used for this study.

Appendix M: Experiment 4.2 Instructions.

The object of the primary task is to move the blocks in the Current-State Window around until they match the configuration of the blocks in the Goal-State Window. Only one move can be made at a time, by moving a coloured block into the currently empty white space. The coloured block must be adjacent to the empty white space to move into it.

Interruption
On occasion, you will be interrupted and taken away from the primary task . . .

Secondary task
Having been interrupted, you will be required to complete a quick secondary task lasting 30 seconds.

Memory task
After completion of the secondary task, your memory for the primary task pre-interruption will be assessed. You will be asked to recall the patterns in both the Goal- and Current-State Windows, and any immediate plans for moves you were going to make before you were interrupted. Once each of these memory tests have been conducted, you will be returned to the primary task, but starting the next trial afresh.

General points
You will be solving the same problem several times over. An eye-tracker will be used for this study.