LEABHARLANN CHOLÁISTE NA TRÍONÓIDE, BAILE ÁTHA CLIATH | TRINITY COLLEGE LIBRARY DUBLIN
Ollscoil Átha Cliath | The University of Dublin

**Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin**

**Copyright statement**

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

**Liability statement**

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

**Access Agreement**

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

# REASONING WITH INSIGHT PROBLEMS

Maire Aislinn Murray

PhD

University of Dublin, Trinity College

2005

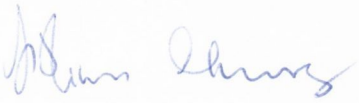
TRINITY COLLEGE
05 MAY 2005
LIBRARY DUBLIN
THESIS
7625

## DECLARATION

(a) The work contained in this thesis has not been submitted as an exercise for a degree at this or at any other University

(b) This thesis is the result of my own investigations, except where otherwise stated. The contributions of others are duly acknowledged in the text wherever included.

(c) I give my permission to the Library to lend or copy this thesis upon request.

Signed 

Date 9/3/05

## SUMMARY

This thesis reports the results of seven experiments that explore insight problem-solving. Insight problems are characterised as ill-defined problems that present difficulty to most people even though they possess the knowledge needed to solve the problem. Insight problems were first described in the early part of the twentieth century by Gestalt psychologists who emphasised that the answer tended to suddenly occur to people in an "aha experience".

In Chapter 1, we review early research into insight problems and we compare contemporary theories, which fall into two main sorts: theories that emphasise the role of mental representation, and theories that emphasise the role of search processes. We discuss the evidence that supports each of these alternative perspectives.

In Chapter 2, we investigate the role of mental representation in insight problem-solving. In three experiments, we examine the common errors people make when interpreting a range of insight problems, and we test simple manipulations that can reduce the negative effect of these errors on correct solution rates. We conclude that representation has an important role in insight problem-solving. However, we also show that it is unlikely to account for all of the difficulty experienced by participants.

In Chapter 3, we turn to the role of search space in determining the difficulty of insight problems. In two experiments we examine the effect of manipulations that limit the search space and that can improve correct solution rates. We conclude that limiting

search space helps problems for which the difficulty is one of choosing a correct move from many possible moves.

In Chapter 4, we report two experiments that investigate what skills are associated with the ability to solve insight problems. In one experiment, we focus on imagery and verbal skills and in the other on the executive functions of attention and working memory. We show that attention switching and working memory abilities are important for insight problem-solving.

In Chapter 5, we summarise the findings of the seven experiments and discuss their implications for understanding insight problem-solving. We consider the contribution of these experiments to understanding insight and their consequences for the main theories of insight. We also consider ideas for future research on insight problem-solving.

## ACKNOWLEDGEMENTS

I will be eternally grateful to my supervisor, Ruth Byrne, who was always generous with her considerable expertise and helped to elevate me from student to scholar, despite my own best efforts to the contrary at times. Neither would I have got this far without the support of my mother, Rose and my husband, Gerry who always believed I could do it, even when I did not believe it myself. I am also indebted to my friends and colleagues, Suzanne, Mary, Laura and Lorraine who along with giving me lots of practical help and support helped to keep me sane during the whole process. A special thank you to Laura McAvinue, Ian Robertson and Richard Roche for sharing with me their expertise on measures of imagery, attention and memory respectively. Thanks also to Michelle Flood and Gry Wester for their assistance with experimental coding and data entry. A more general thank you to all the people who listened to my ideas at various stages and gave me useful feedback, to all the participants who sat patiently through my experiments and to the staff in the Psychology Department, Trinity College who were always willing to help me in any way they could. Last, but not least, this project would not have been possible without the financial support of the Irish Research Council for the Humanities and Social Sciences, Enterprise Ireland and the Trinity College Arts and Social Sciences Benefaction Fund.

## TABLE OF CONTENTS


***CHAPTER 1*** . . . . . **1**

The Idea of Insight . . . . . 2

Theoretical Background to Insight Problem-Solving . . . . 8

Current Theories of Insight . . . . . 11

Properties of Insight Problems and Problem-Solvers . . . . 24

Aims of the Thesis . . . . . 27

***CHAPTER 2*** . . . . . **31**

**Experiment 1** . . . . . **35**

Method . . . . . 36

Results and Discussion . . . . . 38

**Experiment 2** . . . . . **43**

Method . . . . . 45

Results and Discussion . . . . . 46

**Experiment 3** . . . . . **52**

Method . . . . . 53

Results and Discussion . . . . . 55

**General Discussion** . . . . . **63**

***CHAPTER 3*** . . . . . **71**

**Experiment 4** . . . . . **75**

Method . . . . . 77

Results and Discussion . . . . . 79

**Experiment 5** . . . . . **84**

Method . . . . . 85

Results and Discussion . . . . . 87

**General Discussion . . . . . 93**

***CHAPTER 4* . . . . . 97**

**Experiment 6 . . . . . 106**

Method . . . . . 109

Results. . . . . 112

Discussion . . . . . 118

**Experiment 7 . . . . . 119**

Method . . . . . 122

Results. . . . . 128

Discussion . . . . . 137

**General Discussion . . . . . 137**

***CHAPTER 5* . . . . . 140**

Summary of Findings . . . . . 141

Implications . . . . . 159

Future Work . . . . . 165

Conclusion . . . . . 165

**REFERENCES . . . . . 167**

**APPENDIX A . . . . . 178**

**APPENDIX B . . . . . 193**

**APPENDIX C . . . . . 202**

**PUBLICATIONS . . . . . 226**

## List of Tables

**Table 2.1:** Percentages of correct solutions for single-step and multiple -step problems in Experiment 1 . . . . . 39

**Table 2.2:** Percentages of correct solutions for standard and improved wording of single- and multiple-step problems in Experiment 2 . . 48

**Table 2.3:** Percentages of correct solutions for paper-and-pencil and physical prop versions of the single-step and multiple-step problems in Experiment 3 after 10 minutes, with correct solutions after 2 minutes provided in parentheses . . . . . 57

**Table 2.4:** Mean latencies to produce the correct solutions in seconds (log transformed to the base e) for the problems in Experiment 3 in bold (with standard deviations in brackets) and pre-transformation means in italics . 58

**Table 2.5:** Mean latencies to correct solutions for problems solved on the first attempt in Experiment 3 (log transformed to the base e) with pre-transformation means in parentheses . . . . . 59

**Table 2.6:** Mean difficulty ratings for each problem in both conditions in Experiment 3 based on a scale of 1 – 3 where 1 = easy . . . 61

**Table 3.1:** Percentages of correct solutions for the no hint and hint versions of the single-step and multiple-step problems in Experiment 4 . . 80

**Table 3.2:** Mean latencies to correct solution for each problem in Experiment 4 (in seconds log transformed to the base e) with standard deviations in parentheses and pre-transformation means in italics . . 81

**Table 3.3:** Percentage of correct solutions for each problem in the three conditions in Experiment 5 . . . . . 88

**Table 3.4:** Percentage of 'good quality' solutions for each problem in the three conditions in Experiment 5 . . . . . 89

**Table 3.5:** Mean latencies to produce the correct solution measured in milliseconds (log transformed to the base e) with standard deviations in parentheses and pre-transformation means in italics for each problem in Experiment 5 . . . . . 91

**Table 4.1:** Mean scores, standard deviations and possible range of scores for each task in Experiment 6 . . . . . 112

**Table 4.2:** Correlations between correct scores on insight problems, overall and for the two sub-categories, and the other tasks in Experiment 6 . . 114

**Table 4.3:** Correlation matrix of non-insight measures in Experiment 6 . 117

**Table 4.4:** Mean scores, standard deviations and possible scoring range for all problem categories in Experiment 7 . . . . . 129

**Table 4.5:** Mean scores, standard deviations and possible scoring ranges, where applicable, for all attention measures in Experiment 7 . . 130

**Table 4.6:** Correlation between insight problems and measures of attention in Experiment 7 . . . . . 131

**Table 4.7:** Mean scores, standard deviations and possible scoring ranges, where applicable, for both working memory measures in Experiment 7 . 131

**Table 4.8:** Correlations between insight problems and measures of working memory in Experiment 7 . . . . . 132

**Table 4.9:** Summary table for results of stepwise regression in Experiment 7 with insight problem score as dependent variable . . . . 133

**Table 4.10:** Correlation between non-insight problems and measures of attention and working memory in Experiment 7 . . . . 135

**Table 4.11:** Correlations between the other measures, attention and working memory in Experiment 7 . . . . . 136

**Table 5.1:** Summary of percentage increase in solution rates from control to experimental conditions for each problem in each experiment in which it was used . . . . . 150

## LIST OF FIGURES

**Figure 1.1:** The nine-dot problem . . . . . 1

**Figure 1.2:** The solution to the nine-dot problem . . . . . 2

**Figure 4.1:** Plot of insight scores and MRT scores with the position of the two outlying cases indicated by their case numbers . . . . . 114

**Figure B.1:** The T-Puzzle segments and solution . . . . . 195

**Figure B.2:** Horse and Rider problem. (A) Panel Presentation and (B) Solution. Shown smaller than actual size . . . . . 196

**Figure B.3:** The necklace problem. (A) Starting array (B) Solution . 197

**Figure C.1:** Horse and Rider problem. (A) Panel Presentation and (B) Solution. Adapted from Weisberg and Alba (1981). Shown smaller than actual size . . . . . 204

**Figure C.2:** Illustration used for matchstick equation problem. Shown slightly smaller than actual size . . . . . 204

# CHAPTER 1

Have you ever worked at a problem that seemed to get the better of you for a while then suddenly the answer just seemed to occur to you, as if a light bulb lit up over your head? Perhaps the solution seemed so straightforward that you wondered why you had failed to see it for so long. This subjective experience of suddenly reaching a solution after a period of bafflement has been of interest to problem-solving researchers for several generations. It is interesting because even though individuals may possess the knowledge to solve a given problem, there seems to be some kind of obstacle that prevents them getting straight to the solution. Consider the problem shown in Figure 1.1. The task is to join all nine dots using four connected straight lines without lifting the pen from the page.

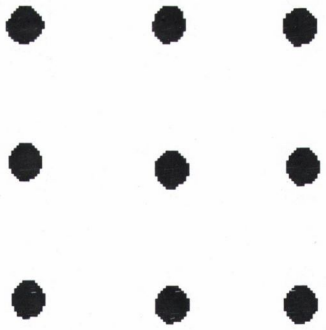


**Figure 1.1:** The nine-dot problem

It sounds straight-forward and everyone knows how to draw lines, but yet very few people solve the nine-dot problem without help. Trying to find out why this problem is so difficult has fed research since Maier's first published accounts (1930). The solution

is given in Figure 1.2. You may notice that the solution requires extending some of the lines past the square-formation. For many years, it was thought that this extension requirement was the main difficulty in the nine-dot problem because people simply did not consider options where lines went beyond the square (e.g. Scheerer, 1963).

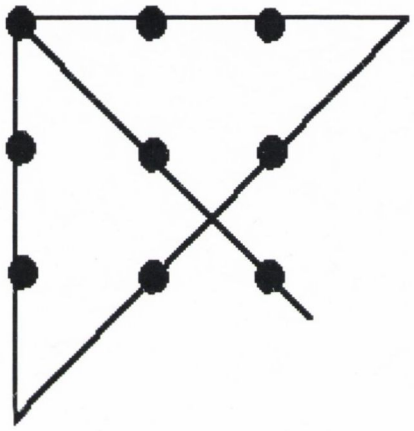


**Figure 1.2:** The solution to the nine-dot problem

## The Idea of Insight

### *What is an insight problem?*

There are a great variety of problems that have been labelled insight problems. Since the time of the Gestalt school (circa 1940) insight problems have typically been problems like the nine-dot problem that are very difficult and require several steps to go from problem to solution. The problems require an insight, such as going outside the square, to proceed but after this step there is some more work to be done. In the nine-dot problem, you need to figure out what direction and in what sequence to draw the

lines. More recently, however, problems such as the ping-pong ball have been used in the research literature (e.g. Ansburg and Dominowski, 2000). In this problem, you are asked to describe how to throw a ping-pong ball so that it will travel a short distance, come to a dead stop and then reverse itself. You are not allowed to bounce it off any surface or tie anything to it. Common suggestions include putting a back-spin on the ball, which will not actually work, or throwing it to another person, which violates the requirement for the ball to stop and return by itself. Although the problem initially seems impossible, there is a simple answer. Just throw the ball straight up in the air and gravity will cause it to stop for an instant before falling back to earth. For this problem, the difficulty seems to be that people assume the ball must travel forward and back but once they realise it can also go up and down, the solution follows readily. In other words, once the insight has been achieved there is little if any further work to be done to reach a solution. We term these problems 'single-step' problems, in contrast to problems such as the nine-dot, which we term 'multiple-step'.

One important stage in successfully solving a problem is to recognise a solution as such or to recognise when a move will lead to a solution. We consider that multiple-step problems may be more difficult because this recognition is not as straight-forward with these problems as it is with single-step problems. For example, extending lines outside the square may not be as obviously a promising move in the nine-dot problem as throwing a ball up is in the ping-pong ball problem. For this reason the distinction goes beyond simple problem difficulty.

But how can researchers more accurately define insight problems other than through examples? The truth is that a precise, accepted definition of insight problems is still lacking in the research literature. We can, however, summarise some generally

accepted characteristics. First, insight problems usually require some kind of fundamental change to the initial interpretation. Weisberg (1995) describes the existence of a discontinuity between the initial analysis of the problem and the analysis needed to actually solve the problem. For example, the ping-pong problem requires a change from a horizontal trajectory to a vertical trajectory, which opens up a whole new set of options to consider in the search for a solution. This change in analysis can be thought of as a "restructuring", a change in the way the problem is understood that has implications for how a person looks for a solution.

Second, for insight to occur, it is necessary for the person to be in possession of the knowledge to solve the problem. If, for example, someone had no information about the effect of gravity on objects then that person would not solve the ping-pong problem. Therefore, a person can only break out of an 'impasse', that is make new progress when problem-solving activity has stopped, if they already possess the requisite knowledge.

Third, another common feature of insight problems is that they tend to be 'ill-defined', that is there is some ambiguity about what the problem requires, or what form the solution will take. Take the nine-dot problem for example. Even though one knows what is required for a solution one does not know what exactly the solution will look like. Some people are surprised that the solution, as shown in Figure 1.2, has an arrow-shape. When problems are ill-defined they are more open to interpretation, interpretations that may be incorrect. Contrast this with a well-defined problem: $4x = 8$, what is x? Anyone who has encountered basic algebra problems before will know that the solution will be a number, a value for x, and will know what steps to take to get that value.

Other features that are associated with insight problems are that they are often more difficult to solve than they might at first appear, as already described in relation to the nine-dot problem, and that sometimes the solution or the key to a solution appears to occur quite suddenly. This last feature has been referred to as the 'aha' moment. It appears subjectively as a sudden transition from bafflement to enlightenment, however, this suddenness may not reflect the true nature of the processes underneath (Novick & Sherman, 2003). Progress towards solution may in fact be incremental rather than a leap from problem to solution.

In summary, insight problems often require a change in understanding or representation. The person can possess the knowledge needed to solve the problem but has trouble utilising it, possibly because the problem or its solution is ill-defined. When a solution does occur it often appears suddenly, leading to the 'aha' experience.

***Insight versus non-insight***

The idea of insight as something special has not been universally accepted (Weisberg & Alba, 1981; Mayer, 1995). There are, however, several studies that have noted distinctions between problems deemed to use insight and other problems.

Participants can more accurately predict their future success given a second chance at an unsolved non-insight problem compared to an unsolved insight problem (Metcalfe & Wiebe, 1987). Their ratings of how close they feel they are to a solution when they are working on a non-insight problem exhibit an incremental pattern as they near a correct solution, but for insight problems it remains uniformly low until a sudden increase just before the solution (Metcalfe & Wiebe, 1987; Davidson, 1995). Problem solving deteriorates as a result of verbalisation in think-aloud studies for insight

problems but not for non-insight ones (Schooler, Ohlsson & Brooks, 1993). Conversely, a concurrent working memory task (counting tones) interferes with performance on non-insight problems, but not with insight problems (Lavric, Forstmeier & Rippon, 2000). In an individual differences study, Gilhooly and Murphy (2004) were able to identify different clusters of insight and non-insight problems.

There is also evidence from neuroscience research that there are differences in brain activation for insight compared to non-insight problems (Lavric et al, 2000; Jung-Beeman, Bowden, Haberman, Frymiare, Arambel-Liu, Greenblatt, Reber & Kounios, 2004).

***Why is insight interesting?***

There are a number of reasons why investigating insight problem-solving is worthwhile, apart from pure interest. First, as we will see below the Gestaltists suggested that it allows us to observe how people reason with novel situations where there is no practised plan to fall back on. In everyday life, people often encounter problems that are new to them in some way or are ill-defined. As Wickelgren (1974, p.37) comments, "vagueness regarding goals is often a feature of real problem-solving". For example, you want to prepare a meal for a vegan couple who are coming to dinner. You may know that the finished meal should not contain any animal products but you still have to figure out what exactly you are going to cook. Or you may have a nobler cause, such as how to design a radio that will be useful in the remote areas of Africa where there is no electricity and nowhere to buy batteries. Radios are no good without a power-source. The inventor, Trevor Bayliss, came up with a novel

way to overcome this problem by using a clockwork wind-up mechanism (Richmond, 2003).

Insight is also of interest to those examining scientific or design breakthroughs. Historical accounts include numerous anecdotes about insights that have dramatically changed our knowledge about the world (e.g. Shrady, 1972; Perkins, 2000). In interviews with nine individuals from the arts and sciences who had made an important, creative contribution to their field, Csikszentmihalyi and Sawyer (1995) noticed that many great ideas seemed to come into consciousness suddenly. The respondents reported that often they had an 'aha' moment when they were resting after a period of intensive work. One such moment was described by a physicist who was returning from a vacation in California, "I got on the Greyhound bus to come back to Princeton and suddenly, in the middle of the night, when we were going through Kansas, the whole thing sort of became crystal clear, so that was sort of the big revelation for me, the eureka experience or whatever you like to call it" (p.350).

From Archimedes rushing naked from his bath to the apple that fell on Newton's head, understanding the processes that lead up to and facilitate important insights is of great interest. There is always the tempting possibility that it might be possible to speed up the process of scientific discovery. Scientific insights may not be separable from domain-specific knowledge, and may result from collaboration (Dunbar, 1995) but there may be a common experience of sudden illumination to which people can relate.

In order to examine how an insightful solution may come about, it is not always feasible to shadow experts for a protracted period observing the birth of an insightful idea. Moreover we are also interested in how everyday reasoning that does not rely on

expertise can result in insight. We turn instead to more concise and manageable problems like that of the nine-dot and ping-pong problems. These problems do not require extensive knowledge of a particular domain and it is possible for at least some people to solve them in a single testing session. Researchers have been carrying out experiments on insight using problems like the nine-dot since the early part of the last century (e.g. Maier, 1930). We turn now to a review of the theories that have been put forward in regard to insight from the early years up to an account of the current major theories.

## Theoretical Background to Insight Problem-Solving

### *The early years*

Many accounts of insight (e.g. Mayer, 1995) start with the observations by Kohler in 1925 of a kind of novel problem-solving in apes whereby a caged ape devised a means of reaching an item of interest outside of the cage by using a stick to bring the item within grasping distance. This observation suggested to some early experimental psychologists, particularly within the Gestalt school, that it was possible to generate a novel solution when it came to tackling a problem for which one had no previously learned method of reaching a solution. Wertheimer (1971) proposed a distinction between productive thinking where some new approach to a problem is devised, and reproductive thinking whereby a problem is solved using previously learned methods. He felt that the then (in 1945) existing approaches of logic and associationism did not adequately explain productive thought. Many of the terms, and indeed many of the problems, we use in relation to insight problem-solving today originate from this period. For example, Duncker (1945) described an obstacle to problem-solving called

'functional fixedness' whereby participants were unable to use an item in a new way when its typical function had recently been confirmed. In the candle-box problem, Duncker asked participants to find some way of attaching three candles to a door using the materials laid out before them. These items included candles, tacks, matches and three small boxes. In one condition, the boxes were used to contain the listed items but in another condition the boxes were empty. Participants who were shown the boxes as containers found it more difficult to arrive at the correct solution of tacking the boxes to the door and using them as candle-holders.

Duncker (1945, p.11) also described how participants could modify the original demands of a problem without even realising they had done so, and in this way made the problem more difficult for themselves, "by substituting unawares a much narrower problem for the original, he will therefore remain in the framework of this narrower problem, just because he confuses it with the original". This notion of an overly constrained version of the problem emerges again in more recent theories. In addition, Duncker describes the moment of sudden comprehension, the 'Aha experience', as the time when a "sudden restructuring of the thought material takes place" (p.29).

In Duncker's seminal work we can also find embryonic notions of the importance of planning ahead, "the man whose vision is not limited to the few feet just ahead, but who directly takes in the more distant possibilities as well, will most surely and quickly find a practicable path through difficult terrain" (p.39), as well as an early intuition of what we might now term 'spreading activation'. The term spreading activation describes a process where the activation of a memory spreads to other associated memories, making these memories more likely to be activated as well (Reber & Reber, 2001). Duncker describes "resonance" as a "process of recognition

that can be crudely described as follows: a peripheral stimulation sends into the nervous system a particular wave of excitation . . . and to this wave kindred traces respond as though by resonance" (p.77).

Although the Gestaltists made an important contribution to the study of insight, they have been criticised for the vagueness of some of the terms they used and the lack of detail on how mental processes actually occur (Eysenck & Keane, 2000; Robertson, 2001). With the increasing dominance of behaviourism and associationism in the post-war period, the study of insight stagnated somewhat.

***The revival of insight***

In the 1960's and 1970's, problem-solving research focused on processes in well-defined problems with the publication of Newell and Simon's information-processing theory of problem-solving (1972). Despite a drop-off of interest in insight problems, some notable papers were published. For example in 1963, Scheerer reviewed some of the problems used by the Gestaltists including the nine-dot problem.

Later in 1981, Weisberg and Alba sought to rebut the idea of insight as a special process in problems such as those reviewed by Scheerer. With problems such as the nine-dot, they argued that if extending lines outside the square was the key to a solution, then everyone who is informed of this fact should be able to solve the problem. They found that such comprehensive facilitation was not the case, and that this instruction only increased solution rates from near 0% (floor-level) to about 20%, while although a substantial increase, was not a reliable difference. They suggested that people had difficulty solving these problems because the domain of possible solutions was so extensive. Furthermore, participants might not become aware that the

domain they are exploring is incorrect or that another domain exists. Weisberg and Alba also attempted to dismiss two other problems with a similar argument but the six matches problem (see Appendix A) did reliably benefit from a key instruction and testing on the horse-and-rider problem (see Appendix B) was abandoned because it was easier than expected. Although they were arguing the case against insight as a special process, Weisberg and Alba's results did not convince everyone.

Lung and Dominowski (1985) countered the findings of Weisberg and Alba in relation to the nine-dot and concluded, "it is the act of extending a line, no matter right or wrong, that is predictive of success on the problem" (p.809), although they did propose that there were additional sources of difficulty including the need to intersect lines at a non-dot point. The controversy surrounding the nine-dot problem served to re-ignite interest in the area of insight problem-solving and by the 1990's insight problem-solving had once again established itself as a topic meriting rigorous investigation.

**Current Theories of Insight**

The dominant theory for problem-solving in general has been information-processing theory since Newell and Simon's (1972) problem space theory. This theory suggests that that there are two processes involved in solving a problem, *understanding* and *search*. When a person encounters a problem, reading and *understanding* it allows them to create a mental representation of that problem that includes a 'problem state' and a 'goal state'. This representation constitutes a 'problem space' within which the person can *search* for a way to get from the problem state to the goal state, that is, reach a solution. To help navigate the space between problem and solution, people often use

strategies or heuristics such as hill-climbing and means-end analysis. Hill-climbing involves selecting moves that seem promising and making progress towards the goal but without much planning. Means-ends analysis is a more structured approach where sub-goals are identified and progress is made by reaching a number of sub-goals, moving closer to the final solution.

Current theories of insight can be divided into those that emphasise the role of representation and those that emphasise the role of heuristics and search processes.

***Theories emphasising representation***

What has evolved into the representational change theory started with Ohlsson's (1992) proposal that an information-processing approach could also be applied to insight problem-solving. He suggested that ill-defined problems allow different interpretations of the problem and/or the goal, which can in turn lead to a poor representation of the problem.

Through a process of spreading activation, the problem representation activates a set of operators, or knowledge structures, that seem to be relevant to the problem. If the representation is of poor quality the actual operators needed to solve the problem remain inactive, even though the person may be competent to solve the problem.

From this explanation, Ohlsson (1992) defined insight problems as "problems which have a high probability of triggering an initial representation which has a low probability of activating the knowledge needed to solve the problem" (p.10). In the example of the ping-pong ball problem the solver knows that if an object is thrown up in the air then gravity will cause it to return to earth. However, if he or she is only thinking in terms of ping-pong balls travelling parallel to the ground, then the

knowledge relating to gravity will remain inactive.

An impasse occurs when problem-solving behaviour comes to a halt. Ohlsson suggested that a new pattern of activation is necessary to break the impasse, which in turn requires a new, or improved, representation. Initially three mechanisms for re-representation were proposed: elaboration, re-encoding and constraint-relaxation.

*Elaboration* requires adding new information to the existing representation such as when a person notices something new in the problem that they had not previously considered. For example, if you were trying to remove a ring that is stuck on your swollen finger, you may recall the knowledge that something cold reduces swelling. You can now use this information to form a solution to your problem whereby you apply ice to the finger until it shrinks sufficiently to remove the ring. *Re-encoding* involves correcting a mistake in the representation. For example, in the ping-pong ball problem you may realise that you were wrong to assume that the ball would travel back and forth. *Constraint-relaxation* occurs when a person realises that they have set unnecessary limits on the problem or its solution and subsequently relaxes them. For example, a person may constrain solution attempts on the nine-dot to within the square formed by the dot array. Once a correct representation is found, navigation of the problem-space continues as it might for a well-defined problem.

More recently, Ohlsson and his colleagues have focused on another mechanism for representation change, that of *chunk-decomposition* (e.g. Knoblich, Ohlsson, Haider & Rheinus, 1999), which requires breaking up the elements in a perceptual chunk. This has been applied in relation to matchstick equations constructed of roman numerals, for example in the form XI = III + III. When asked to move one match to balance the equation, participants are more reluctant to break up the 'chunk' that is the 'X' than

they are to move a match that is a chunk in itself like the 'I' (the answer is VI = III + III, moving the leftwards-slanting match of the 'X' further left to form a 'V' instead). The research on chunk-decomposition has been conducted primarily in relation to these matchstick equations.

Knoblich et al (1999) proposed that prior knowledge of equations as having variable values but constant arithmetic signs would constrain participants to changing value signs in a matchstick equation of the form III = II – I, and that only when these constraints were relaxed would the participant reach a solution (III – II = I). In addition, participants would be more reluctant to break up 'tight chunks' such as the X in the earlier equation described above. They were able to verify their predictions that matchstick equation problems of these types are more difficult to solve than equations where a 'loosely' chunked value was the crucial match, for example, IV = III + III to VI = III + III (moving the first 'I' from before the 'V' to after). Not only is this key match a loose chunk but it is also in keeping with the bias of changing values in an equation rather than signs. In other words, "a solution that violates few constraints [on breaking up chunks] should be easier to think of than one that violates several constraints" (p.1535).

Knoblich, Ohlsson and Raney (2001) confirmed the relative difficulty of these categories of equations. In an eye-tracking study, they found that for the more constrained categories of equations, successful problem-solvers spent longer fixating on the crucial elements as they approached solution than did unsuccessful problem-solvers. This result was viewed as supporting the idea that having reached an impasse in relation to trying to change the values, successful problem-solvers then relaxed their constraints and turned their attention to the operators and/or tight chunks. Unsuccessful

problem-solvers, on the other hand, did not spend any longer on the previously-constrained, crucial elements than would have been expected by chance. Knoblich and Wartenburg (1998) reported that performance on equations with constraints increased with exposure to pre-conscious primes (a clue presented too quickly to be processed consciously). Performance on the easier, value-based equations did not improve as a result of primes.

The preceding studies have been interpreted in favour of the representation theory. However, it may also be the case that people embark on a search of all the possible ways of recombining the matches and use some sort of heuristic to refine their search. If they start with a promising match, such as an easily transferable 'I' match, and this one turns out to be the correct match then they will solve the problem relatively quickly. If, however, the correct match is further down in the selection criteria, it will take longer for the participant to get as far as testing possible combinations using this match and they may not reach a solution in the time allowed, or give up.

Ohlsson's (1992) information-processing approach heavily influenced other theories of insight. Suzuki, Abe, Hiraki and Miyazaki (2001) propose that there can be multiple constraints in one problem, those that apply to the objects in the problem, the relations between the objects and on the ultimate goal. The representation of the problem and of the solution is affected by these constraints and the relaxation of constraints is the principal means of improving the representation. They also add that constraint relaxation may only occur after initial attempts have failed to reach a solution.

Gick and Lockhart (1995) also subscribe closely to the information-processing account with the most important processes being the generation of a representation of a

problem and the search for a solution within the constraints of that representation. There are three potential areas of difficulty: accessing an appropriate representation, constructing an appropriate novel representation and the application of this representation to reach a solution. Identifying the common elements in past failures, 'noticing invariants' as proposed by Kaplan and Simon (1990), is a useful heuristic for changing poor representations. In contrast to the other representation-based theories, Gick and Lockhart put a greater emphasis on the affective aspects of insight problem-solving, especially the 'aha' experience. They conceptualise three component properties of insight: a failure to solve or a lack of understanding, then transition to a goal-state, at least some part of which does not take place consciously, and lastly the affective response characterised mainly by suddenness and surprise. The surprise aspect varies as a function of the incongruity between the original and new representation such as is seen with jokes, where an incongruity arises between the representations presented in the lead-up and the punch-line. The ease with which the new representation leads to a solution gives the subjective feeling that the solution comes suddenly.

The Three-Process Theory (Sternberg & Davidson, 1989; Davidson, 1995) proposes processes similar to those described by Ohlsson (1992) to facilitate overcoming impasses in insight problem-solving. *Selective encoding* occurs when a person notices elements of a problem that were not obvious before or were considered irrelevant. This kind of insight leads to a restructuring of the problem representation using the now-relevant information. *Selective combination* occurs when the person recombines elements of a problem in a way that was not previously obvious to them. *Selective comparison* occurs when the person sees a relationship between the problem and previously acquired information that they had not noticed before, perhaps allowing

them to use an analogous solution. Insightful thinking is said to have occurred when these processes are successfully applied in problems where the person does not have a standard set of procedures.

The theories in this section emphasise representation as the primary influence on the way people approach problems and the information they draw on to help them solve problems. It is only when the representation is changed to a more accurate one, frequently through the relaxation of inappropriate constraints, that a problem can be solved correctly.

***Theories emphasising heuristics and search processes***

Chronicle, Ormerod and MacGregor (2001) proposed a computational model of insight comprising of three components: (a) scope of search; (b) move selection and (c) move evaluation. Participants select a move based on their evaluation of the current position compared to the goal that maximises the progress they can make towards the goal. This strategy is termed 'locally-rational' and is similar to the hill-climbing heuristic mentioned earlier. In the example of the nine-dot problem, people draw lines that will join the maximum number of dots (maximisation). This model predicts that the nine-dot problem is difficult because a person will have to plan ahead for at least four moves in order to produce a correct solution otherwise they will not realise that the maximisation strategy will not lead to a solution. Extending lines beyond the square would not be identified as necessary to achieve maximum dot cancellation except when the participant had a look-ahead of four moves. They predicted accurately that hints aimed at encouraging people to extend lines beyond the square of the nine-dot problem would fail to significantly improve solution rates because people would rarely plan four

moves in advance, although solution rates in one condition did reach 33%. MacGregor, Ormerod and Chronicle (2001) were also successful in using the model to predict moves on the nine-dot problem and its variants using different numbers of dots but with the same solution pattern.

Ormerod, MacGregor and Chronicle (2002) extended testing of their computational model in relation to a new problem, the eight-coin problem. Participants have to move two coins in an array so that each coin touches exactly three others. The solution requires stacking some of the coins, such that participants must work in three dimensions instead of two. It also requires dividing coins into two groups. The model suggests that participants will initially try to make three coins touch in two-dimensions if such moves are available (the locally-rational strategy). When criterion-failure occurs, that is when the participant realises that this strategy will not work, it motivates him or her to relax the constraint that a coin must touch two others in the first move. They found that when the coins were arranged so that the locally-rational, but erroneous, strategy of moving one coin to touch others two-dimensionally was not available, more participants solved the problem correctly than when such a move was available. When this one-touching-two move was available, more participants chose this move as their first move than would have been expected by chance alone. Few participants in either group, however, solved the problem without a verbal hint to use three dimensions that was introduced some minutes into the attempt. The authors suggest that the no-move-available participants who experienced the criterion-failure early on were better prepared to use the verbal three-dimensional hint when they received it.

Overall, the eight-coin problem provides further evidence for the theory that

participants will use a locally-rational strategy to select moves that appear to bring them closer to the solution. Only when participants realise that such moves do not work, that is they experience criterion-failure, will they consider alternative moves. In addition, they will be unlikely to make good use of constraint-relaxing hints until this point has been reached. Ohlsson (1992) also proposed constraint-relaxation as a mechanism for making progress but he referred to constraints on the problem representation whereas the current theory focuses on relaxing the constraints on available moves.

Chronicle, MacGregor and Ormerod (2004) showed evidence for the use of a hill-climbing heuristic in relation to a six-coin problem where the solution could either be a ring arrangement or a grouping arrangement depending on how many moves were required. When working towards a ring arrangement, participants were inclined to select moves that maximised the number of coins in that solution pattern. The authors also indicated, however, a role for the solution representation in this problem because whether participants were working towards a ring or grouping solution significantly influenced the solution rates (the former being more difficult) and also the likelihood that participants would employ the hill-climbing heuristic. Participants were not making an error in their representations as such, but were working towards a particular solution representation that was more difficult to achieve than other possible solutions. The authors also observed a hill-climbing heuristic in participants attempting another coin problem with ten coins.

Heuristics have also featured in other hypotheses of insight problem-solving. Kaplan and Simon (1990) described the heuristic of 'noticing invariants' in relation to the mutilated chequerboard problem. This strategy requires identifying what element is

common to all earlier failures. Kaplan and Simon observed participants attempting to solve the problem of the mutilated chequerboard. In this problem, you are told that 32 dominoes will cover all the squares on a normal chequerboard with 64 squares. If, however, two diagonally opposite squares on the board are removed, will 31 dominoes cover the remaining 62 squares? Not only do you have say 'yes' or 'no', the answer must be proved. The answer is 'no' because every domino covers one black and one white square. The two diagonally opposite squares are of the same colour, so if two white squares are removed there will be two black squares leftover and one domino cannot cover two black squares. Participants typically find this problem very difficult but it can be made easier if the idea of parity is introduced, for example if the alternate squares are labelled bread-and-butter instead of black and white. When two 'butters' are removed there will be two 'breads' leftover.

The usual initial approach to the problem is to try and cover 62 squares with the 31 dominoes, however, there are so many possible layouts that it is virtually impossible to sample and eliminate them all. The authors argue that the size of the search space is the key difficulty and instead of too many constraints, there are too few. Chronicle et al (e.g. 2004) have proposed that a similar difficulty exists in the nine-dot and various coin problems. Heuristics are employed to make the search more manageable, but these heuristics may not always be the best strategy. However Kaplan and Simon believe that the noticing invariants heuristic can be a useful strategy for insight problems. Like Chronicle, Ormerod and MacGregor, Kaplan and Simon also view insight problem-solving as similar to other forms of problem-solving except that with insight problems, one has to search *for* the problem space or representation. The heuristic and search accounts have been tested for problems where one has to choose

from a multiplicity of possibilities. However some insight problems initially appear to have no possibilities (Dominowski & Dallob, 1995), for example how can one man legally marry twenty women? With these problems it can be difficult to think of any possibilities rather than have to limit the choice between many.

Gilhooly and Murphy (2004) found that a scaled-down version of the mutilated chequerboard problem with 36 squares instead of 64 still resulted in almost no correct solutions. This finding could suggest that reducing the search space for this problem may help participants to realise that one less domino will not cover two fewer squares, but it may not help them devise the proof based on parity that is needed for a full solution.

Other authors have also considered the idea of a role for search in insight problem-solving. Smith (1995) suggested that people search memory for relevant knowledge to help with the solution of a problem. However, other more recently learned, more dominant knowledge can interfere with the search for the more obscure knowledge. Keane (1989) suggested that people search long-term memory for a potential solution plan that can be evaluated and then either used, rejected or adapted. Dealing specifically with the Gestaltist construction-type problems, such as Duncker's candle-box problem (see Appendix A), Keane suggested that the objects featured in these problems do not act as good cues for retrieving appropriate solution plans. He also suggested that when people access information about an object, they do not retrieve everything they know about that object every time they encounter it. Instead, they access the salient properties as suggested by the context within which it is presented. To take Keane's rather gruesome example, if a person sees a pair of pliers in the context of an electrician's toolbox, the pliers' property of being a gripping tool

becomes salient. If, however, that person sees the pliers by the head of a bleeding corpse, then its 'heaviness' property becomes salient. Insight problems sometimes require that objects be used in a way that is not suggested by the context in which they are presented, as in the candle-box problem for example.

Accounts emphasising heuristics and search processes see the problem space as highly influential in insight problem-solving. To make the search of the problem space more manageable, people employ various strategies or heuristics that may be appropriate with other problems and are quite sensible. However, insight problems are such that these strategies actually hinder the problem-solving efforts rather than helping them. To reach a solution, a person must recognise that his or her current efforts are not working and be prepared to try another strategy.

***Common themes in contemporary insight theories***

A common theme that we identified in the previous section is the role of mental representation. How a problem is understood or represented is central to the subsequent solution process. The problem representation is the tool that guides the search of memory, consciously or unconsciously, for relevant information or potential solution plans (Ohlsson, 1992; Keane, 1989). If the representation is flawed in some way, that is incorrect, incomplete or overly constrained, then the solution attempts will be similarly flawed. Processes that improve the representation improve a person's capacity to utilise the information they already possess (Davidson, 1995). The lack of definition distinguishes insight problems from more well-defined problems such as algebraic equations. The representation of these problems is usually concise and not open to misinterpretation and so it is more likely to activate the information needed for

its solution, if the person possesses that information.

Another frequently occurring theme is the importance of experiencing, and recognising, failure. Failure may be the catalyst that prompts us to change something in our approach. Experiencing 'criterion-failure', that is, being unable to find an acceptable move in line with one's current strategy, may be sufficient to cause doubt in the strategy and to consider alternatives (Chronicle et al, 2001). Problem solvers may be more receptive to clues after failure (Suzuki & Hiraki, 1997; Ormerod et al, 2002). They may also be able use information from failed attempts in new attempts: actual failures may be 'flagged' so that when helpful information is encountered later, there is a short-cut back to earlier work (Seifert, Meyer, Davidson, Patalano & Yaniv, 1995), and analysis of failed attempts may highlight a common error (Kaplan & Simon, 1990). Therefore, meta-cognition may play an important role in insight problem-solving by helping people to monitor what they have tried and failed and to switch strategies accordingly.

A third theme that runs through the literature is that of constraints. One type of constraint operates on the representation (Ohlsson, 1992; Suzuki et al, 2001). Another type of constraint operates on the kinds of moves that are considered (Chronicle et al, 2004). In both of these situations, relaxing the constraints assists in progress towards a solution by allowing more possibilities to be considered, possibilities that include the correct solution or the next step towards solution. Although for the mutilated chequerboard problem, Kaplan and Simon (1990) assert that insight problems are not sufficiently constrained.

Although there may be some consistency in the themes that emerge in the theoretical analysis of insight, there has been diversity in the problems used to

investigate insight. The nine-dot problem has regularly featured in studies of insight but other problems, like the matchstick equations, feature in relation to just one account. This diversity is a complicating factor when trying to compare theories, and when evaluating experimental results that suggest particular properties of insight problems, and problem-solvers, as outlined in the next section.

**Properties of Insight Problems and Problem-Solvers**

Other investigations have identified some properties of problems and problem-solvers. For example, people can benefit from training in insight problem-solving. Ansburg and Dominowski (2000) found that training participants on insight problems by getting them to explore a set of problems ('verbal' problems, usually with a single-step to solution) and their solutions improved performance on a subsequent set of insight problems. The authors also found that participants benefited from strategic instructions, such as that the right answer would seem to 'click', to be careful in interpreting the problem and that the first answer that occurred to them could be incorrect. The best solution rates were found when strategic instructions and training were combined. The authors suggest that training and instructions may help participants to achieve an understanding of the deeper structure of the problems and also encourage them to monitor their progress.

Kershaw and Ohlsson (2001) found that training could also improve solution rates on the nine-dot problem. Participants who trained by working on problems that involved procedures similar to those required in the nine-dot such as turning on a non-dot point, solved a problem based on the nine-dot problem faster and more often than participants who received hindering training problems or no training at all. The

hindering problems were dot-linking problems whose solution did not require moves such as turning on a non-dot point that would be needed to solve the nine-dot-like problems. However, the positive training effects depended on how successful participants had been on the training problems.

In contrast to the apparent benefits of training, researchers have found that people do not seem to be able to utilise relevant information, for example in the form of hints very effectively. Perfetto, Bransford and Franks (1983) found that presenting relevant information in a sentence-rating task did not help a subsequent insight problem-solving task unless participants were informed of the connection prior to attempting the problems. They also found that once uninformed participants had attempted the insight problems, they could no longer benefit from the earlier information even when belatedly informed and given another go at the insight problems. It seemed as if these participants could not suppress their earlier, incorrect attempts. In a subsequent, similar experiment Lockhart, Lamon and Gick (1988) found that participants were more likely to recall relevant information from an earlier task if it was presented in the form of a puzzle rather than in a simple statement, as it had been in the previous study. The same puzzling information did not benefit participants when it was presented in simple, declarative statements, even when those statements were repeated.

Both of these studies used insight problems of the kind we earlier described as single-step and that are also referred to elsewhere in the literature as 'verbal' problems (Dominowski & Dallob, 1995). The nine-dot problem has also proved resistant to hints. Weisberg and Alba (1981) found that direct hints to extend lines outside the square did not reliably improve solution rates. Similarly, a hint to the shape of the

solution of the nine-dot problem using a shaded area did not reliably benefit participants (Chronicle et al, 2001).

One study that did, however, find some positive effects of hints is reported by Davidson (1995) who compared the performance of participants divided into lower and higher IQ categories. Only the lower IQ participants (although these participants were still of average intelligence) benefited from the hints given to the mainly single-step, 'verbal' problems while the higher IQ group performed better than the lower IQ group regardless of whether or not they got a hint. Other studies on hints have mainly used student populations, which may have a higher mean IQ, whereas this study used members of the general public. This difference in the effect of the hints may suggest that participants at different levels of intelligence may approach problems differently. From examination of individual protocols, the author suggested that the higher IQ group were more inclined to try various ways of restructuring the representation of the problem, as seen in the number of different strategies employed. This group may also have better self-monitoring skills.

Individual differences studies have looked at what skills are associated with being able to solve insight problems in the hope that they will illuminate how people reason with insight problems. As Davidson (1995) proposed, higher IQ individuals may be better at monitoring their own progress during the problem-solving process, a proposal echoed by Ansburg and Dominowski (2000).

Another skill that has been associated with success on insight problem-solving is the ability to generate a larger set of associations to a target. For example, in the Remote Associates Test (originally developed by Mednick, 1962) a participant has to find the common associate of three concepts, such as 'common', 'work' and 'play'. In

this example, the solution is 'ground': 'common ground', 'ground work' and 'play ground'. For each of these concepts, however, the word 'ground' has different meanings. Better performance on these types of problems has been found to be associated with better performance on insight problems by both Schooler and Melcher (1995) and Ansburg (2000). Similarly, Gilhooly and Murphy (2004) found that participants who were able to generate more alternative uses for objects, for example a paperclip, solved more insight problems than participants who generated smaller sets of alternative uses.

So far the available evidence suggests that people who are successful at solving insight problems might be better able to manage their problem-solving attempts and to access a wider range of knowledge about the elements of a problem. Efforts to help people towards a correct solution through the use of hints have, however, had less than comprehensive success.

**Aims of the Thesis**

Our overall aim in this work was to investigate how people reason with insight problems, with an emphasis on how people succeed in solving insight problems rather than why they fail to solve them. We report the results of seven experiments that aimed to improve performance on insight problems through simple manipulations, and to identify what skills are associated with success.

Our aim in the first experimental chapter, Chapter 2, was to explore the role of mental representation in insight problem-solving. As was evident from our review, many accounts include a prominent role for representation (e.g. Ohlsson, 1992; Knoblich et al, 1999) and so we examined whether manipulations to improve the

quality of participants' representations would boost the solution rates on a range of insight problems. We report three experiments in this chapter. The first experiment sought to verify that participants actually make the kind of errors that they are assumed to make, and to see how their error is related to performance. The second experiment investigated how the representation of a problem is influenced by how the problem is presented. We compared the effects on performance of altering the wording of a range of problems to encourage a more accurate representation. The third experiment in the chapter looked at transforming the problem representation through the use of physical props.

Many accounts also include a prominent role for the size of the potential problem-space. When participants have a large number of possibilities to search through they may employ heuristics to narrow the search but in a way that makes the problem more difficult (e.g. Chronicle et al, 2004). Therefore in Chapter 3, our aim was to examine the role that search space plays in insight problem-solving.

We report two experiments in this chapter. Experiment 4 examined the effect of limiting the search space in problems with physical moves. We hypothesised that having a wide choice of possible moves could contribute to the difficulty of a problem and we predicted that limiting the number of possible moves would make the problems easier. Experiment 5 investigated whether limiting the number of possibilities would have a similar effect on problems that did not involve physical moves.

In both series of experiments we used a range of insight problems to check whether all problems react the same way to similar manipulations. There are many examples of insight problems in the literature and any future theory of insight will have to adequately explain all of them or else redefine the concept of an insight problem to

include a more select group.

In these two experimental series, we considered various ways to improve people's ability to reason with insight problems, and the effect of these manipulations on different kinds of problem. Our third series examined the skills associated with being able to solve insight problems. We reasoned that if a particular skill or set of skills seemed to play a key role in determining how successful a person is at solving insight problems, it might provide some clues as to what people need to do to reach a correct solution.

Chapter 4 reports two studies of individual differences. We used a selection of insight problems, including problems that seem to have a visual or spatial element and those that are primarily verbal. Experiment 6 investigated three key skills. The first was whether being able to use mental imagery was useful in reasoning about the situations described in insight problems. We contrasted this with performance on a more verbal test, a version of the Remote Associates Test described earlier. The third element in this experiment was a measure of absent-mindedness (the Cognitive Failures Questionnaire).

Experiment 7 was influenced by the suggestion that being able to monitor one's progress and change strategies accordingly was important for success in insight problems (e.g. Davidson, 1995). We hypothesised that being able to consider a number of possible moves and switch between them would require both working memory and attention resources. We included several measures of attention and working memory in this experiment.

All the experiments focused on how people can successfully solve insight problems. The three experiments reported in Chapter 2 examined how problems can be

manipulated so that participants' mental representations of them facilitate better problem-solving efforts. The two experiments reported in Chapter 3 examined how participants' performance can be improved by narrowing the size of the problem-space that they need to explore when searching for a solution. The two experiments in Chapter 4 attempted to identify what skills contribute to individuals' ability to reason with insight problems. In the final chapter, Chapter 5, we discuss the implications of our findings for research on insight problem-solving.

The overall aim was to explore ways in which people can be helped to successfully reason with insight problem. We sought to do this by identifying what kinds of simple manipulations would improve performance and what particular skills were associated with individuals who are good at solving insight problems.

# CHAPTER 2

## Representation in Insight Problem-Solving

The principal aim of the series of experiments described in this chapter is to examine the role of mental representation in insight problems and the ways in which representations can be manipulated to facilitate performance on such problems. The representation of a given insight problem has a key role in the solution process according to several theories of insight (see Chapter 1). According to Ohlsson's (1992) theory, the representation of a problem activates knowledge in long-term memory. If the representation is flawed, for example, a horizontal trajectory is assumed for the ping-pong problem, it does not activate the relevant knowledge, for example about gravity. Activating the wrong knowledge, and leaving the relevant knowledge inactive, may explain why people find insight problems difficult even though they possess the knowledge to solve them. Most insight problems do not require special knowledge; for example most people know that when an object is thrown up, gravity will cause it to come back to earth. This characteristic is one that has been especially associated with insight problems (e.g. Dominowski, 1995).

When progress towards solving a problem comes to a halt and the individual can see no way forward, an impasse has been reached. Ohlsson asserts that an impasse can be broken by a change in representation. An improved representation can bring into consciousness the previously neglected, but necessary knowledge that will bring the individual to a solution or at least set him or her on the path to a solution. We can illustrate this idea in more detail with the example of the ping-pong problem. If the representation includes the assumption that the ball is travelling in a horizontal direction, then only knowledge concerning this direction will be activated. Knowledge

relating to how an object moves up and down, and the effect of gravity remains inactive and the problem is unsolved. If the constraint on ball trajectory is relaxed in some way, for example through the provision of a hint, then the activation may extend to include vertical travel and a solution may be reached.

Weisberg (1995) also advances the notion of a need to change or restructure a flawed representation of a problem to progress towards the solution in his taxonomy of insight problems. He suggests that if a comparison of the initial solution attempts and the correct solution requires a change in the way a problem is analysed, for example a change from ways a ball can travel back and forth to up and down, this change is evidence of restructuring.

***Representation and hints***

Our aim in this first series of experiments was to examine ways in which people may be facilitated in constructing better representations of insight problems in order to reduce the hindrance posed by flawed representations. We wished to target the potential for misrepresentation from the very start of the problem-solving process as earlier work suggests that manipulations to change the representation once it has been formed have limited success. For example, Perfetto et al. (1983) compared participants' ability to use relevant information that they had acquired during a rating task in a subsequent insight problem-solving task. Participants who were not told of the connection before they made their first attempt at the problems were still not able to utilise the information even after they were given a second attempt at the problems. In contrast, participants who were informed of the connection before their first attempt were able to use the earlier information. Perfetto et al. concluded that the previously uninformed

participants' attempts to access the earlier, helpful information were blocked by the incorrect solutions they had made on their first attempt at the insight problems. Chronicle et al (2001) found that encouraging participants on the nine-dot problem to extend lines by shading in the area occupied by the solution failed to reliably improve solution rates. Also on the nine-dot problem, Weisberg and Alba (1981) found that an explicit hint to go outside the square did not reliably improve solution rates (although Ohlsson, 1992, argues that as solution rates did increase by about 20%, this result suggests that the within-line assumption is a factor of difficulty). Therefore, our emphasis in the following experiments was on starting participants off with a good representation of the problem rather than trying to get participants to reshape a bad representation with hints.

Over the years a rather large battery of problems assumed to involve insight has accumulated in the literature. Some problems seem to require only a single insight and once the representation has been adjusted accordingly, the solution is readily apparent. In other problems, however, there is more work to be done even after insight has been achieved, and the initial impasse broken. We included both multiple-step and single-step problems in our problems set and because of the extra work needed in the former, we expected them to be more difficult.

**Overview of Experiments**

The aim of this series of experiments was to examine the role of representation in insight problem-solving. Our first aim was to identify the most common erroneous assumption made for a range of insight problems. Previously, deciding on what errors to tackle with hints has been largely based on conjecture as to what errors participants

make when they are representing a given insight problem. Our primary aim in the first experiment, therefore, was to establish the common errors made on different insight problems. We designed the first experiment so participants would externalise their representation of each problem by either drawing a diagram or explaining it to a friend. This explicit representation allowed us to check the nature of errors. In addition, we manipulated whether participants externalised their representations either before or after attempting a solution. We considered that externalising a representation first might help participants to recognise their erroneous representation and adjust it accordingly.

In the second experiment, we hypothesised that the phrasing of insight problems may serve to make some elements of the problem more salient than others, or fail to highlight the salience of other elements. Hence, the wording of a problem may have a strong effect on the way the problem is mentally represented. For a set of multiple-step and single-step problems, we changed the wording in ways we expected would help participants to construct a better representation.

In the third experiment, we provided participants with physical props to improve the quality of their representation. We anticipated that having an external representation of the problem via the physical props and being able to interact with the external representation would help participants to avoid or overcome errors in their mental representations.

Our strategy in this series of experiments was to identify initially the specific erroneous representations people tend to construct in relation to a variety of insight problems. We then used this information to design manipulations to improve participants' representations as measured by an observable increase in correct solution

rates or a reduction in the time taken to solution.

## Experiment 1

The aim of the experiment was to examine the nature of the mental representations that people construct for a variety of problems, in two ways. First, we established the errors people make on a variety of single-step and multiple-step insight problems. We used eight problems in the experiment, including four single-step problems: the ping-pong ball, window-cleaner, camping trip and woods walk problems, and four multiple-step problems: the nine-dot, six-matches, candle-box and radiation problems. (The eight problems are described in full in Appendix A). For each problem there is an obvious error that could lead participants astray (see the common error listed for each problem in Appendix A). The errors depend on participants assuming an additional constraint in their mental representation of the problem (Ohlsson, 1992). For example in the ping-pong problem, many participants add the constraint that the ball must travel horizontally. The constraint must be relaxed if the correct solution is to be reached. Some of the errors have been conjectured to underlie the difficulty of the problems in the previous literature (e.g., Weisberg, 1995).

We attempted to confirm that participants made the expected errors by requiring them to construct external representations of the problems. They constructed two sorts of external representations. One sort was based on communicating their understanding of the problem by drawing a diagram of the situation. The second sort was to explain the problem 'as if to a friend'. We selected problems that could be represented readily in either diagrams or verbal explanations. The external representation should help establish what mistaken assumptions people make when they attempt to solve these

problems.

In addition, we varied whether people generated an external representation prior to solving the problem or after solving the problem, to assess whether generating external representations helped people to solve the problems. The experiment was thus also designed to test whether external representations, diagrammatic or verbal, improve correct solution frequencies for single-step and multiple-step problems.

## Method

### *Participants*

The participants were 27 psychology undergraduates from Dublin University, Trinity College, 7 men and 20 women, aged between 18 and 45 with a mean age of 22, who volunteered to participate for course credits. One participant was eliminated because he indicated he had seen all 8 problems before.

### *Materials and design*

We constructed a set of eight insight problems, four multiple-step problems (based on the nine-dot, six-matches, candle and radiation problems) and four single-step problems (based on the ping-pong ball, window-cleaner, camping trip and woods walk problems). Most of these problems have been examined previously in the insight literature. Appendix A provides the description of each problem presented to participants, its correct solution, and the conjectured error. Each participant carried out two tasks: (1) they attempted to solve the problem and (2) they generated an external representation of it, either in a drawn diagram or in a verbal explanation to a friend. Half the participants were instructed to attempt to solve the problem first, and then to

generate an external representation of it; the other half were instructed to generate an external representation first, and then to attempt to solve it. For half of the problems the participants were asked to draw a diagram of the problem and for the other half they were asked to explain it to a friend. To control for order effects, half of the participants drew diagrams for the first four problems and explained the last four, and the other half did the tasks in the opposite order. Participants recorded their solutions, diagrams and explanations on separate sheets in an answer booklet.

***Procedure***

Participants were tested in small groups of one to three individuals. The participants first read the instructions (see Appendix A). Each problem was then displayed on an overhead projector at the front of the room. Each problem was presented for a set period of time: the length of time is indicated in the Appendix for each problem, and it was determined in a pilot study where 8 volunteers read each problem aloud slowly and carefully and their average reading time was doubled.

Participants were instructed by way of an example (not later used in the experiment) and they were told that they would be given different sorts of problems, each presented on a screen for a set period of time, and they were asked not to take notes. Participants who generated an external representation of the problem first and then solved it, were told their first task was to 'communicate your understanding of [the problem]...For some of the problems you will be asked to draw a simple diagram of the problem. For others you will be asked to explain the problem in writing as if you were giving it to a friend, giving as much detail as possible but without revealing the answer directly'. They were told they would have one minute to finish this task, and if the

solution occurred to them during this task, they were to record it on the sheet along with their diagram/explanation. Their second task was to try to solve the problem. They were asked to write their solution attempts and all their rough work. They were told they would have one minute, thirty seconds to work on a solution for each problem. The length of time was chosen on the basis of previous research which showed that 97% of all correct solutions to insight problems occur within the first two minutes of problem-solving time allowed, including time to read the problem (Lockhart et al, 1988). The shortest reading time allowed in the experiment was thirty seconds, and combined with a solution time of one minute, thirty seconds, provides at least two minutes in total. Participants who solved the problem first and then represented it were given these task instructions in the opposite order. The order of tasks, of attempting a solution first or representing the problems first, was counterbalanced.

The problems were presented in a different random order on the overhead projector for each small group. The participants were asked to complete the 8 problems in the order they were given, not to skip any, and not to go back to previous problems to change an answer once they had completed it. At the end of the experiment participants indicated if they had seen any of the problems before. They were then given the solutions to the problems.

**Results and Discussion**

Participants had seen on average 1.8 of the problems before, with a range of 0-3 (21 of the 26 participants had seen the nine-dot problem before), and we analysed only the 160 unseen problems (of a potential set of 208 problems, that is, 27 participants given 8 problems). Scores for each participant were calculated based on the proportion of

unseen problems that were solved correctly.

**Table 2.1:** Percentages of correct solutions for single-step and multiple-step problems in Experiment 1

| | | **Represent then solve** *Total* | | | | **Solve then represent** *Total* | | |
|---|---|---|---|---|---|---|---|---|
| | N | ***Diagram*** | ***Explain*** | | N | ***Diagram*** | ***Explain*** | |
| **Single step** | | | | | | | | |
| Ping pong | 13 | 50 | 29 | *39* | 12 | 0 | 40 | *17* |
| Window cleaner | 14 | 50 | 50 | *50* | 12 | 57 | 100 | *75* |
| Camping trip | 14 | 50 | 88 | *71* | 12 | 43 | 80 | *58* |
| Woods walk | 14 | 50 | 50 | *50* | 12 | 43 | 40 | *42* |
| *Total* | | *50* | *55* | *53* | | *36* | *65* | *48* |
| **Multiple step** | | | | | | | | |
| Nine-dot | 4 | 0 | 50 | *25* | 1 | 0 | 0 | *0* |
| Six matches | 14 | 0 | 0 | *0* | 12 | 0 | 20 | *0* |
| Candle | 7 | 100 | 67 | *86* | 5 | 67 | 50 | *60* |
| Radiation | 9 | 25 | 0 | *11* | 5 | 0 | 0 | *0* |
| *Total* | | *31* | *17* | *24* | | *14* | *22* | *17* |

* Key: N reflects total number of participants unfamiliar with this problem

***Single-step versus multiple-step problems***

Participants solved 40% of the problems overall. As can be seen from Table 2.1, they solved more single-step than multiple-step problems overall (50% versus 21%, Wilcoxon's $z = 3.93$, $N = 23$, $p < 0.0005$, and all tests are one-tailed unless otherwise stated). The result confirms that multiple-step problems are more difficult than single-step ones and it is consistent with our suggestion that multiple step problems require more processing than single step problems.

***Conjectured errors***

We classified participants' external representations as indicating the conjectured error or not. For example, for the ping-pong ball problem, a diagram of a ball travelling in a

horizontal direction was classed as the conjectured error; likewise an explanation that mentioned the horizontal direction was classed as the conjectured error (Appendix A describes the conjectured errors for each problem). An independent rater scored over one third of the booklets, and inter-rater reliability approached 100%. Participants made the conjectured errors more often for multiple-step problems than single-step problems (72% versus 53%, Wilcoxon's $z = 3.75$, $N = 22$, $p < .001$, two-tailed). For the 96 incorrect solutions produced overall, participants' external representations contained the conjectured error for 87 cases (and we were unable to classify 1 case). For the 64 correct solutions produced overall, their external representations contained the conjectured error for only 8 cases (and we were unable to classify 3).

We examined the tendency for expected errors to be associated with failure in a 2x2 table (expected error present or not and solution correct or not) and found that this tendency was reliable for each of the single-step problems: people who failed to solve the ping-pong problem tended to represent the ball travelling in a horizontal direction (Fisher's exact test, $p < 0.001$, and this test was used in all cases where there were cells with an expected frequency of less than five), people who failed to solve the window cleaner problem tended to represent the man at the top of the ladder (Fisher's exact test, $p < 0.0005$), people who failed to solve the camping trip problem tended to represent the 'thing' as animate (Fisher's exact test, $p < 0.01$), and people who failed to solve the woods walk problem tended to represent the two men as setting out together (Fisher's exact test, $p < 0.0005$). The results confirm the conjectured errors for single-step problems.

The difference was also reliable for the multiple step problems: people who failed to solve the candle problem tended to represent the box as a container ($\chi^2 = 6.52$,

df =1, p = 0.055), people who failed to solve the radiation problem tended to represent the rays as high intensity (although the result was marginal in this case, $\chi^2$ = 13.00, df =1, p = 0.077). For the six-matches problem every participant represented the triangles in two dimensions only, including the one person who got the correct solution, and for the nine-dot problem there were too few data points to test, because most of the participants had seen the problem before. The results confirm that participants make the conjectured errors, and that their tendency to make these errors is related to their tendency to solve the problems. The results support the suggestion that people add erroneous constraints to the mental representations of these problems and the additional constraints hinder their ability to solve the problems.

***External representations***

We established that the counterbalanced order of representing by a diagram or an explanation did not influence performance (35% vs 45%, Mann -Whitney Z = .979, $N_1$ = 13, $N_2$ = 13, p = 0.328, two-tailed) before collapsing this distinction in further analyses. Participants solved the same number of problems when they represented them first and then tried to solve them as they did when they tried to solve them first and then represented them (42% versus 38%, Mann-Whitney Z = 0.362, $N_1$ = 14, $N_2$ = 12, p = 0.718, two-tailed) and there was no difference for single-step problems (53% vs 48%, Mann-Whitney Z = 0.237, $N_1$ = 14, $N_2$ = 12, p = 0.813, two-tailed), or for multiple-step problems (24% vs 17%, Mann-Whitney Z = 0.791, $N_1$ = 14, $N_2$ = 12, p = 0.429, two-tailed) as Table 2.1 shows. The result shows that drawing a diagram or providing a verbal re-description of a problem does not help in its solution, for this set of problems. In fact most participants who reached a correct solution tended to do so during the first

phase, whether they were producing a diagram or re-description, or whether they were attempting to solve the problem.

The experiment provides three new pieces of evidence about insight problem solving: (1) most participants make the conjectured error, and their tendency to do so is related reliably to whether they solve the problem or not, (2) solving these problems is not helped by constructing an external representation, whether diagrammatic or explanatory, and (3) multiple-step problems are harder to solve than single step problems. The experiment provides evidence on the nature of the erroneous constraints that participants introduce into their mental representations of the problems.

The results support the accounts such as that of Ohlsson (1992) which assert that an incorrect representation of a problem reduces the likelihood of a correct solution. Search and heuristic-based theories such as that of Chronicle et al (2004) suggest that participants refine their search of possibilities by choosing the most promising first move. The suggestion has been applied to multiple-step problems such as the nine-dot problem where there are a number of possible moves to choose from, but not to single-step problems such as the ping-pong ball. For these problems, it may be difficult to think of any possible starting points rather than it being difficult to choose from a myriad of options. As the analysis of their external representations indicates, in some problems such as the six-matches, participants might not get beyond a two-dimensional solution attempt because they seem to make progress within this dimension, and so they follow a 'locally-rational' strategy. But for other problems, search based accounts do not offer an explanation for why throwing a ball forwards could be judged to be more promising than throwing a ball upwards. Instead, participants who do not reach a correct solution have failed to consider the correct

representation completely. The results confirm that the expected representational error consistently occurs and this error is associated with failing to solve the problem, and so they strongly suggest a role for representation in the solution process. We will return to the role of search in the next series of experiments. Our aim in the next experiment was to examine ways to help participants to avoid introducing additional assumptions that lead to errors in their mental representations.

## Experiment 2

Insight problems may be difficult in part because people make erroneous assumptions that constrain their mental representations of the problems (Ohlsson, 1992). They may be difficult because of false assumptions in understanding the problem and the search for a solution (Segal, 2004). When an impasse is reached, incubation (taking a break from active problem-solving attempts) can help relax the false assumption. The aim of the second experiment was to help people to solve insight problems by ensuring that they avoided making the erroneous assumption in the first place. The first experiment provided evidence on the nature of the additional constraints that participants introduced into their mental representations and in this experiment we relied on that evidence to manipulate their initial representations to avoid the additional constraints. For example, in the ping-pong problem many participants add the constraint to their representation that the ball must travel on a horizontal trajectory. This assumption may be made because ping-pong balls generally follow a horizontal trajectory when in play. The assumption may block the solution, which depends on a vertical trajectory. Suppose instead you had been asked to describe how to throw a basketball so that it will go a short distance, come to a dead stop, and then reverse itself. Basketballs often

follow a vertical trajectory when in play and they are heavier than ping-pong balls, and so we predicted that participants would be unlikely to add unhelpful assumptions about a horizontal trajectory to their mental representation. The basketball version helps to relax the constraints otherwise placed on the solution by the background knowledge in the ping-pong ball version.

We re-worded various problems, sometimes by replacing just a single word (basketball instead of ping pong ball), to manipulate the assumptions that participants would make (see Appendix A). Our manipulations were based on the errors identified for specific problems in Experiment 1. However, we reduced the set of problem to four of the original eight, two each of the multiple-step and single-step categories. We selected the problems that were most suitable for our intended manipulations of wording changes and the provision of physical props (in Experiment 3). We compared these improved rewordings or 'constraint-relaxed' problems with the standard wording or 'constraint-assumed' problems. We predicted that participants would solve more of the constraint-relaxed problems than the constraint-assumed ones.

We also expected that the helpful re-wordings should be more effective for single-step problems than multiple-step problems. For single step problems, when participants can form an initial mental representation that does not contain unhelpful assumptions, they should be able to solve the problems readily because once the insight is achieved the solution is readily available. But for multiple-step problems, even when participants can form an initial mental representation that does not contain unhelpful assumptions, there remain several further steps to complete to reach the solution. For example, in the nine-dot problem, even when participants know they may draw lines outside of the square, they must still work out exactly which lines to draw (Chronicle et

al, 2001).

As a further exploration of the sources of difficulty in these problems, we also asked participants to suggest a hint for a friend attempting to solve the problem, after the participant knew the solution. Our aim was to examine what aspects of the problem the participants thought posed the greatest obstacle to a solution.

## Method

### *Participants*

The participants were 40 members of the psychology department's participant panel (recruited from the general population through national newspaper advertisements), who were paid a nominal fee of 8 euro per hour. There were 29 women and 10 men (and one participant did not provide information about gender). They were aged between 27 and 74 years, with a mean age of 54 (and two participants did not provide age information). Participants were assigned at random to the standard version of the problems (n = 22) or the reworded version (n = 18). Two participants from the reworded condition were subsequently eliminated because of an administrative error in their booklets.

### *Materials and design*

We constructed two sets of problems. Each set consisted of two single-step problems (the ping-pong and window-cleaner problems), and two multiple-step problems (the nine-dot and six-matches problems). The control 'constraint-assumed' set was based on the standard wording of the problems, and the experimental 'constraint-relaxed' set was based on an improved wording of the problems, designed to lead to a better

representation of the problem (e.g., for the ping-pong problem the reworded version referred to a basketball instead). The standard and reworded versions of the problems are presented in full in Appendix A. Participants were assigned to the control group or the experimental group at random. Each participant completed all four problems in a different random order.

The participants' first task was to attempt to solve the problem. Each problem was presented on a separate sheet of paper with space to attempt a solution. Their second task for each problem was to provide a hint to a friend. After they submitted their solution, they were given a second sheet of paper (which contained the solution to the problem on it). They were asked to imagine they were to provide a suitable hint to a friend on a game show who is allowed to ring for help with the problem, but who cannot be told the answer directly.

### *Procedure*

Participants were tested in small groups of two or three individuals. Each person completed four problems, either the standard wording 'constraint-assumed' versions or the improved wording 'constraint relaxed' versions. Participants were given a sheet with the problem on it. They had two minutes to read the problem and attempt a solution. This sheet was then taken up before they were given the second sheet with the solution. They were allowed one minute to read the solution and write down a hint before moving on to the next problem.

## Results and Discussion

We analysed only the 139 unseen problems (of the potential set of 152 problems, that

is, 38 participants given 4 problems), and scores based on the proportion of unseen correct solutions were calculated for each participant.

***Constraint relaxation***

Participants who were given the improved wording versions solved more problems compared to the participants given the standard wording versions (31% versus 18%, Mann-Whitney $Z = 2.019$, $N_1 = 22$, $N_2 = 16$, $p < 0.05$, and all tests are one-tailed unless otherwise stated) as Table 2.2 shows. Participants solved more single-step problems in the reworded versions than in the standard versions (57% versus 33%, Mann-Whitney $Z = 1.863$, $N_1 = 22$, $N_2 = 16$, $p < 0.05$). This trend of better performance with the improved wording held for both the window-cleaner (60% versus 36%)and ping-pong ball problems 53% vs 29%) but individually the value for chi-square did not reach significance levels: ($\chi^2 = 2.006$, df = 1, $p > .05$ and $\chi^2 = 2.258$, df = 1, $p > .05$, respectively). However, participants did not solve more multiple-step problems in the reworded versions than the standard versions (3% versus 0%, Mann-Whitney $Z = 1.863$, $N_1 = 22$, $N_2 = 16$, $p = 0.121$). The results indicate that the erroneous assumptions that people add to their mental representations of the problems can be manipulated, at least for some problems, and that when these erroneous constraints are relaxed people solve the problems more readily.

The multiple-step problems did not benefit from the re-wording, which could reflect a floor effect. It may be the case that two minutes is sufficient time for the single-step problems but not for the others. Alternatively, it may be that the representations of the nine-dot and six-matches problems are simply not influenced by the wording used.

**Table 2.2:** Percentages of correct solutions for standard and improved wording of single- and multiple-step problems in Experiment 2

| | | Standard | Improved |
|---|---|---|---|
| **Single-step** | Ping pong | 29 | 53 |
| | Window cleaner | 36 | 60 |
| | *Total* | *33* | *57* |
| **Multiple-step** | Nine-dot | 0 | 0 |
| | Six matches | 0 | 7 |
| | *Total* | *0* | *3* |
| *Overall* | | *18* | *31* |

***Single-step versus multiple-step***

As in the previous experiment, participants solved more single-step than multiple-step problems (42% versus 2%, Wilcoxon's $Z = 4.138$, $N = 23$, $p < 0.0005$) both in the standard version of the problems (33% versus 0%, Wilcoxon's $Z = 2.762$, $N = 9$, $p < 0.01$) and in the 'constraint relaxed' improved version (57% versus 3%, Wilcoxon's $Z = 3.106$, $N = 14$, $p < 0.01$) as Table 2.2 shows. The result replicates the findings of the previous experiment and supports the distinction between these two types of problems.

***Conjectured errors***

Participants' solution attempts were classified as indicating the conjectured misrepresentation or not based on their written rough work in the booklet, e.g. solutions to the ping-pong ball problem were classified as indicating a horizontal trajectory or not (see Appendix A for the conjectured errors). An independent rater classified solutions for over one-third of the booklets and agreement was 97%. Participants made the

conjectured errors of representation more often for multiple-step problems than single-step problems (98% versus 52%, Wilcoxon's $z = 3.543$, $N = 17$, $p < 0.001$, two-tailed). For the 32 correct solutions produced overall, participants' produced the conjectured error for only 5 cases, whereas for the 107 incorrect solutions produced overall, they produced the conjectured error for 83 cases (a further 24 cases had no rough work to classify). Participants who misrepresented the ping-pong problem as indicating a horizontal trajectory had more incorrect than correct solutions (92% vs 8%, $\chi^2 = 14.00$, $df = 1$, $p < 0.0005$) as did those who misrepresented the window-cleaner problem as indicating the man was at the top of a high ladder (80% vs 20%, $\chi^2 = 20.19$, $df = 1$, $p < 0.0005$). For the six-matches problem, the only participant to produce a correct solution was also the only one not to misrepresent it. For the nine-dot problem, there were no correct solutions and every participant misrepresented it in the conjectured way. The result replicates the findings from Experiment 1 that erroneous assumptions in the mental representation of insight problems play a functional role in hindering their solution.

A second measure of participants' mental representations is available in the 'hints to a friend' that participants provided when they knew the solution. Most participants provided hints that targeted the conjectured misrepresentation for the single step problems. An independent rater classified solutions for over one-third of the booklets and agreement was more than 90%. The most frequent hint for the ping-pong ball problem referred to the ball travelling on the vertical rather than the horizontal plane (42%), and hints referring to gravity were also common (39%). Most of the remaining hints put an emphasis on throwing the ball (11%). The most frequent hint for the washing windows problem referred to the man's height off the ground (78%). The

next most frequent hint emphasised ground level (11%). The most frequent hint for the multiple step problem of the six-matches was to refer to the shape of pyramids (43%), either indirectly by giving a clue about Egypt (29%) or directly by mentioning pyramids (14%). The next most frequent hint was to suggest using three dimensions (31%). Perhaps unexpectedly, in the nine-dot problem, most of the suggested hints referred to the general shape of the solution, especially the triangular portion (63%). The next most frequent hint was to extend lines beyond the square (23%).

The experiment provided three further pieces of evidence about the nature of the mental representations that people construct for insight problems: (1) the erroneous constraint that participants add to their mental representation can be relaxed by re-wording the problems. Improved wording that removes common assumptions (e.g., the trajectory of the ball, or the height of a man on a ladder), successfully helps people to solve some insight problems. Participants solved more problems when they were given the improved wording compared to the standard wording. (2) The constraint relaxation manipulation of improved wording helped with the single-step problems, but not with the multiple-step problems. As in the previous experiment, participants solved more single-step problems than multiple-step ones. (3) Participants' most frequent hints to a friend targeted the conjectured error. As in the previous experiment, participants who made the conjectured error, as evidenced in their rough work, tended not to solve the problems. These converging measures -- of external diagrams and re-descriptions in the first experiment, and rough work and hints to a friend in the second experiment -- provide strong evidence of the nature of the erroneous assumptions that people add to their mental representations of these insight problems.

The experiment shows that it is possible to improve performance on insight

problems, at least for the single-step variety, by helping people to avoid the typical erroneous representational assumptions. People did not include the erroneous constraints in their mental representations, such as the constraint that the ball had to travel on a horizontal trajectory, when the problem was re-worded in such a way that participants could access different information from memory (basketballs often travel in the vertical trajectory in play). The attempt to help people to avoid adding the erroneous constraints to their mental representations was successful for the single-step problems but not for the multiple-step ones. The failure to improve on the multiple-step problems could be due to insufficient time allowed or simply that the representation of these problems is not influenced by the wording. The representation of these problems may rely on past experience of superficially similar problems, for example, join-the-dots problems. Alternatively it could be that the difficulty of these problems is more influenced by the need to search a large problem space and less so by the representation.

As the solution rates failed to reach 100%, even for the problems that did benefit from the wording changes, this shortage leaves room for the possibility that another factor is contributing to the difficulty of these problems, perhaps the need to generate and search through various possibilities. Nonetheless, the examination of conjectured errors and the hints offered by participants indicate a role for misrepresentation in each of the problems. Our next experiment attempts to help people to avoid adding erroneous constraints in their mental representations of multiple-step problems as well as single-step problems.

## Experiment 3

The aim of the experiment was to help people to solve insight problems by ensuring that they relaxed any additional constraints that they had erroneously introduced in their mental representations of the problems. It attempts to improve the mental representations by ensuring that any erroneous assumptions that are made are immediately salient and so they can be readily identified and relaxed. We gave participants the physical elements of the insight problems to accompany the written description of each problem (see Appendix A for a description of the props used). For example, for the six-matches problem we gave participants six small rods and some blue-tack. They were asked to show how the rods can be used to form a shape that is made up of four equilateral triangles (and each side of every triangle must consist of just one unbroken match). Many participants add the erroneous constraint that the shapes must be constructed in two-dimensions (for example flat on a table), as we showed in the first experiment. If they add this constraint, their exploration with the physical materials is likely to ensure that they relax the constraint and consider three-dimensional shapes. Weisberg and Alba (1981) and Tsai (1987) have both previously used physical props with the six-matches problem, and both studies found that additional hints were needed to increase solution rates. However, we identified practical difficulties in getting the tetrahedron solution for this problem to stand independently without the use of blue-tack and so those participants in those earlier experiments may have been discouraged from attempting to complete solutions in three-dimensions. Hence, in our physical props we included both rods and blue-tack. Physical props have not previously been used in any of the other three problems, to our knowledge. We compared these physical insight problems with the standard pen-and-

paper problems. As in Experiment 2, the props used were based on tackling the errors identified in Experiment 1.

We also took the opportunity to check an untested assumption in experiments on insight problems. Previous studies have tended to limit the time available to participants to solve the problems, for example, 3 minutes to solve the nine dot problem (Chronicle et al., 2001), or 6 minutes each for a mixture of problems (Schooler et al., 1993). It has been suggested that longer times do not result in more solutions (Lockhart et al, 1988). We gave participants an extended time period (10 minutes), and we gave them feedback when their initial solution attempts were wrong to encourage them to continue working throughout this time. We kept a record of the time each participant took to reach the correct solution. In this way we can identify and compare how many problems were solved within the standard time frame for example, of two minutes (Lockhart et al, 1988), and how many were solved within the extended time frame of ten minutes.

## Method

### *Participants*

The participants were 40 students from Dublin University, Trinity College who participated voluntarily and in some cases in return for course credits. There were 23 women and 17 men, and their ages ranged from 17 to 49 years, with a mean age of 19. They were assigned at random to the written version (n = 20) or the physical version (n = 20).

### *Materials and design*

We used four of the problems used in the previous experiments (the nine-dot, six-

matches, ping-pong and window-cleaner problems). The control group of participants were given the problems in the standard wording version, in written form (see Appendix A). Each problem was presented on a separate page in a booklet, on which they were asked to record their rough work and solution. The experimental group were given physical props that they could use to solve the problem, but were not allowed pen-and-paper. For the six-matches problem, the prop was six wooden rods (measuring 8mm in diameter and 250mm long) and a packet of blue-tack. For the nine-dot problem the prop was a corkboard with nine red pins arranged in the matrix formation, some extra pins in a cup, a ball of string and scissors. For the ping-pong problem the prop was a standard white ping-pong ball. For the window cleaner problem the prop was a model wooden ladder (measuring 595mm long and 40mm wide) and a small figure of a man (measuring 70mm in height). Participants in both conditions received the identical wording for each problem (the standard wording that appears in Appendix A, with the exception that the wording of the six matches problem in this experiment referred to 'rods' instead of 'matches', and blue tack was referred to in both the physical and pen-and-paper versions). The participants' task was to solve the problems. They were assigned to the control group or the experimental group at random. Each participant was given the four problems in a different random order. In addition, they were given a rating sheet at the end of the experiment to rate the difficulty of each problem (easy, average or difficult).

***Procedure***

Participants were tested individually. After reading the instructions (see Appendix A) and confirming their understanding of them, they were given each problem on a

separate sheet. For participants in the physical condition, the materials for each problem were placed on the table in front of them along with the written description and with the comment that here were 'some materials you might like to work with'. Participants in the physical condition were not allowed to use pen–and-paper.

Participants were allowed to work on each problem for ten minutes. They were timed using a digital stopwatch. When they thought they had the correct answer, they announced it and the stopwatch was paused. If the solution was correct, the time of the announcement was recorded and they moved on to the next problem. If the solution was incorrect, they were given this feedback, they were asked to refer back to the written description of the problem, and the stopwatch was restarted. There was no limit on the number of attempts a participant could make within the 10 minute time period. They were told when they had one minute remaining.

The experimenter recorded whether or not a correct solution was reached, the time taken to reach the correct solution (in seconds), and the number of attempts made. Participants were asked if they had seen each problem before and no attempts were made for familiar problems. Participants were given the solutions to the four problems at the end of the testing session, and then they provided their difficulty ratings for all four problems.

**Results and Discussion**

142 unseen problems (of a potential set of 160 problems) were analysed. A score was calculated for each participant based on the proportion of unfamiliar problems solved correctly.

***Physical props and correct solutions***

As expected participants solved more problems when they had physical props than when they relied on pen and paper, after 10 minutes (75% vs 58%, Mann-Whitney Z = 2.116, $N_1 = 20$, $N_2 = 20$, $p < 0.05$, and all tests are one-tailed unless otherwise stated). Physical props improved performance reliably on the multiple-step problems (57% vs 37%, Mann-Whitney Z = 1.654, $N_1 = 20$, $N_2 = 20$, $p < 0.05$). Physical props appeared to improve performance somewhat on the single-step problems (89% vs 79%), but the difference was not reliable (Mann-Whitney Z = .853, $N_1 = 20$, $N_2 = 20$, $p = 0.196$). In fact, as Table 2.3 shows, participants solved more problems when they had physical props for one of the single step problems, the window-cleaner problem (100% vs 79%, Fisher's exact test, $p = 0.05$) but not the other, the ping pong problem (79% for both, $\chi^2 = 0$, df = 1, $p = 0.65$). Likewise, they solved more problems when they had physical props for one of the multiple-step problems, the six matches (78% vs 35% $\chi^2 = 7.01$, df = 1, $p < .005$), but not the nine-dot problem (25% vs 38%, Fisher's exact test, $p = .39$). The results show that it is possible to improve performance on at least some multiple-step and some single-step insight problems by providing physical props. The physical props ensure that participants can relax any additional erroneous constraints that they have added to their mental representation of the problem, for example, they can remove the constraint that the shapes from the six-matches must be in two-dimensions.

However, the props were not successful in helping all of the problems. The props used in the nine-dot problem did not allow participants to record their previous attempts, in comparison to the pen-and-paper condition, and so participants may not have realised that they were trying the same futile moves that they had already eliminated. The props provided in the ping-pong ball problem were not readily utilised

- participants did not tend to throw the ball to test their solution suggestions. However, the prop did improve solution times as we will describe shortly.

Once again, participants solved more single-step problems than multiple step problems (84% vs 45%, Wilcoxon's $z = 3.231$, $N = 25$, $p < 0.01$). The result is in keeping with our earlier hypothesis that there is a valid distinction to be drawn between these two categories of insight problems.

**Table 2.3:** Percentages of correct solutions for paper-and-pencil and physical prop versions of the single-step and multiple-step problems in Experiment 3 after 10 minutes, with correct solutions after 2 minutes provided in parentheses.

| | | **Paper** | **Physical** |
|---|---|---|---|
| **Single-step** | Ping pong | 79 (53) | 79 (63) |
| | Window cleaner | 79 (37) | 100 (68) |
| | *Total* | *79 (47)* | *89 (66)* |
| **Multiple-step** | Nine-dot | 38 (6) | 25 (0) |
| | Six matches | 35 (5) | 78 (22) |
| | *Total* | *37 (6)* | *57 (13)* |
| *Overall* | | *58 (27)* | *75 (43)* |

There were more correct solutions after 10 minutes then after 2 minutes (66% vs 35%), and this was so for the single-step problems (84% vs 57%) and the multiple-step problems (45% vs 9%). It is common in insight studies to provide participants with just a few minutes to solve the problem, and this result suggests that previous research may have underestimated participants' ability to solve insight problems by placing a restrictive time limit on their solution attempts, especially for multiple-step problems. Although the solution rate for the single-step problems was greater when extra time was allowed, the rate was still well above floor level even after only two minutes. As can be seen in Table 2.3, the biggest improvement was seen in the multiple-step problems with noticeably few solutions in the first two minutes, especially in the

control condition.

*Solution times*

**Table 2.4:** Mean latencies to produce the correct solutions in seconds (log transformed to the base e) for the problems in Experiment 3 in bold (with standard deviations in brackets) and pre-transformation means in italics

| | | **Paper** | **Physical** |
|---|---|---|---|
| **Single-step** | Ping pong | **4.39 (.92)** | **3.98 (1.19)** |
| | | *116.80* | *107.40* |
| | Window cleaner | **4.65 (.85)** | **4.15 (1.10)** |
| | | *140.13* | *112.05* |
| **Multiple-step** | Nine-dot | **5.71 (.61)** | **5.73 (.79)** |
| | | *163.44* | *366.00* |
| | Six matches | **5.71 (.81)** | **5.19 (1.08)** |
| | | *377.43* | *264.07* |

Time in seconds to reach a correct solution was log transformed to the base e before analysis in a 2 (paper vs physical) by 2 (single-step vs multiple-step) ANOVA with repeated measures on the second factor. Participants produced the correct solutions somewhat faster when they had a physical prop then when they relied on pen-and-paper, but the result is not reliable (4.73 sec vs 5.2 sec, $F(1,22) = 2.996$, $Mse = .879$, $p = .097$), as Table 2.4 shows. (And throughout the latencies reported are the log-transformed latencies). They took longer to produce the correct solution to the multiple-step than the single step problems (4.44 sec vs 5.50 sec, $F(1,22) = 13.76$, $Mse = .932$, $p < 005$). The two factors did not interact ($F(1,22) = .01$, $Mse = .932$, $p = .92$). Separate t-tests on each of the four problems confirmed that the props did not result in significantly faster times on any of the problems (and all tests are one-tailed): nine-dot ($t = 0.37$, $df = 7$, $p = .486$), six-matches ($t = 1.132$, $df = 19$, $p = .136$), ping-pong ($t =$

1.067, df = 28, p = .148), although the t-value for the window-cleaner approached significance (t = 1.460, df = 32, p = .077) and in the expected direction. However, the construction of boxplots on untransformed times suggested three outliers for each of the ping-pong and window-cleaner problems. These participants had taken an anomalously long time to reach a solution compared to the other participants in the group, suggesting they may have been distracted at some point. T-tests were re-ran with these outliers removed and the difference between the mean transformed time with props was significantly lower than without props: ping-pong (t = 2.286, df = 23, p < .02) and window-cleaner (t = 2.878, df = 26, p < .005).

**Table 2.5:** Mean latencies to correct solutions for problems solved on the first attempt in Experiment 3 (log transformed to the base e) with pre-transformation means in parentheses

| | | **Paper** | **Physical** |
|---|---|---|---|
| **Single-step** | Ping pong | 4.10 (96.30) | 3.24 (32.86) |
| | Window cleaner | 4.57 (193.38) | 3.64 (42.02) |
| **Multiple-step** | Nine-dot | 5.66 (335.60) | 5.73 (366.00) |
| | Six matches | 5.44 (290.80) | 4.67 (143.78) |

Participants produced a correct solution at their first attempt faster when they had a physical prop than when they relied on pen-and-paper. (We analysed these data in a series of one-tailed t-tests because there were few data points in some instances for ANOVA). As can be seen in Table 2.5, they were faster for each of the single-step problems: ping-pong (3.24 sec vs 4.10 sec, t = 1.995, df = 15, p < .05) and window-cleaner (3.64 sec vs 4.57 sec, t = 2.92, df = 24, p < .005). The difference approached significance for the multiple-step six matches problem (4.67 sec vs 5.44 sec, t = 0.47, p = .80), but there was no difference for the nine-dot problem (5.73 sec vs 5.66 sec, t =

0.13, df = 6, p = .45). The results suggest that for some insight problems physical props can help participants to reach a solution faster.

Solution rates for the window-cleaner problem increased reliably with the use of physical props and the time to solution was faster with outliers removed and when the problem was solved on the first attempt. A similar pattern was observed for the six-matches problem but the time difference did not quite reach significance. Props did not increase the frequency of correct solution for the ping-pong ball problem but solution times were faster with props when outliers were removed and when participants who solved the problem on their first attempt. The nine-dot problem did not benefit from the provision of props on any measure.

***Difficulty ratings***

Participants rated the difficulty of each problem as easy, average or difficult (scored 1, 2 and 3). We analysed their ratings in a 2 (paper vs physical) by 2 (single-step vs multiple-step) ANOVA with repeated measures on the second factor. They rated the multiple-step problems as more difficult than the single-step problems (M = 2.33 vs 1.45, F(1,38) = 37.12, Mse = .413, p < .0005), as Table 2.6 shows. Physical props did not affect the difficulty rating of the problems (M = 1.85 versus 1.93, F (1, 38) = 0.73, Mse = .155, p = .4). The two factors did not interact (F(1, 38) = 0.30, Mse = .413, p = .86).

As expected, the difficulty ratings for each of the four problems were negatively correlated with participants' success at solving the problem (Spearman's rank and n = 40 in all cases). This pattern applied to both of the multiple-step problems, nine-dot (r = -.371, p < .05) and six matches (r = -.727, p < .0005). It was also the case for both of the single-step problems, window-cleaner (r = -.339, p < .05), and marginally for the

ping-pong problem ($r = -.257$, $p = .06$). The result shows that participants rated the problems they had not solved as the more difficult ones.

**Table 2.6:** Mean difficulty ratings for each problem in both conditions in Experiment 3 based on a scale of 1 – 3 where 1 = easy

| | | **Paper** | **Physical** |
|---|---|---|---|
| **Single-step** | Ping pong | 1.42 | 1.47 |
| | Window cleaner | 1.58 | 1.37 |
| **Multiple-step** | Nine-dot | 2.25 | 2.67 |
| | Six matches | 2.55 | 2.06 |

The experiment provides some further evidence about the constraints that people introduce to their mental representations of insight problems and how they may be relaxed. (1) Participants solved more problems when they had physical props than when they relied on pen and paper. They solved more problems when they had a physical prop for single-step and multiple-step problems overall. However, the effects of physical props are clearly dependent on the content of the problem, since the props helped only one of the single-step problems (the window cleaner problem) and one of the multiple-step problems (the six-matches problem).

Why did the other props not work? Perhaps they did not highlight that the additional constraint was erroneous (in retrospect the provision of a ping pong ball and nothing else may help people include the given constraint not to attach anything to the ball, but it may not help them to realise that the ball need not travel on the horizontal trajectory) or it may only have helped people who would eventually have reached a solution to produce it faster. Perhaps not being able to record previous attempts or ideas led to a working memory load that negated the positive effect of the props, particularly

in relation to the nine-dot problem.

(2) Participants produced a correct solution at their first attempt faster when they had a physical prop, reliably for both single step problems, and marginally so for one of the multiple step problems (six-matches). The result provides converging evidence that physical props can help people to solve insight problems. (3) Participants produced more correct solutions after 10 minutes then after 2 minutes: twice as many solutions for single step problems, and five times as many solutions for multiple step problems. The result suggests that people may have a greater ability to solve insight problems once they are given the time to do so, particularly in the case of the multiple step problems. People may need more time particularly when they receive feedback and make more attempts. (4) People solved more single step problems than multiple step problems, and this result is consistent with our previous experiments. They also produced the correct solutions more quickly for the single step problems than the multiple step problems, and they rated the single-step problems as less difficult than the multiple-step ones. These converging measures validate the idea of two distinct categories of insight problems.

We suggest that the physical props help people to relax the constraints on their mental representations. An alternative explanation for the efficacy of props is that participants have a limited ability to reason spatially in the absence of props, for example in the six-matches problem. However, the window-cleaner problem benefited from the props and from the wording changes, which suggests that its difficulty does not specifically depend on spatial reasoning ability. Nonetheless we refer to the role of spatial ability in later experiments.

Another possibility is that physical props for the six-matches problem provided

a clue to work in three dimensions regardless of their nature. However, we observed that even participants who had the physical props for the six-matches problem started with 'flat' configurations before moving on to work in three dimensions. Participants who received the booklet had the problem worded as using 'rods' and 'blue-tack' as well, which did not seem to suggest a three-dimensional solution in itself.

The findings suggest that the provision of physical props can help participants to solve insight problems, but they may not be appropriate for every insight problem. For example, in the marrying-man problem one is asked how a man can marry twenty women without breaking the law, with no divorce or bereavements. The answer is that the man is the priest who married twenty brides to twenty different grooms, but providing twenty female figurines and one male figurine may not help participants to reach the correct solution more often or any faster.

## General Discussion

The experiments reported in this chapter confirm that how an insight problem is represented is important to the solution process. Experiment 1 showed that participants tend to make the expected misrepresentation for each problem in a range of insight problems and this error has a significant impact on the likelihood that they will reach a solution. Most participants who misrepresented a problem failed to solve it within the time allowed in both Experiments 1 and 2.

We also demonstrated that manipulations aimed at overcoming each expected error helped to improve solution rates on insight problems. Experiment 2 showed that the wording of some problems can make some aspects of the problem appear more salient than they should be, leading to an overly constrained representation of the

problem. In the ping-pong ball problem, participants were constrained by their knowledge of how a ping-pong ball normally travels. In the window-cleaner problem, participants assumed that references to the height of the ladder and the office-block were important and constrained their representation to one where the man falls the full distance. Changing the wording slightly, in a way that did not rule out the need for insight, was sufficient to boost performance on these two problems.

The nine-dot and six-matches problems, however, did not benefit from a wording designed to help participants create a better representation. The nine-dot problem rewording was aimed at trying to relax the emphasis on the geometrical properties of the square-shape and the shape of the solution by getting participants to draw their own matrix and refocus the aim as marching through aliens rather than drawing lines. Participants were still inclined to make the error of confining lines within the square, however. Similarly, participants who were told to form a shape and that they could also use blue-tack on the six-matches problem still made the two-dimensional error. The failure of the rewording in relation to these problems may be because the standard wording does not in itself promote the representational error in the same way that it does for the other problems. Unlike the ping-pong and window-cleaner problems, the wording does not serve to focus attention on particular aspects of the problem such as the specific kind of ball, or the specific height of the ladder. Therefore, changes to the wording that do not explicitly tackle the conjectured error may not be useful to participants.

Alternatively, it may be that these problems are inherently more difficult, perhaps because of the need for multiple-steps, and that two minutes was not enough time for people to work through to a solution even with the rewording. In Experiment

3, where participants were allowed up to 10 minutes and multiple attempts, solution rates were noticeably higher in both the experimental and control conditions for the nine-dot and six-matches problems compared to Experiments 1 and 2.

Experiment 3 demonstrated that targeting misrepresentations through the use of physical props benefited participants on some insight problems. We expected that the physical props would help participants to recognise their own mistaken assumptions by allowing them to interact with the problem. For example, when provided with just a ladder and the figure of a man for the window-cleaner problem, participants are helped to focus on these elements and deterred from introducing other elements not explicitly mentioned in the problem. For this problem, when given the physical materials, all participants were able to reach the correct solution and they were faster to solution on the first attempt. Correct performance on the six-matches problem was also improved by the provision of actual rods and blue-tack and first attempt times were marginally faster.

Participants did not solve the ping-pong ball problem more often when they were given the prop but those who did solve it on their first attempt (as is the usual way of collecting answers) did so faster when they had the prop. The aim of this prop was to allow participants to see for themselves that it was not possible to throw the ball horizontally and have it return by itself. We observed, however, that participants were reluctant to throw the ball. Even though it may have helped them to rule out ideas such as throwing the ball to another person, the ball in itself may not have helped them to generate other possibilities to consider.

The nine-dot problem was the only problem that did not benefit in any way from the props in terms of correct solutions or time to solution. The prop was a board

with nine pins, some string and extra pins. Participants given the prop had no means of recording their previous attempts and so may unwittingly have found themselves trying the same moves repetitively. In contrast, participants in the control condition had pen-and-paper and could record their earlier attempts. The recording of moves may not be a difficulty for some multiple-step problems such as the six-matches problem because there are not so many promising moves, or because one does not get too far beyond the first move before realising that a solution is impossible.

The three experiments provide support for the idea that how a problem is represented plays a central role in how people reason with insight problems. The ping-pong ball problem had more correct solutions when it was reworded and faster time to solution when props were used. The window-cleaner problem had more correct solutions with both rewording and props. The six-matches problem had more correct solutions when props were used, but not when the wording was changed. The nine-dot problem was the only problem that was not improved by either props or wording. However, erroneous mental representations may not be the only difficulty that participants have to contend with when they reason with insight problems. Not all problems benefited equally well from manipulations to alter the misrepresentation. Moreover, in only one problem (the window-cleaner problem in Experiment 3) did solution rates for a problem reach ceiling levels in the experimental condition, and so for all other problems there were some participants who still failed to solve even when they were assisted in improving their representations of the problem.

In relation to the representation-based theories, the results of our experiments are consistent with the mechanism of 'constraint-relaxation' (Ohlsson, 1992). Participants' initial representation of the problem includes constraints that are not made

explicit in the problem, but may be suggested. For example, there may be a tendency for participants attempting the ping-pong ball problem to focus too much on the ping-pong ball itself and constrain their solution attempts to scenarios that are consistent with their existing knowledge of how ping-pong balls normally travel.

Although our findings are consistent with the representation-based theories, neither are they inconsistent with the search-based theories. After all, the solution attempts offered are usually sensible, even if they are incorrect. A common suggested solution to the ping-pong ball problem was to "put a back-spin on the ball", which would work if one was rolling the ball across a surface, but the problem clearly states that the ball must be thrown, not rolled. There may well be a tendency to offer a solution that seems, loosely, to fulfil the criteria for a solution and discontinue the search for a solution, as opposed to rigorously searching for the optimal solution. Crucially, however, this search for a satisfactory solution often stems from the initial, incorrect representation. Heuristics or strategies may then be employed but not independently of the representation. For some insight problems at least, the representation is an important factor.

The first important finding is that the data suggest that how the error in representation comes about may vary between problems. In the ping-pong ball and window-cleaner problems, the wording directs the individual's attention to aspects of the problem that seem important but actually are not (a ping-pong ball, a sixty-foot ladder). Once attention is drawn to these aspects, participants' representations may be biased by their existing knowledge or previous experience of them (e.g. going back and forth in play, using a ladder to reach upper-storey windows). For other problems, such as the six-matches and nine-dot problems, participants may draw solely on their

previous experience of related materials. Although the wording does not specifically direct participants to a particular aspect of the problem, neither does it help to direct them away from the idea of tackling the problem in a similar way to conventional join-the-dots tasks, for example. Furthermore, for multiple-step problems one has to start somewhere and starting with something from past experience seems like a good idea.

Possibly the props help participants to concentrate on the actual elements of the problem and not introduce other elements that are not mentioned in the problem (e.g. a mattress in the window-cleaner problem). The props may also help participants to test hypotheses about what might work and get more accurate feedback (such as trying to throw the ping-pong ball with a back-spin). Such feedback may not be available when participants try to mentally simulate the elements and the outcome of actions.

A second important finding of the experiments was the confirmation of the distinction between multiple-step and single-step problems. In all three experiments, the multiple-step problems were more difficult. Only in the third experiment when more time and multiple attempts were allowed did solution rates really get above floor level for these problems. We speculate that multiple-step problems are more difficult primarily because there is more work to be done than in the single-step problems. For these problems there is an additional requirement to plan a path to solution and it may be necessary to investigate a potential solution path to some depth before one realises that it will not, after all, lead to a correct solution. Furthermore, a good initial move may not be properly evaluated, and hence discarded, because its quality is not readily apparent. For example, the move of drawing a central diagonal line in the nine-dot problem may not initially seem any better than drawing a line along the side. In contrast, the process of evaluating an idea in a single-step problem is simpler because it

may be easier to make a judgement about whether it could work or not.

We also observe, however, that multiple-step problems tend to have a strong visual or spatial element that is not so consistently found in single-step problems. Therefore there is a possibility that their lower solution rates may reflect a general difficulty with spatial reasoning among participants. This idea may provide an alternative explanation for why the physical props were helpful for the six-matches problem. However, the window-cleaner problem benefited from both props and wording changes, and so the spatial distinction is unlikely to be the sole explanation of the differences in solution rates. In the next series of experiments, we examine some single-step problems that have a visual or spatial element. Multiple-step problems that are primarily verbal in nature are, however, less common in the literature (and in fact we could find only one).

Multiple-step problems are consistently more difficult and the reason for this difficulty, whether it is the number of moves, the spatial element or a combination, may prove an important distinction between the two sets of problems. The implication of this distinction between multiple-step and single-step is that any general theory of insight has to account for both sorts of problems. Results of research using one type may only be completely generalisable to insight problems of the same sort. A general understanding of the process of 'insight' may require attention to both sorts of problems.

In summary, this chapter identified a distinction between insight problems that can be termed multiple-step and those that are single-step. Although we speculated that the reason multiple-step problems are more difficult is due to the distance in terms of further required steps between the insight and solution, and recognising the insight as

such, other explanations for this difference are possible. For example, the difference between the two problem types may be due to multiple-step problems typically being spatial and single-step typically verbal. We will examine this possibility in the next chapter. For both kinds of problems, an error in the initial representation was associated with subsequent failure to solve the problem. Manipulations aimed at altering the misrepresentation through wording changes and physical props reliably improved performance on a range of problems, but not all problems benefited equally well from all manipulations. Our results indicated an important role for mental representation in the solution process but also suggest that other factors, such as the difficulty in navigating the problem-space and the use of heuristics may have a role to play. In the next chapter, we investigate the possible role for search space.

# CHAPTER 3

Insight problems tend to be less well-defined than some other problems, for example mathematics problems. Sometimes a problem is ill-defined in terms of the way it is worded, encouraging a particular interpretation of the situation that is not fully accurate (e.g. the window-cleaner problem). In other situations, it is the nature of the solution that can be open to interpretation (e.g. the nine-dot problem). In between the problem and the solution is the 'problem space' (after Newell & Simon, 1972). The navigation of the problem space requires evaluating possible moves so that one can progress from the problem to the solution. As we saw in Chapter 1, some theories place emphasis on how the problem and the solution are represented as these representations influence what knowledge is brought in to try and solve the problem (e.g. Ohlsson, 1992). Other theories, however, emphasise the problem space and particularly the processes employed in its navigation (e.g. Ormerod et al, 2002).

In Chapter 2, we saw that it was possible to improve participants' performance on a range of insight problems by targeting their erroneous initial mental representations of a problem. However, some of the problems, such as the nine-dot, were not improved by manipulations targeting the mental representation. Furthermore, only one of the problems reached a hundred percent solution rate suggesting that the mental representation alone may not fully account for the difficulty of an insight problem. In the series of experiments reported in this chapter, our aim was to explore what role the search of the problem space played in a selection of insight problems.

### *Problem space in insight problems*

In the case of the nine-dot problem, Kershaw and Ohlsson (2004) have identified

multiple sources of difficulty to be overcome including not only the square shape of the matrix but also a reluctance to turn on a non-dot point. This finding is in keeping with our earlier assertion that multiple-step problems such as the nine-dot are more difficult than single-step problems because there is more work to be done, that is more steps to negotiate, even after the initial insight. One other aspect that multiple-step problems tend to have in common is the requirement to make moves: draw four lines, assemble six matches, move two coins. A participant has to plan a sequence of moves where there is an array of possible starting moves, a number of pieces to select from and numerous locations to which it can be moved. Chronicle et al (2002) have argued for a role for 'locally-rational strategies' whereby participants select a starting move on the basis of how close it appears to bring them to the goal state: a maximising strategy. They have predicted the likelihood of certain moves in the nine-dot problem (MacGregor et al, 2001) and with coin problems (e.g. Chronicle et al, 2004). With the eight-coin problem, they found that re-arranging the starting array of coins so that the most desirable maximising move (which was not actually the correct starting move) was no longer available, improved solution rates for this problem (Ormerod et al., 2002). There is also a possibility, however, that the large number of choices of moves may be a difficulty in itself and that just reducing the number of move options to be considered, without necessarily targeting the erroneous, maximising move could improve solution rates. Weisberg and Alba (1981) suggested that, in relation to the nine-dot problem, the size of the domain of possible within-square moves meant that participants may never exhaust all the possibilities and so fail to even become aware that another domain of outside-square moves existed. Although Lung and Dominowski (1985) argued that the domain of potential moves was not as vast as suggested by

Weisberg and Alba.

However, not all insight problems require the physical manipulation of objects and not all insight problems can be considered 'move' problems in the same sense as, for example, the six-matches problem. A problem such as the marrying-man relies on the limited interpretation of the verb 'to marry'. At first glance, it seems that a theory based on searching and evaluating possible moves cannot account for problems such as this one.

There is a common link, however, between insight problems that focus on interpretation and those that focus on moving pieces. One of the processes necessary for insight may be the ability to consider alternative possibilities. People may need to search through 'paths' in a problem space that could lead to a solution, and to judge whether a path merits further exploration, in other words to consider a possibility initially and decide whether it is worth developing further. Smith (1995) refers to 'probes', which are sets of information elements that are used as cues to search memory for knowledge that can then be analysed. Similarly, Keane (1989) suggests that the elements of a problem act as cues in searching long-term memory for potential solution plans. The ill-defined nature of insight problems results in many alternative possibilities, or 'paths' to choose. For example, in the ping-pong ball problem the solution is that the ball is thrown up. On initial consideration of the problem there are a number of possibilities that could come to mind and merit further exploration. Participants appear to consider many potential 'solution paths' or possibilities: the solution might be something to do with the properties of the ball (a ping-pong ball is small and very lightweight); it might be something to do with how the ball is thrown (you put a back-spin on the ball); it might be something in the environment (a gust of

wind or a friend who catches it and throws it back). The possibility that is given priority may be dictated by the participants' representation, but there is a choice to be made nonetheless. Participants frequently produce the incorrect solution to put a back-spin on the ball, as the previous experiments have shown.

In a similar way, participants attempting the nine-dot problem have to choose which dot to start on, but the dots might seem equally feasible at first glance. Having elected a given dot, they must then choose a direction in which to draw a line: north, south, east or west. There are a huge number of possibilities to consider. Both problem-types, 'move' and 'non-move' require that a number of possibilities be considered when trying to work out a solution.

***Assisting in the search for a solution***

In Chapter 2 we looked at ways of helping participants to solve insight problems by improving their representation of the problem. In this series of experiments we focused instead on helping participants by trying to make their search of the possibilities for each problem easier by reducing the number of possibilities.

It may be helpful to limit a participant's search to ensure they do not explore unfruitful paths, such as the back-spin possibility in the ping-pong ball problem. But 'positive' hints that tell a participant what to consider have not been very effective (see Chapter 1). We took the opposite tack. A 'negative' hint that tells a participant what not to consider may be effective. Wiley (1998) suggests that participants may assume that a particular domain contains the solution to a problem and confine their search to that area, leading to a kind of fixation or inappropriate focusing on one particular domain. She found that participants who had an extensive knowledge of baseball were

more likely to incorrectly suggest a baseball term as an associate of three target words when the first two of these could be baseball terms, than participants who were not as familiar with baseball terminology. Therefore manipulating a search for a solution by diverting efforts away from an unhelpful domain may benefit performance on insight problems, where there is a tendency to assume the solution lies in a particular domain when in fact it does not. Similarly Knoblich et al (2001) observed in an eye movement study using matchstick equations problems that participants were inclined to concentrate on the matches they assumed would be the key to the solution. Participants who did less well on the problems where the key match was not one of the obvious candidates were less inclined to shift their attention to consider the other matches that they had been ignoring.

In summary, our aim in this chapter was to discover if search space was a significant factor in the difficulty of insight problems by manipulating the range of possibilities participants have to consider. In the first experiment of this series, Experiment 4, we constrained the number of possible moves participants had to consider in insight problems that involved the physical manipulation of pieces. In the second experiment, Experiment 5, we constrained the search by providing negative hints to problems that did not involve moving physical pieces, but still had a number of possible solution paths to be considered.

## Experiment 4

Our aim in the first experiment of this series was to examine whether simply limiting the range of possibilities a participant has to consider can improve solution rates. Following from the existing research of Ormerod and colleagues, we used problems

that involve moving pieces, but different types. In addition, we used not only problems with multiple moves but also examples of problems where only a single move is necessary to reach a solution. Search-based theories have dealt primarily with problems that require more than one move and so involve an element of planning in addition to the searching of a problem space. We expected that using problems that did not require a sequence of moves would minimise the effect of having to plan moves.

For each problem we placed restrictions on the number of possible moves to be evaluated by the participant. Consider the 'T-puzzle' where there are four pieces to be arranged into the shape of a capital letter T (Suzuki & Hiraki, 1997). The puzzle is deceptively simple (see Appendix B). Each of the four pieces can be rotated around 360 degrees and their straight sides provide few clues about what goes where. People become fixated on filling the notch in the only piece that does not have a completely straight side (Suzuki, Miyazaki, & Hiraki, 1999; Suzuki et al., 2001). Placing restrictions on available moves can help participants. For example, given a cardboard cut-out of a T shape, and the puzzle pieces to physically manipulate, the T shape provides some indirect clues for positioning the pieces and may help to narrow down the potential possibilities for the different pieces. Suzuki et al. (1999) suggested that even a half-size template could improve performance.

In this first experiment of the series, we used four problems that require the physical manipulation of objects. Some of the problems require multiple moves to reach a solution (such as the T-puzzle and the necklace problem), and some require only one move (such as the horse-and-rider problem and the matchstick equation problems, see Appendix B). We gave one group of participants the standard version of the problems accompanied by the necessary physical elements of the problem, and for a

second group we devised means of limiting the moves to be considered. We predicted that the participants who had fewer possibilities to consider would solve the problems more often or faster than the control participants.

## Method

### *Participants*

The participants were 65 members of the psychology department's participant panel (members of the general population recruited from advertisements in national newspapers). They volunteered to take part in the experiment and were paid a nominal fee of 8 euro. There were 46 women and 19 men and their ages ranged from 18 years to 73 years, with a mean age of 48 years. The participants were assigned at random to one of two groups, the control 'no-hint' group (n = 35) or the experimental 'physical hint' group (n = 30).

### *Materials and design*

We used four insight 'move' problems, two multiple step problems (the T-puzzle and the necklace problem), and two single step problems (the horse-and-rider and the matchstick equation). One of the multiple step problems required participants to match patterns (T-puzzle) and the other required some arithmetic (necklace); likewise, the single step problems required pattern matching (horse-and-rider) and arithmetic (matchstick equation) (see Appendix B). We chose two different classes of problem to see whether the search restriction hypothesis would extend to a range of move problems.

The participants in the experimental condition were given additional materials. For the T-puzzle they were given an actual-size cardboard cut-out of a T shape. For the

horse-and-rider they were given the two panels pinned through the centre (so that they could only be rotated). For the necklace problem, they were given a fully-formed chain linked by three coloured links to encourage them to limit their efforts to opening and closing only three links and to emphasise that three links were all that were necessary to reach a solution. For the matchstick equation, they were given three coloured matches forming the two mathematical symbols and they were told that the crucial match was one of these coloured matches (see Appendix B).

Participants were given two problems each (because of anticipated time constraints) in a between-subjects design. They were given one multiple step and one single step problem each (at random either the two pattern problems, T-puzzle and horse-and-rider or the two arithmetic problems, necklace and matchstick equation), with the caveat that any participant who was not familiar with 'the roman numerals one would see on a clock' was not assigned to the arithmetic-type problems (2 people indicated they were not familiar with roman numerals). The problems were presented in a different random order to each participant.

***Procedure***

The participants were instructed that they could have up to 10 minutes to work on each of their two problems. They were told they could move on to the next problem once they had solved the previous problem. They were also told they could choose to leave a problem before they had reached a solution. The solution to the problem was given to the participant when the time limit was up, or when they indicated they wished to give up on it. If an initial attempt was incorrect, participants were allowed to keep working on it if they wished to do so. Feedback was not given on individual attempts but

participants were allowed to ask questions. We recorded whether or not a correct solution was reached by the participant and the time they spent on each problem (in seconds) using a stopwatch.

**Results and Discussion**

The 65 participants attempted two problems each (providing a set of 130 observations) and none of them had seen any of the problems before.

***Correct solutions***

As expected the participants who were given the physical hint solved more problems compared to the participants who were not given the hint (75% vs 46%, $\chi^2 = 11.56$, df = 1, $p < .001$, and all tests are one-tailed unless otherwise stated). Participants solved more problems when they were given a physical hint compared to no hint for the multiple-step problems (50% vs 6%, $\chi^2 = 16.404$, df = 1, $p < .0005$). As can be seen in Table 3.1, participants solved more problems when they had the physical hint for each of the multiple step problems, the T-puzzle (53% vs 5%, Fisher's exact test, $\chi^2 = 10.483$, df = 1, $p < .005$) and the necklace problem (47% vs 7%, Fisher's exact test, $\chi^2 = 6.136$, df = 1, $p < .05$).

Participants also solved more problems when they were given a physical hint compared to no hint for the single-step problems overall (100% vs 86%, $Z = 2.138$, $N_1 = 35$, $N_2 = 30$, $p < .05$; we tested the difference using Mann Whitney Z rather than chi-square because the distribution left one cell empty). However, the high solution rates to these problems in both conditions made it difficult to detect these differences in the correct solutions for each of the problems separately.

**Table 3.1:** Percentages of correct solutions for the no hint and hint versions of the single-step and multiple-step problems in Experiment 4

| | **No Hint** | **Physical Hint** |
|---|---|---|
| **Multiple-step** | | |
| T-puzzle | 5 | 53 |
| Necklace | 7 | 47 |
| *Total* | *6* | *50* |
| **Single-step** | | |
| Horse-and-rider | 95 | 100 |
| Equation | 73 | 100 |
| *Total* | *86* | *100* |
| **Total** | **46** | **75** |

Participants solved more problems when they had the physical hint than no hint for each of the single-step problems. The difference in solution rates was reliable for the matchstick equation (73% vs 100%, Fisher's exact test, $p = .05$) but not for the horse-and-rider problem (95% vs 100%, Fisher's exact test, $p = .571$). For the horse-and-rider problem, the improvement is evident instead in the length of time participants took to reach the correct solution, as we will now see.

***Latencies***

No comparison of the length of time to the correct solution was possible for the multiple step problems because there were so few correct solutions in the no hint condition. All latencies given are in seconds, log transformed to the base e (see Table 3.2). Pre-transformation latencies are also given in Table 3.2. For the single step problems, the mean time taken to a correct solution was faster when participants were given a hint compared to when they were not (3.31 sec vs 4.39 sec, $t = 3.694$, $df = 58$, $p$

< .0005, and all tests are one-tailed).

**Table 3.2:** Mean latencies to correct solution for each problem in Experiment 4 (in seconds log transformed to the base e) with standard deviations in parentheses and pre-transformation means in italics

| | **No Hint** | **Physical Hint** |
|---|---|---|
| **Multiple-step** | | |
| T-puzzle | 4.88* | 5.75 (0.62) |
| | *2m 22s* * | *6m 3s* |
| Necklace | 5.71* | 5.93 (0.40) |
| | *5m 2s** | *6m 40s* |
| **Single-step** | | |
| Horse-and-rider | 4.44† (0.73) | 2.81† (0.84) |
| | *2m 2s* | *33s* |
| Equation | 4.23† (1.61) | 3.57† (1.17) |
| | *2m 43s* | *1m 10s* |

Key: * Only one correct solution
† Latencies shown are with outliers removed

Participants solved the horse-and-rider problem faster when they were given the hint compared to no hint (3.06 sec vs 4.49 sec, t = 4.544, df = 32, p < .0005). We used a boxplot to check the pre-transformation latencies for outliers. This method indicated three outliers and when these were removed the transformed means were 2.81 and 4.44 for the hint and no hint conditions respectively (t = 5.670, df = 28, P < .0005, one-tailed). These participants had taken an anomalously long time to reach a solution compared to the other participants in the group, suggesting they may have been distracted at some point. Participants also solved the equation problem faster when they were given the hint compared to no hint but the difference was not reliable (3.57 sec vs 4.23 sec, t = 1.215, df = 24, p = .118). A boxplot also indicated outliers on this problem and when these three cases were removed the difference was reliable (3.09 sec vs 4.23 sec, t = 2.213, df = 21, p < .05).

***Single-step and multiple-step problems***

Participants solved more single-step problems than multiple-step problems (92% vs 26%, Wilcoxon's $z = 6.557$, $N = 43$, $p < .0005$), and we were able to confirm that this was so for both the pattern-type problems (97% vs 26%, Wilcoxon's $z = 5.000$, $N = 25$, $p < .0005$) and the arithmetic-type problems (87% vs 27%, Wilcoxon's $z = 4.243$, $N = 18$, $p < .0005$) as Table 3.1 shows. These results confirm that there is a valid distinction to be drawn between single and multiple step insight problems for this new set of problems. Even when the single-step problems are visual or spatial in nature, they are still easier than multiple-step problems. These data also show that participants generally reach a correct solution to a single-step problem in less time than for multiple-step problems. The relatively long solution time (ten minutes) that participants were allowed may account for the ceiling effects in the single-step problems. The additional time may allow participants to test their hypotheses about a solution exhaustively. Perhaps a two-minute time limit would have revealed a greater distinction between correct solution rates in the control and experimental groups on the single-step problems. However, two minutes would probably have been insufficient time for the multiple-step problems, and different time limits for the single and multiple step problems could have given participants a clue as to which were the more difficult problems.

The results show that people solve more problems when they are given a physical hint that limits the possibilities they must search compared to when they are given the standard problem. The experiment provides the following information: (1) Physical hints help people to solve the multiple-step T-puzzle and necklace problems.

Participants solved most of the easier single step horse-and-rider and matchstick problems whether they were given a hint or not, but nonetheless the hints helped people to solve these easier problems more quickly. (2) Single step problems were easier than multiple step problems, for this new set of move problems. The physical hints reduce the number of options that people need to consider. The experiment indicates that when people have fewer options to consider they solve the problem more readily. The result supports the suggestion that people can solve problems that require physical moves better when their search of the possibilities is constrained.

The results support search-based theories of insight problem-solving. Nonetheless, an alternative view is that limiting the possibilities to be considered may also steer participants towards a better representation of the problem. For example, the actual size template in the T-puzzle may help participants to better represent the final solution. Even so it is likely that participants still modify their search of the problem-space in keeping with the solution representation, so although the representation may be altered, the search will also be affected. Similarly, the limits on the matchstick equation may change participants' representation of the operator signs compared to the numerals, but we did not confine participants to moving a match to another operator sign. They could have moved one of the highlighted matches to anywhere of several locations within the equation. It is possible that the problem representation influences the kind of search that takes place, and that modifying the search options in advance can help participants to reach a better representation.

In the next experiment, we examine whether the results of Experiment 4 showing the effect of limiting possibilities can be extended from move problems to single-step, non-move problems.

## Experiment 5

Our aim in the second experiment of this series was to try to direct participants away from an incorrect possibility, in the expectation that they would then consider some of the alternative possibilities. This re-direction might help people to narrow down the possibilities they need to consider, and ensure that they search a better potential 'solution path', similar to limiting the available moves in problems where participants must choose from a selection of pieces and possible physical moves.

We used four single step problems, three from the previous experiments in Chapter 2 and one other (the marrying-man problem), well documented in the literature (e.g. Weisberg, 1995). We selected these problems because, unlike the problems used in the previous experiment, one cannot readily plan out all the possible moves or combination of moves. For these problems, a number of possibilities have to be generated before they can be evaluated. Restricting search space may seem sensible in problems with physical moves but a general account of insight should extend to all problems including single-step problems and problems not based on physical moves.

We measured participants' correct solutions and the latencies to reach them. We also measured the quality of the solutions, to examine whether the hints were successful in directing participants away from the most common erroneous solution (even if the new possibility that they considered did not provide the correct solution). An example of a poor-quality solution for the ping-pong ball problem is any answer that assumes the ball must travel horizontally, as we saw in the earlier experiments. An example of a good-quality solution (but one that is not the correct solution) is to roll the ball up a slope. The solution indicates the participant is thinking about the possibility of gravity and vertical (or semi-vertical) trajectory. Nonetheless, the problem specifies that the

ball must be *thrown*, and it must not bounce off any surface, and so the solution is not correct.

We compared solutions produced by a group of participants who were given no hint, and a group of participants who were given a hint after they read the problem. We also included a third group of participants who were given the hint before they read the problem. Participants may form an erroneous representation of the problem when they read it that they then find difficult to alter, and so we provided the hint before the problem for this group to examine whether there was any difference between a hint provided at a juncture that may pre-empt the erroneous representation and a hint provided after an erroneous representation has likely to have been formed. In the baseball-set fixation study by Wiley (1998) described above, she found that warning high knowledge participants not to use their baseball knowledge did not reduce the number of errors, suggesting that knowledge activation may be spontaneous and outside conscious control.

## Method

### *Participants*

60 participants were recruited from Trinity College, Dublin University through posters and they were paid a nominal fee of two euro for their participation. There were 35 women and 24 men. Their ages ranged from 15 to 47 years, with a mean age of 26 years. One participant was eliminated from the analysis because that person did not respond to all the problems. The participants were assigned at random to the three groups, hint-before (n = 20), hint-after (n = 20), and no hint (n = 19).

### *Materials and design*

We gave participants four single-step problems: the ping-pong, window-cleaner,

camping-trip and marrying-man problems (see Appendix B). There were three groups. The control group were given the standard wording of the problems. The two experimental groups were given a hint, one group read the hint before they read the problem, and the other group read the hint after they read the problem. The hints were negative hints, designed to tell participants what not to do (rather than positive hints). The hints were designed to direct participants away from the more common erroneous solutions, established in the first experiment for the first three problems (see Appendix A), and derived for the fourth problem from elsewhere (Weisberg, 1995). The hints are described in Appendix B.

***Procedure***

The problems were presented using the application Superlab 1.75 on a Macintosh desktop computer. Participants worked through one practice problem under the direction of the experimenter and then completed the four test problems in a different random order for each participant. First all participants were shown each problem on the first screen for 30 seconds. They were asked only to read the problem on its first presentation and not to try to solve it, and to press the space bar to move onto the next screen. For participants in the hint-after condition, the second screen contained the hint and they had up to 30 seconds to read it. The third screen showed the problem and the hint together and they were given up to three minutes to offer a solution. They were asked to press the spacebar as soon as they thought they knew the solution. For participants in the hint-before condition, the first screen contained the hint for 30 seconds, the second screen contained the problem for 30 seconds, and the third screen contained the problem and the hint together for 3 minutes. For participants in the

control group, the first screen contained the problem to read for 30 seconds, and the second screen contained the problem to solve for 3 minutes. Participants could offer their solution before the three minutes were up, and all participants were prompted at the end of the 3 minutes to tell their answer to the experimenter who recorded it. The experimenter recorded whether the solution was correct, and participants were shown the correct answer on the screen and pressed a button on the keyboard to record whether their solution was correct or not, and to indicate whether they had seen the problem before. The program recorded the time taken to reach a solution.

**Results and Discussion**

***Correct solutions***

A total score for each participant was calculated based on the proportion of unseen problems they had answered correctly. No participant had seen more than one problem before. There were 229 unseen problems in total. We carried out a Kruskal-Wallis analysis to test for differences between the three conditions. Participants solved somewhat more problems in the hint-after condition (65%) and the hint-before condition (67%), compared to the no-hint condition (53%), as can be seen in Table 3.3, but the difference was not reliable ($\chi^2 = 1.882$, df = 2, p = 0.390).

Separate analyses on each of the four problems showed that two of them were helped by the hints. Participants solved the window-cleaner problem more often in the hint-after (84%) and hint-before (90%) conditions compared to the no-hint condition (56%, $\chi^2 = 7.181$, df = 2, p. < .02, and all tests in this section are one-tailed unless otherwise stated). Participants solved the camping trip problem more often in the hint-before condition (80%) than in the hint-after condition (60%) or the no-hint condition

(56%). There was no difference between the three conditions overall ($\chi^2$ = 2.917, df = 2, p = .117), but more participants solved the problem in the hint-before condition compared to the no-hint condition (Mann -Whitney Z = 1.597, $N_1$ = 20, $N_2$ = 18, p = 0.055) and in the hint-before condition compared to the hint-after, although the difference is marginal (Mann -Whitney Z = 1.363, $N_1$ = 20, $N_2$ = 20, p = 0.086).

**Table 3.3:** Percentage of correct solutions for each problem in the three conditions in Experiment 5

| | **Hint After** | **Hint Before** | **No Hint** |
|---|---|---|---|
| Window-cleaner | 84 | 90 | 56 |
| Camping-trip | 60 | 80 | 56 |
| Ping-pong | 39 | 35 | 37 |
| Marrying-man | 74 | 63 | 63 |
| *Total* | *65* | *67* | *53* |

There were no differences in the hint-before, hint-after and no hint conditions for the ping-pong problem (39%, 35% and 37% respectively, $\chi^2$ = .062, df = 2, p = 0.475), and no differences between the three conditions for the marrying-man problem (74%, 63% and 63%, $\chi^2$ = .632, df = 2, p = .365), as Table 3.3 shows, and none of the pairwise comparisons were reliable for these two problems. The result shows that the negative hints helped people to solve some of the problems. The provision of a hint before or after helped people to solve the window cleaner problem, and the provision of a hint before (but not after) helped them to solve the camping trip problem. The hints did not help the ping-pong or marrying man problem.

***Good quality solutions***

We coded participants' answers, both correct and incorrect, as good quality or poor

quality. We included only unseen problems, and we did not include cases where no answer was offered. An example of a poor quality solution for the camping trip problem is any solution that assumes the 'thing' is a living creature such as a worm, as we saw in Experiment 1 in Chapter 2 (see Appendix A for common errors). An example of a good quality solution (but which is not the correct solution) is any solution that assumes the 'thing' is a non-animate object. For example, some participants identified the 'thing' as a pencil (which has two ends but is not commonly referred to as having a head and a tail). Inter-rater reliability on a sample of half the participants' responses was 100%.

**Table 3.4:** Percentage of 'good quality' solutions for each problem in the three conditions in Experiment 5

| | **Hint After** | **Hint Before** | **No Hint** |
|---|---|---|---|
| Window-cleaner | 89 | 94 | 61 |
| Camping-trip | 94 | 100 | 59 |
| Ping-pong | 56 | 89 | 58 |
| Marrying-man | 74 | 69 | 67 |
| *Total* | *78* | *88* | *61* |

As can be seen in Table 3.4 there were more good quality solutions in the hint-after and hint-before conditions (78% and 88%) than in the no-hint condition (61%), and this difference was reliable as shown in a Kruskal-Wallis test ($\chi^2 = 6.183$, df = 2, $p < .05$). Once again we carried out separate analyses for each of the four problems. For the window cleaner problem participants produced more good quality solutions when a hint was given, either before or after (94% and 89% respectively) compared to no-hint (61%, $\chi^2 = 7.609$, df = 2, $p < .02$, and all tests are one-tailed). The camping-trip problem showed the same pattern (100%, 94%, 59%, respectively, $\chi^2 = 13.502$, df = 2,

$p < .001$). For the ping-pong problem participants produced more good quality solutions when a hint was given before the problem (89%) compared to a hint after, or no-hint (56% and 58%, respectively, $\chi^2 = 5.649$, df = 2, $p < .05$). There was no reliable difference for the marrying man problem, (69%, 74%, and 67%, respectively ($\chi^2 = .236$, df = 2, $p = .445$). The result shows that the negative hints helped people to produce more good quality solutions for most of the problems. The provision of a hint before or after helped people to produce more good quality solutions for the window cleaner and camping trip problems. The provision of a hint before (but not after) helped people to produce more good quality solutions for the ping pong problem. The hints did not help people to produce good quality solutions for the marrying man problem.

***Latencies to correct solutions***

The time spent on the problem presentation screen, the hint presentation screen (for the experimental groups) and the solution screen was recorded by Superlab and log transformed to the base e prior to analysis. When presented with the solution screen participants were instructed to stop the timer when they thought they knew the solution, and they had up to a maximum of three minutes to spend on this screen. Only latencies for correct solutions were analysed.

We subtracted the time spent reading the hint (which was measured for each participant up to a maximum of 30 seconds) from the total time spent on each problem for participants in the hint-before and hint-after conditions in order to compare them to the no-hint condition. We carried out a one-way ANOVA on the mean correct solution time for each participant based on all the problems they solved correctly (see Table 3.5). There was no reliable difference between the three conditions, F (2,52) = .077,

MSE = .260, p = .926.

**Table 3.5:** Mean latencies to produce the correct solution measured in milliseconds (log transformed to the base e) with standard deviations in parentheses and pre-transformation means in italics for each problem in Experiment 5

| | **Hint After** | **Hint Before** | **No Hint** |
|---|---|---|---|
| Window-cleaner | 10.46 (.75) | 10.50 (.53) | 10.57 (.62) |
| | *45.94* | *41.27* | *47.13* |
| Camping-trip | 10.22 (.58) | 10.33 (.61) | 10.31 (.67) |
| | *32.89* | *35.33* | *37.22 (* |
| Ping-pong | 10.71 (.70) | 10.33 (.29) | 10.45 (.41) |
| | *55.09* | *25.59* | *41.26* |
| Marrying-man | 9.82 (.51) | 9.93 (.68) | 10.06 (.53) |
| | *20.59* | *25.37 (17.50)* | *27.52* |

We carried out separate analyses for each of the four problems but the differences between the three conditions were not reliable for any one of them: ping pong ball, F(2,18) = 2.711, MSE = .246, p. = .94; window-cleaner, F(2,41) = .93, MSE = .405, p = .911; camping-trip, F (2,35) = .111, MSE = .379, p = .896; marrying-man, F(2,35) = .579, MSE = .329, p = .565. See Table 3.5 for the pre-transformation latencies for correct solutions.

Boxplot graphs based on the pre-transformation latencies for each of the problems indicated the presence of outliers on three of the problems. We ran the analysis of the transformed latencies again with these outliers removed. There was no evidence of any difference between the conditions: window-cleaner, F(2,38) = .912, MSE = .265, p = .410; camping-trip, F (2,33) = .341, MSE = .319, p = .355; marrying-man, F(2,32) = .455, MSE = .248, p = .639.

The experiment provides the following information: (1) Negative hints help people to solve some single step problems. The provision of a hint before or after the

problem helped people to solve the window cleaner problem, and the provision of a hint before (but not after) helped them to solve the camping trip problem. In the design of the experiment, we included the hint-before as well as the hint-after group in case participants were unable to adjust their representation of the problem once it had been formed. In the case of the camping-trip problem, participants who read the problem first may have assumed the 'thing' was a living creature. Perhaps, when they were then informed that the 'thing' was not alive, they may have adjusted their representation to a dead creature rather than broadening it to include inanimate things as well. The hints did not improve correct solution rates for the ping-pong or marrying man problem. People did not solve any of the problems more quickly when they were given negative hints.

(2) Negative hints helped people to produce more good quality solutions for most of the problems. The provision of a hint before or after helped people to produce more good quality solutions for the window cleaner and camping trip problems. The provision of a hint before (but not after) helped people to produce more good quality solutions for the ping-pong problem. A common good quality solution for this problem was to roll the ball up the slope, which is very close to the correct solution except that the ball must be *thrown* rather than *rolled.* Participants who arrived at the rolling solution may have thought that it met the problem requirements and so they did not continue searching for a solution. This misunderstanding of the solution requirements could explain the discrepancy between the correct solutions and good quality solutions.

The hints did not help people to produce more good quality or correct solutions to the marrying man problem. The hint was that the man's culture does not permit multiple wives. It was intended to direct people away from the idea that it may be

acceptable to have multiple wives but in retrospect it may have drawn attention to the distinction between a cultural rule and a legal rule. Some participants offered a solution whereby the man's culture did not allow polygamy but it was allowed by law.

## General Discussion

The aim of this chapter was to investigate the role that search plays in insight problem-solving by examining the effect of limiting the search space on the solution rates of insight problems. We predicted that all problems would improve as a result of manipulations to limit the number of possible options that participants needed to consider when they are trying to solve a problem. The first experiment in this series, Experiment 4, examined problems that involve moving physical pieces. This experiment showed that correct solution rates or time to correct solution for all four move problems (T-puzzle, necklace, equation and horse-rider) can be improved by a restriction on the possible moves available.

In Experiment 5, we examined whether participants' performance on problems that did not involve physically moving pieces also benefits from restrictions placed on the search space. The improvement on these problems was not as comprehensive as the improvement seen with the move problems in the previous experiment. The correct solution rates for two of the problems improved with hints (window-cleaner and camping-trip) and participants had good quality solutions for both of these problems and the ping-pong ball problem when they received a hint. However, there was no improvement on the marrying-man problem.

While the results of the experiment indicate support for the search-based theories, they also suggest that search space alone is unlikely to completely account for

the difficulty in single-step problems without physical moves. The difference between the two problem sets of move versus non-move may be due to the fact that for the move problems there is a wide choice of all available moves, and it is theoretically possible to test all the possible sequences and combinations. For example, in a move problem like the T-puzzle one could feasibly try out, and make a note of, all the possible ways of combining the four pieces until eventually one hit upon the correct combination through trial and error. The initial problem in this case is to select the best move from a large array of possibilities. In contrast for many of the non-move problems, there are fewer possibilities and the initial array may not include the actual correct move. The initial problem here is to actually generate the possibility before it can be evaluated. For example in the ping-pong ball problem, if one does not conceive of the possibility that the ball could travel vertically as well as horizontally then this possibility cannot be considered.

Hence for the move problems in Experiment 4, narrowing the options to be considered increases the probability of discovering the right solution in the time available by trial and error. This explanation seems to rule out a role for insight. However, the relatively short time in which people reach a solution suggests that they may discover the solution without exhausting every possible combination, at least for the easier single-step problems. There is also the possibility that restricting the search may have the effect of making some important aspects more salient or diminishing the importance of other non-salient aspects. For example, in the matchstick equation a participant is encouraged to consider changing the mathematical signs whereas previously they may have first attended to the matches making up the numerals. These move problems may fall into the category of 'hybrid' insight problems described by

Weisberg (1995) where a problem may be either solved expediently by an insight into the key element of a problem or by an eventual process of trial and error. The ten-coin problem may also be a hybrid problem (Chronicle et al, 2004).

For the non-move problems used in Experiment 5, restricting the possibilities to be considered using negative hints may help participants to avoid spending time on incorrect potential solutions. However, there is still a need to generate the correct possibility. For example, on the ping-pong ball problem just because a participant has eliminated certain options such as the wind or a back-spin technique does not necessarily mean that he or she can rule out the option of throwing the ball up instead of forward. Dominowski & Dallob (1995) identify problems where one seems to be lead into a wrong answer, as in "There were thirty sheep in a field and all but nine escaped. How many are left?" (The answer is nine, but it is tempting to say twenty-one). They distinguish such problems from those problems that seem impossible and initially it is difficult to think of any possibilities, such as how the window-cleaner falls without injuring himself. For this latter type of problem it may be that highlighting the important elements (or downgrading the seemingly important elements) by positively targeting the representation is more helpful. For example, Experiment 2 showed that changing the ping-pong ball to a basketball helped participants to ignore the specific characteristics of the former when reasoning about the problem. Nonetheless directing attention away from unhelpful solution paths did at least result in better quality solutions for most of these problems. Hence searching inappropriate problem spaces may be a factor in all insight problems but it plays a bigger role in move problems or 'hybrid' problems.

In summary, restrictions on the possible moves to be considered positively

affected all the problems with physical moves in Experiment 4. This manipulation could have worked because all of these problems are open to a trial-and-error solution and participants might normally attempt to reduce the search space through the use of heuristics, such as trying to fill the notch in the T-puzzle. When the search is restricted, participants are given a readymade shortcut, and one that is accurate.

For the non-move problems in Experiment 5, restrictions on search space led to an increased correct solution rate on the window-cleaner and camping-trip problems; the ping-pong ball problem showed an increase in good quality solutions and the marrying-man problem showed no improvement on either measure. These results suggest that there might be some benefit from search restrictions in assisting participants to avoid exploring incorrect possibilities, but search restrictions may not help them to generate alternative possibilities.

The five experiments reported thus far in Chapters 2 and 3 have highlighted that it is possible to improve correct performance on insight problems through manipulations aimed at improving mental representations and manipulations aimed at restricting search space. However it is difficult to find a single type of manipulation that affects all problems equally, which indicates both mental representation and search space may play a role in understanding insight problem difficulty. Similarly a number of processes may combine to facilitate reaching a solution in insight problems, some of which may relate to the content of individual problems and others to the way participants approach insight problems. In the next chapter, we turn our focus to the skill or combination of skills that is associated with successful insight problem-solving.

# CHAPTER 4

What skills does a person need to be good at insight problem-solving? The previous two chapters examined ways of improving participants' performance on a range of insight problems. In this chapter we compared the ability to solve insight problems with abilities in other areas in two experiments. Our aim was to advance our understanding of how people correctly solve insight problems by investigating what skills might be required to be a good insight problem-solver.

Researchers sometimes use the individual differences approach to better understand what skills are needed for different kinds of reasoning. The approach assumes that people will perform at different levels and receive different scores on the measures used in the study. For example, Newstead, Handley, Harley, Wright and Farrelly (2004) found that participants differed in their ability to reach correct conclusions on a deductive reasoning task. People who got higher scores on the reasoning task also tended to get high scores on a measure of intellectual ability and on a measure of the ability to generate alternative representations. Miyake, Friedman, Emerson, Witzki and Howerter (2000) employed an individual differences study to distinguish between three proposed executive functions: shifting between mental tasks or sets, updating and monitoring of working memory contents, and inhibition of prepotent responses (i.e. responses that tend to be made more often than other possible responses). Hypothesising that such functions are important for reasoning, they sought to link measures of performance on these functions to performance on three tasks commonly used in the assessment of normal and clinical populations. They found that people who did better on the Wisconsin Card Sorting Task also did well on the measure of the ability to shift mental sets. Likewise performance on the Tower of Hanoi task

was linked to the updating of working memory, and Random Number Generation to both inhibition and working memory updating. However, in this study we are primarily interested in how people reason with insight problems.

***Individual differences and insight***

Schooler and Melcher (1995) compared individuals' performance on insight problems with their performance on three other measures of elements hypothesised to be instrumental in insight problem-solving. The proposed elements were perceptual restructuring, overcoming a context-induced set and memory retrieval. The first element, perceptual restructuring, was measured through participant's ability to recognise out-of-focus pictures and had the strongest association with correct solutions to insight problems, with better perceptual ability associated with better insight score. For the second element of context-induced set the authors used two measures of field dependence/independence. This factor relates to how participants' perception of a display is influenced by cues in the environment with less influence indicating greater field independence. One of these measures, the Embedded Figures task, indicated that people with greater field independence solve more insight problems. This task requires people to detect a given shape in a more complex array made up of lots of small shapes. Similarly, Antonietti and Gioletta (1995) found that field-independent participants were more likely to provide an analogical solution to Duncker's radiation problem than were field-dependent participants.

Better performance on the Remote Associates Test, where participants must find a fourth word that links to three others (e.g. common/level/play = 'ground') was also associated with a better insight score. Schooler and Melcher (1995) interpreted

this measure as one of memory-retrieval. There was no relationship between insight and several other memory-retrieval measures.

Other correlates of correct insight problem solutions in this study were the vocabulary score on the Student Aptitude Test and correct solutions to analytic problems. However, the pattern of correlates with the measures of perceptual restructuring, overcoming sets and memory retrieval was quite different for the insight and analytic problems suggesting that while they share some processes in common, there are also differences. Gilhooly and Murphy (2004) also found that there was a different pattern for the correlates of insight and non-insight problems, even though there were correlated with each other. Better scores on insight problems were associated with better scores on measures of figural fluency (tests where participants must copy or draw figures according to instructions), vocabulary and alternative uses. Working memory was related to both insight and non-insight performance.

Ansburg (2000) found evidence of a positive relationship between scores on insight problems and the Remote Associates Test, which in this case was interpreted as reflecting fluency of thought. People who correctly solved insight problems were also those who showed an ability to apprehend relations, measured using a verbal analogies task (e.g. month is to year as hour is to day) and a series-completion task (e.g. two w four r one o).

Overall being able to recall a comprehensive set of knowledge relating to an object or word is important for insight problem-solving. The Remote Associates Test, which requires the recall of a wide set of associates for each word has been found to be positively related to insight problem score (Schooler & Melcher, 1995; Ansburg, 2000). The ability to generate a broad set of alternative uses for objects has also been found to

be associated with insight performance (Schooler & Melcher, 1995, Gilhooly & Murphy, 2004). It is possible that working memory is needed to maintain and process these large sets of information and so it will be one of the factors that we will examine in relation to performance on insight problem performance. Working memory has been found to be associated with performance on the Tower of Hanoi and Random Number Generation tasks (Miyake et al, 2000) and with both insight and non-insight problem scores (Gilhooly and Murphy, 2004). However, there are relatively few studies that have concentrated on skills that are necessary to be good at insight problem-solving.

We hypothesised that three factors are necessary for insight problem-solving. If mental representation plays a key role in insight problem-solving then we could expect good insight problem-solvers to be competent in forming mental images. Hence we predicted that people who are able to solve insight problems would also have better imagery ability. However, if managing the search for a solution is important in insight problem-solving then we could expect executive functions like attention and working memory to be related to insight. Over two experiments, we examine the three factors of working memory, attention and imagery and we give a brief outline of each here.

***Working memory and problem-solving***

Working memory can be thought of as a limited capacity store for information while it is being processed (Reber & Reber, 2001). Gilhooly and Murphy (2004) found that working memory capacity was related to both insight and non-insight problems. Markovits, Doyon and Simoneau (2002) found that ability to reason with conditional reasoning problems was correlated with better scores on measures of verbal working memory. Capon, Handley and Dennis (2003) found a similar relationship between

working memory measures and performance on tasks of syllogistic reasoning and spatial reasoning. Süß, Oberauer, Wittmann, Wilhelm and Schulze (2002) found that there was a strong positive relationship between reasoning ability (including verbal and figural tasks) and performance on a range of working memory measures. Working memory as measured by a sentence span task was found by Garcia-Madruga, Gutierrez, Carriedo, Luzon and Vila (2004) to be related to reasoning on tasks involving 'if . . . .then" statements, although measures that combined working memory and reasoning were more strongly related. Borella, Jouffray and de Ribaupierre (2003) found that working memory, as measured by a sentence span test, accounted for a sizeable part of the variance in a test of fluid intelligence (Raven's Matrices) with higher span associated with greater intelligence.

On the role of working memory in problem-solving, Passolunghi, Cornoldi and De Liberto (1999) found that a listening span test that required both storing and processing information in working memory discriminated between good and poor problem-solvers among primary school children. In the listening span test, poor problem-solvers made more intrusion errors, recalling words that had been presented but were not the end-of-sentence targets, suggesting that this group had trouble suppressing irrelevant information.

In a review, Conway, Kane and Engle (2003) conclude that working memory tasks that require processing as well as storage (e.g. sentence span) load on a factor separate from that of simple storage tasks (e.g. digit span). Furthermore, they assert, "what seems to be important about working memory span tasks is that they require the active maintenance of information in the face of concurrent processing and interference and therefore recruit an executive attention-control mechanism to combat interference"

(p. 551). Problem-solving requires both the recall and control of information, and in the case of insight problem-solving a potentially large set of information, which may explain why working memory measures have been associated with success on insight problems. Nonetheless Lavric et al (2000) found that a concurrent working memory task (counting tones) disrupted performance more on an analytic problem than it did on an insight problem.

Working memory has previously been related to several different reasoning tasks but there is a distinction between tests of storage-only and storage-plus-processing. The results regarding working memory and insight problems are sometimes conflicting.

***Attention and problem-solving***

Attention can be thought of as a mechanism by which people select information for further cognitive processing. Eysenck and Keane (2000) use the metaphor of changing channels to select what appears on the television screen for the changing of focus from one source of information to another. As with the capacity of the television, attention is limited and selective (Willingham, 2004).

The role of attention in problem-solving has not been investigated to the same extent as working memory, but it has not been completely ignored. The existing research has indicated that directing a participant's attention to a relevant component of the problem makes a solution to that problem more likely. Glucksberg and Weisberg (1966) using the candle-box problem found that drawing attention to the tack-box through labelling encouraged participants to use the box in the solution. This trend continued even when the problem requirements were altered so that preventing the

candle dripping wax on the floor was unnecessary, making more alternative solutions available. When attention was drawn to the box through labelling, the box solution was still the most popular.

In a more direct way, Grant and Spivey (2003) perceptually highlighted the critical component in a diagram-based problem. The diagram was based on Duncker's (1945) radiation problem where lasers are used to kill a tumour. An initial eye-tracking experiment where participants attempted to solve the problem using the diagram found that participants' gaze focussed on the part of the diagram representing the skin prior to reaching a correct solution. A subsequent experiment where the skin-part of the diagram pulsated facilitated a new group of participants in reaching a correct solution. The results of the two experiments suggested that helpfully directing participants' attention to a key component of the problem, as identified by the eye-tracking study, facilitated the solution of that problem. This finding could also suggest that the independent identification of a problem component worthy of further attention is an important phase in the problem-solving process.

Eye-tracking equipment to determine the focus of attention has also been used in relation to matchstick-equation type problems by Knoblich et al (2001). These authors found that successful participants spent more time looking at the constrained, but crucial, components than did unsuccessful solvers. Diverting attention away from an unhelpful focus may help to relax constraints and promote solution to a problem. Segal (2004) proposed that a break in activity focused on a problem on which a person has reached an impasse (otherwise known as incubation) might allow the individual to relax their incorrect assumptions about a problem. Diverting the attention of participants who had reached an impasse on a geometrical problem requiring insight

was beneficial regardless of whether the intervening activity was taxing or not, or of a shorter or longer duration.

It may also be that self-distraction that results in a broader attentional sweep is beneficial to problem-solving processes. Ansburg and Hill (2003) examined attentional resources and creative problem-solving and found that higher creativity scores (as measured by the Remote Associates Test) were reliably predicted by a measure of greater sensitivity to peripheral cues. This latter involved anagram solutions being presented on a to-be-memorised list or on a to-be-ignored audio-tape. The measure used was how many of the peripherally presented anagrams were later solved. Their conclusion was that participants who divided their attention so as to process information other than that which they were supposed to be focusing on were better able to use that information when it became relevant. Choudhury and Gorman (2000) examined the relationship between sustained attention and a problem-solving task for children aged 17-24 months. Perhaps surprisingly, the children with more frequent off-task glances had longer attention spans, performed better at the problem-solving task and had higher scores on an infant mental development scale.

The findings on attention and problem-solving so far indicate that directing participants' attention to a crucial element that might otherwise be neglected can help participants to reach a solution. However, in the absence of a specific redirection, the indications are that diverting rather than focusing attention may help to break a fixation or to allow participants to pick up information that they can use later.

***Imagery and problem-solving***

Imagery refers to the process of mentally recreating sights, sounds, smells, tastes, touch

and other bodily sensations (Robertson, 2002) although in common use the term imagery is often used to refer to a visual image. Imagery differs in some important ways from an actual image: there is not an actual projected image but rather a cognitive process that is an analogue of an actual scene; there may be construction of an image rather than a reproduction; and it may be possible to mentally adjust the image (Reber & Reber, 2001).

There has been some debate about how similar a mental visual image is to an actual visual image. For example, instructions to mentally draw what turns out to be a visual illusion does not always produce the illusion in the same way that it occurs when it is physically drawn out and viewed (Pylyshyn, 2003). However, several studies have supported the idea that a mental image can be examined and manipulated in a similar way to actually perceived objects. For example, Finke, Pinker and Farah (1989) reported that participants were able to rearrange familiar patterns and to describe the results such as rotating the letter D and attaching the letter J, then naming the resulting image as an umbrella. Shepard and Metzler (1971) reported that participants took longer to mentally rotate an abstract figure when the required degree of rotation was greater, suggesting that participants were simulating what would happen if they were actually rotating the figure through the prescribed rotation. Mental rotation-type tasks have since become a common way of measuring visual imagery ability.

Individual differences have been reported in relation to imagery ability (Kosslyn, Brunn, Cave & Wallach, 1984). However, objective measures of imagery ability do not necessarily correlate with subjective reports of imagery ability (McAvinue & Robertson, 2003). Schooler and Melcher (1995) found no significant correlation between correct solutions to insight problems and a mental rotations task.

However they suggested that this absence may have arisen due to their failure to distinguish between problems with a visual or spatial component and other problems. Evidence of a role for imagery in solving insight problems was found by Adeymo (1994) who compared performance between high and low imagery ability participants on what was termed 'practical construction problems'. These problems consisted of four functional fixedness insight problems including Duncker's (1945) candle-box problem and Maier's (1931) two-string problem. Imagery ability was measured using the Differential Aptitudes Test (DAT) and a questionnaire by Barrat (1953). High imagers scored significantly better on the insight problems compared to low imagers. Furthermore, high imagers benefited more from instructions to imagine the materials and a solution prior to solving the problem than did the low imagers. Instructions to reorganise and restructure the problem helped neither group. Imagery ability may be related to insight problems with a visual or spatial element, if not to all insight problems. We will examine the role of imagery in insight problem-solving in the first experiment in this chapter. The second experiment will examine working memory and attention. In both experiments we will examine not just total insight score but verbal and non-verbal problems separately. This distinction has not previously been made in other individual differences studies of insight problems.

## Experiment 6

The first experiment in this series examines the role of imagery in insight problem-solving. An explicit instruction to draw a diagram prior to solution in Experiment 1, reported in Chapter 2, did not lead to more solutions. However, in Experiment 3, also reported in Chapter 2, we observed that externalising a representation by the use of

physical props was helpful to participants when working with insight problems. We suggested that being able to construct an accurate mental representation of a problem situation, and manipulate the elements within it, might be important in reaching a solution. Thus we hypothesised that imagery ability would be related to the capacity to construct accurate representations and increased solutions on insight problems.

We chose a mental rotation task to achieve an objective measure of visual imagery ability. Shepard and Metzler's (1971) results suggested that participants could mentally simulate the rotation of an abstract figure to the required degree in order to make a judgement on the task. The particular test used in our experiment was a version of the Vandenburg-Kuse figures (1978) redrawn with a CAD package by Peters (1995). This test has been widely used in the literature (e.g. Astur, Tropp, Sava, Constable & Markus, 2004; Malinowski, 2001).

In addition to the mental rotations task we used a self-report measure of imagery ability. We chose the Questionnaire upon Mental Imagery (Sheehan, 1967) as it includes not just visual imagery but imagery ability related to the other senses as well. It is a shorter version of an older questionnaire (Betts, 1909). This measure has also been well used in the literature (e.g. Kwekkeboom, Huseby-Moore & Ward, 1998; Hatakeyama, 1997; Okada, Matsuoka & Hatakeyama, 2000).

We also used two further measures. An adapted form of the Mednick's (1962) Remote Associates Test was selected as a measure of verbal creativity to contrast with the imagery measures. A relationship has been established between this measure and insight (Schooler & Melcher, 1995; Ansburg, 2000). However, our aim was to compare this measure's correlation with verbal and non-verbal problems. The version used here was one compiled by Bowden and Jung-Beeman (2003).

The last measure used was the Cognitive Failures Questionnaire devised by Broadbent, Cooper, Fitzgerald and Parkes (1982) to measure what could be thought of as absent-mindedness (e.g. Garavan, Ross, Murphy, Roche & Stein, 2002). It asks participants to rate how often they have, for example, gone to a particular room and then forgotten what they wanted there. The questionnaire has been widely used, often with clinical populations but also with normal populations (e.g. Wallace, Vodanovich & Restino, 2003; Boomsma, 1998). We hypothesised that scores on this measure would reflect a tendency to be distracted by other information, or to think of more than one thing at a time, which could be helpful when thinking about insight problems. The ill-defined insight problems are characterised by the need to consider a wider range of alternatives and Schooler, Ohlsson and Brooks (1993) found that instructions to try alternative strategies impeded solutions to non-insight problems but did not adversely affect solutions to insight problems. Ansburg and Hill (2002) found that higher creativity scores (as measured by the Remote Associates Test) were reliably predicted by a measure of sensitivity to peripheral cues. There are, of course, numerous anecdotal reports of the absent-mindedness of insightful individuals such as Albert Einstein (Blackwell, 2004), Adam Smith (Landry, 2002) and Andre-Marie Ampere (Asimov, 1992).

In summary, the main aim of Experiment 6 was to examine the relationship between imagery ability and insight problem-solving. A second aim was to examine the relationship between other factors such as verbal creativity and absent-mindedness, and insight problem-solving.

## Method

### *Participants*

Participants were recruited through posters on the campus of Trinity College, Dublin. The thirty participants received either 10 euro or course credits for their participation. There were 6 men and 24 women aged between 17 and 38 years, with a mean age of 20.8 years.

### *Materials*

Participants completed five tasks. The first measure was of insight problem-solving ability. Participants attempted to solve eight insight problems with two minutes allowed per problem. Our earlier experiments (e.g. Experiment 5) and those of Lockhart et al (1988) suggest that this is sufficient time for single-step problems (but not multiple-step problems). Multiple-step problems such as the nine-dot problem tend to display solution rates close to zero and so are unsuitable for gathering data for a correlational study. We used eight single-step problems. We used four single-step problems that were non-verbal in that they required a move of some kind or had visual-spatial requirements. The non-verbal problems were the horse-and-rider, matchstick-equation, ping-pong and inverted-pyramid problems. We used four problems that were verbal in that their difficulty arises from interpretations of the wording. The four verbal problems were the window-cleaner, marrying-man, camping-trip and woods-walk problems. The full wording and solutions for all eight problems is contained in Appendix C. The problems were presented in booklet form with instructions including a sample problem and a request not to skip ahead or return to previously presented problems (see Appendix C for full instructions). The order of the problems was randomised in each

booklet and participants were given an opportunity to indicate if they had seen any of the problems before or if they were unfamiliar with the roman numerals used in the matchstick-equation problem.

*Imagery Measures*

Participants completed two measures of imagery ability. The Mental Rotations Test (MRT), revised by Michael Peters (1995) from the original by Vandenburg and Kuse (1978), requires participants to mentally rotate target abstract block figures. In a set there is one target figure and four views of similar figures rotated about the vertical axis. Sample figures are given in Appendix C. The task is to identify which two of the four views show the actual target figure and which two do not. There are 24 sets, divided into two blocks of 12 with three minutes allowed for each block. There is a short break (30 seconds) between the two blocks. Participants are given three practice problems and their solutions prior to commencing the test. One point is given for each correct pair identified, giving a maximum score of 24.

The second imagery measure was the Questionnaire upon Mental Imagery (QMI), revised by Sheehan (1967). It is a subjective test and is included in Appendix C. Participants close their eyes and imagine each item in the questionnaire, for example the smell of an orange. They then rate the vividness of the image on a 7-point scale where 1 is a perfectly clear image and 7 is no image at all. The 35 items cover seven subscales: cutaneous, visual, olfactory, organic, auditory, glutaneous, kinetic. A low score indicates someone who self-reports his or her imaging ability to be very good with the lowest possible score, 35, reflecting the highest possible report of imaging ability. The maximum score, which indicates a self-report of complete inability to

generate an image is 245.

*Remote Associates Test*

We gave participants a test of verbal ability in the form of an adapted Remote Associates Test (RAT). Originally developed by Mednick (1962), in this task participants are given a set of three words and asked to find the word that is an associate of each. The materials for the RAT were taken from a list suggested by Bowden and Jung-Beeman (2003). In this list, all the items form new compound words or phrases with the solution word as in the example loser/throat/spot where the solution is 'sore'. We chose 24 items from Bowden and Jung-Beeman's tables for the normal percentage of solutions reached in 15 seconds. We took eight items each from the easier (51 - 76% solution rate), medium (26 – 50%) and more difficult items (1 – 25%). Participants had six minutes (24 items x 15 seconds) to complete as many items as they could. After a pilot study, we decided to arrange the items in order of difficulty, starting with the easiest, to avoid the possibility of artificially low scores resulting from people spending too long on a difficult set early in the booklet. Participants were given one practice set and instructed not to spend too long on any one set. The maximum possible score was 24.

*Cognitive Failures Questionnaire*

Participants completed the Cognitive Failures Questionnaire (CFQ) devised by Broadbent et al (1982). This questionnaire asks participants to rate how often they have experienced each of 25 items in the last six months. The items are everyday cognitive failures such as failing to remember appointments or missing signposts. The

items can be thought of as measuring 'absent-mindedness'. There is a five-point scale ranging from 0 for 'never' to 4 for 'very often'. The maximum score is 100 with higher scores representing greater absent-mindedness.

***Procedure and design***

Participants were tested individually or in small groups. There was one session lasting approximately one hour. All participants completed all task. They started with the insight problems and then completed the other tasks in the following order: Mental Rotations Test (MRT), Cognitive Failures Questionnaire (CFQ), Remote Associates Test (RAT), and Questionnaire upon Mental Imagery (QMI). Where required, the experimenter timed tasks using a stopwatch.

**Results**

A total score for insight problems was calculated based on the proportion of unfamiliar problems solved giving a maximum score of 1.00. There were 220 unseen problems from a possible 240. Overall participants solved .47 of the problems. We also derived proportional scores for the verbal and non-verbal problems on the same basis. Participants solved .61 of the verbal problems and .34 of the non-verbal problems. The mean scores for performance on all the other tasks is summarised in Table 4.1.

**Table 4.1:** Mean scores, standard deviations and possible range of scores for each task in Experiment 6

| Task | Mean | SD | Possible Range |
|---|---|---|---|
| Mental Rotations Task | 8.07 | 4.14 | 0 - 24 |
| Q. upon Mental Imagery | 104.80 | 23.33 | 35 - 245 |
| Remote Associates Test | 11.00 | 4.46 | 0 - 24 |
| Cognitive Failures Q. | 40.77 | 11.08 | 0 – 100 |

***Correlation of insight problem score with scores on other tasks***

All correlations used are Pearson's r with n = 30 and are one-tailed unless otherwise stated. In relation to the Mental Rotations Task (MRT), we had predicted that higher performance on this task would be associated with higher performance on the overall insight score. We expected that an ability to accurately imagine the situation described in the problem and to mentally manipulate the image would be reflected in higher scores on the insight problems. However, this prediction was not confirmed by the data as the correlation between insight score and MRT score was quite low ($r = .110$, $p = .282$).

On inspection of a scatterplot of the data we noted that two participants appeared in an extreme position on the graph (see Figure 4.1). These participants had very low insight scores but very high MRT scores, a pattern that was anomalous with the other participants in the study. As both of these participants were visiting American students, we speculated that they may have had prior experience on similar tasks and so we decided to re-run the analysis with these participants removed. Subsequent to their elimination, the value of the correlation between insight and MRT scores improved to reach significance level ($r = .385$, $n = 28$, $p < .05$). The mean score on this task was only 8.07, which compared to the maximum possible score of 24 suggests that participants found this task quite difficult.

On the self-report measure of imaging ability, the QMI, participants reported a mean score of 104.80 indicating that most participants reported themselves as having moderately clear imagery. However, the self-report measure was not correlated with overall insight score either ($r = .111$, $p = .279$).

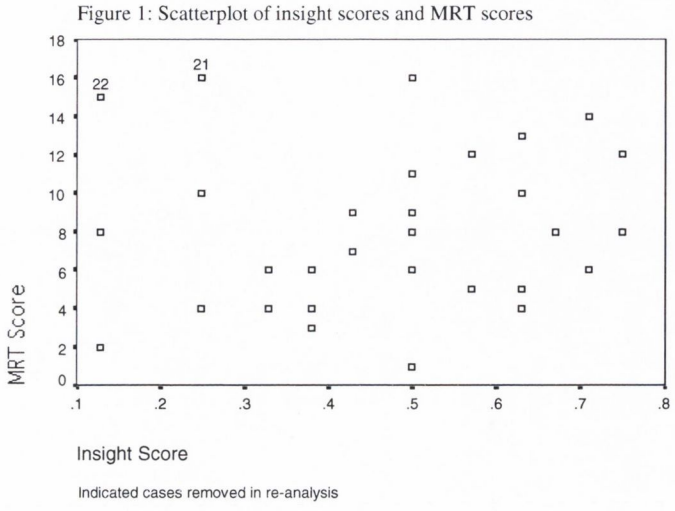



**Figure 4.1:** Plot of insight scores and MRT scores with the position of the two outlying cases indicated by their case numbers

On average, participants correctly solved 11 of the 24 RAT items correctly. Better performance on the RAT was reliably associated with better performance on the insight problems as expected ($r = .327$, $p < .05$).

**Table 4.2:** Correlations between correct scores on insight problems, overall and for the two sub-categories, and the other tasks in Experiment 6

| **Task** | **All Insight** *(8 probs)* | **Verbal** *(4 probs)* | **Non-verbal** *(4 probs)* |
|---|---|---|---|
| Mental Rotations Task | .110 | .025 | .128 |
| Mental Rotations Task† | .385* | .295 | .205 |
| Q. upon Mental Imagery | .111 | .346 | -.256 |
| Remote Associates Test | .327* | .508*** | -.108 |
| Cognitive Failures Q. | .446** | .256 | .263 |

Key: † with outliers removed
*p < .05, one-tailed
**p <. 01, one-tailed
***p <. 01, two-tailed

The mean score on the Cognitive Failures Questionnaire was 40.77 out of a maximum score of 100, suggesting that most participants reported that they experienced absent-mindedness occasionally. More absent-mindedness was reliably associated with better performance on the insight problems as expected ($r = .446$, $p < .01$). Table 4.2 summarises the correlations found between insight problem scores and scores on the other tasks.

***Regression Analysis of Insight Problem Score***

We also analysed insight problem score as the dependent variable in a regression analysis with the two imagery measures, and the RAT and CFQ measures. From a stepwise regression, the most efficient model that emerged accounted for 17% of the variance in insight problem score. The only predictor variable in the model was the CFQ score with higher scores on this measure associated with higher scores on the insight problems. The adjusted R square was .170 and the model was significant ($F[1,28] = 6.943$, $p < .05$).

***Correlation of scores on verbal and non-verbal subcategories with scores on other tasks***

All correlations used are Pearson's r with $n = 30$ and are one-tailed unless otherwise. We examined whether the verbal and non-verbal problems had a similar pattern of association to the other measures. As can be seen from Table 4.2, the pattern was different for the two subcategories on some of the measures.

The MRT, the objective imagery measure, had a low and non-reliable correlation, in the same direction, with the non-verbal problems ($r = .128$, $p > .250$, two-tailed) and the verbal problems ($r = .025$, $p > .05$, two-tailed). When we re-ran the

analysis with the two outlying cases removed as described above, the strength of the correlations with both categories increased but did not reach significant levels for either (non-verbal: $r = .205$, $n = 28$, $p > .05$, two-tailed; verbal: $r = .295$, $n = 28$, $p > .05$, two-tailed).

The QMI, the subjective imagery measure, showed different patterns for the two subcategories. The correlation with the verbal problems was positive but did not quite reach significant levels ($r = .346$, $p = .06$, two-tailed). Recall that a lower score reflects clearer imagery ability in the QMI and so this relationship indicates that participants who solved more of the verbal problems correct had less clear imagery. The correlation with the non-verbal problems was negative, in the opposite direction, and indicates that participants who solved more of the non-verbal problems reported clearer imagery. However, it did not reach the level of significance ($r = -.256$, $p > .05$, two-tailed).

The RAT revealed a similar disassociation between the subcategories. The correlation with the verbal problems indicated that answering more RAT items correctly was associated with correctly solving more verbal problems ($r = .508$, $p < .01$, two-tailed). Answering more items correctly on the RAT was associated with solving fewer of the non-verbal problems, although the correlation did not reach significance level ($r = -.256$, $p > .05$, two-tailed).

The CFQ was positively associated with both the verbal and non-verbal problems but neither of the correlations was reliable. The correlation with the verbal problems was $r = .256$ ($p > .05$, two-tailed) and the correlation with the non-verbal problems was $r = .263$ ($p > .05$, two-tailed).

Surprisingly there was no significant correlation between the verbal and non-

verbal problems (r = -.048, p = .401). This difference may have arisen because participants found the non-verbal problems to be harder than the verbal ones (.34 v .61) as confirmed by a paired t-test (t =3.727, df = 29, p < .01, two-tailed). Also, in this particular sample there were no high performing participants, with the highest overall score recorded being 6 out of 8 problems correct (75%).

***Relationship between measures***

We made no specific predictions about the relationships between the other measures and no significant correlations were found. Individual correlations are given in Table 4.3. The direction of the association between the QMI and MRT was of better self-reported imagery and better spatial ability but the relationship was not significant (r = -.223, p = .236, two-tailed). The relationship between the QMI and the RAT was in the opposite direction indicating a disassociation between self-reported imagery ability and verbal ability, but again the correlation was not significant (r = .266, p = .156, two-tailed). The CFQ's strongest relationship was with the RAT with greater reported absent-mindedness associated with more correct items on the latter, but this relationship was not reliable (r = .222, p = .238, two-tailed).

**Table 4.3:** Correlation matrix of non-insight measures in Experiment 6

| | **QMI** | **RAT** | **CFQ** |
|---|---|---|---|
| Mental Rotations Task | -.223 | .158 | .040 |
| Cognitive Failures Q. | .112 | .222 | |
| Remote Associates Test | .266 | | |

**Discussion**

Our main aim in this experiment was to assess the role of imagery in insight problem-solving. The experiment provided three main pieces of information: (1) Performance on the Mental Rotations Test (MRT), which measured visual imagery ability was moderately associated with insight problem score when two outlying cases were removed. The two outlying cases had high MRT scores but very low insight scores. There were no instances where any participant had scored very well on the insight problems and very poorly on the MRT. This pattern of results could indicate that some imagery ability is necessary to do well on a range of insight problems, but it is not sufficient, that is someone who is a capable imager may not be a good insight problem-solver.

(2) The Remote Associates Test and the Questionnaire upon Mental Imagery had different patterns of association for the subcategories of verbal and non-verbal problems. Better performance on the verbal problems was associated with higher scores on the RAT, which involves identifying a word which is an associate of three others. In contrast, participants who reported clearer imagery solved more non-verbal problems. These results suggest a distinction between these two sorts of insight problem and have implications for the selection of problems in insight studies. For example, RAT type problems have been used as examples of insight problems in some studies (e.g. Jung-Beeman et al, 2004) but have not necessarily been distinguished as *verbal* insight problems.

(3) Greater absent-mindedness as measured by the Cognitive Failures Questionnaire was associated with higher scores on the insight problems. Scores on the CFQ was the best predictor of insight score and accounted for 17% of the variance. We

inferred from the results from the CFQ that the role of attention in problem-solving merited further investigation. The CFQ has been linked to attention (e.g. Robertson, Manly, Andrade, Baddeley & Yiend, 1997; Vom Hofe, Mainemarre & Vanneir, 1998) and in the next experiment we investigate the role of attention, and also of another executive function, working memory.

## Experiment 7

Our aim in this experiment was to examine the role that the executive functions of attention and working memory play in insight problem-solving. Planning a number of moves in advance may be important to successfully solve insight problems such as the nine-dot problem (Chronicle et al, 2001). Working memory capacity may be needed to keep alternative possibilities in mind and attention may play a role in deciding what elements of a problem to focus on or directing the search for relevant information internally and externally. As we saw earlier, some studies have suggested that directing attention to a particular element of a problem can improve performance (e.g. Glucksberg & Weisberg, 1966) and people who pay more attention to peripherally presented information make better use of that information in a subsequent task (Ansburg and Hill, 2002). There are different aspects of attention that are distinguishable parts of an overall attention network as indicated by findings from neuropsychology. For example, an area known as the cingulate gyrus has been implicated as the biological base for identifying and reacting to errors, whereas an area in the parietal lobe has been associated with moving the focus of attention from one item to another (Gazzaniga, Ivry & Mangun, 1998).

The results of the previous experiment indicated that participants who are more

absent-minded are also better at solving insight problems. We considered that participants who did not focus on a single aspect of a problem but instead considered several alternative aspects might solve more insight problems. As a result, we hypothesised that good insight problem-solvers would *not* do well on tasks requiring selective attention, such as searching for targets with particular features while ignoring other distracting information, such as searching a crowd for a particular person. To measure this selective attention we used the Map-Search test, which is a subscale of the Test of Everyday Attention battery (Robertson, Ward, Ridgeway & Nimmo-Smith, 1994). This task requires finding all the symbols for a restaurant on a detailed area map, within a time limit. The participant has to ignore other symbols as well as distracting information such as place names and topographical information (see also Sterr, 2004 and Chan, 2000 for other studies using the TEA).

People are also able to switch attention between different sources of information, for example writing an essay while watching television. Davidson (1995) observed that successful insight problem-solvers were more likely to switch strategies when they realised their current strategy was not working. Good problem-solvers may keep several alternatives in mind and so they may need to switch between them in order to evaluate different possibilities rather than think about only one option. Participants who can switch attention accurately may be able to solve more insight problems. Therefore we also included a measure for the ability to switch attention, the Visual Elevator task, again taken from the Test of Everyday Attention battery. This task requires participants to count the floors an elevator passes, switching from counting up to counting down. In addition to this task, we included a task that involved switching between more complex thoughts. The plus-minus task, used by Miyake et al (2000) to

measure the ability to shift between sets, requires participants to change from adding three to numbers to subtracting three from other numbers.

Our final measure was of sustained attention, or vigilance. Vigilance requires maintaining a sense of alertness while waiting for a signal to act. An everyday example is that train drivers need to watch out for possible red lights on their route and if there is a red light, they may need to stop or change course. The Sustained Attention Response Task (SART) was developed by Robertson et al (1997) to measure this aspect of attention by measuring how successful people are at inhibiting an automatic response when a target appears (see also Moores & Andrade, 2000 and Farrin, Hull, Unwin, Wykes & David, 2003 for other studies using the SART). A series of numbers are presented very quickly and the participant must press a button when each number appears except when that number is three, and instead they must refrain from responding. We expected that good insight problem solvers would not do well on this task because it requires focusing on the task and poorer SART performance has previously been associated with higher scores on the Cognitive Failures Questionnaire (Robertson et al, 1997).

In summary, we expected that participants who solved more insight problems correctly would not perform well on the task of selective attention because they would be distracted by other information. In contrast, we expected that good insight problem-solvers would be better at the attention-switching task because they may tend to switch between strategies. We expected that good insight problem-solvers would not perform well on the SART because this task requires focused attention.

The second element of our investigation was the role of working memory. Other studies have implicated a role for working memory in problem-solving, as we

described earlier (e.g. Gilhooly & Murphy, 2004). We considered that working memory would be related to insight problem performance because of the need to keep in mind multiple alternatives and to the need to retrieve information from long-term memory to assist with solving the problem. We used one measure each of storage capacity alone and storage plus processing capacity as it has been suggested that these aspects of working memory can be differentiated (e.g. Conway et al, 2003). To ensure a clear measure of the role of working memory during the problem-solving process, we asked participants to work on solving each problem without recourse to any memory aids such as pen-and-paper.

We examined attention and working memory for both insight and non-insight problems. We expected that there would be a relationship between the two types of problems. Nonetheless, we expected that there would be some differences in the associations with the attention and working memory measures. We also distinguished between verbal and non-verbal problems, as in the previous experiment. Experiment 6 showed that there are some distinctions between these two categories. The main aims of Experiment 7 were to examine the role of attention and working memory in insight problem-solving. A second aim was to compare how non-insight problems were associated with the same measures.

## Method

### *Participants*

Participants were undergraduates recruited by posters on the campus of Trinity College, Dublin and they were paid 10 euro for one hour. There were 33 participants, 17 women and 16 men. Their ages ranged from 19 to 33 years with a mean age of 21.7 years.

***Materials***

Participants received the same set of insight problems as those in Experiment 6, that is eight single-step problems divided into subcategories of four verbal and four non-verbal problems. They also received four well-defined, non-insight problems. These consisted of a syllogism, an algebra problem, a move problem and an arithmetic problem, all of which are given in Appendix C. In our selection of these non-insight problems, we chose problems that, like the insight problems, required a minimal number of steps to solution and could feasibly be attempted without the use of pen-and-paper. They were administered in the same way as the insight problems. The dependent measure for both sets of problems was the proportion of unfamiliar problems solved correctly. Participants also received four attention tasks and two working memory task as follows:

*Map Search*

This task comes from the Test of Everyday Attention battery (Robertson et al, 1994) and is a measure of selective attention. Participants are asked to find all the restaurants on a detailed map of Philadelphia. A restaurant location is indicated by the presence of a knife and fork symbol and the participants have to circle as many of these symbols that they can find in two minutes. After one minute they swap from a red marker to a blue marker so that the experimenter can determine how many targets were found in the first and second minutes respectively. There are 80 symbols on the map but participants are not informed of this at the start (see Appendix C for more detailed instructions). The task aims to measure how distracted participants are by the other information on the map, such as other symbols and place names, by how many symbols

they can find in the time limit. The dependent measures were how many symbols are found within the time limit, which is then scaled according to the participant's age as prescribed in the test manual.

*Visual Elevator*

This task is also taken from the Test of Everyday Attention Battery (Robertson et al, 1994) and measures the ability to switch attention, in this instance from counting upwards to counting downwards. The task is phrased in terms of the participant riding an elevator on which the counter is broken (see Appendix C for more detailed instructions). In order to determine what floor the lift stops at, they have to count the floors the elevator passes as it travels up and down. The elevator's journey is represented on a page with symbols. A symbol of elevator doors represents each floor the elevator passes and occasionally a large directional arrow appears to indicate what direction the lift is now travelling in. For example, a participant's script might be as follows " 1 . . . 2 . . . 3 . . . down . . 2 . . up . . .3 . . .4".

There were two practice trials followed by ten test trials. Each trial is timed. The dependent measures were an accuracy measure of how often the participant ends on the correct floor after a trial and a time measure based on the total time taken for correct trials divided by the number of switches in counting direction needed in those trials. This measure gives an indication of how long it took the participant to switch from counting up to counting down and vice-versa. Both measures were then transformed to scale scores according to the participant's age as described in the Test of Everyday Attention manual.

*Plus – Minus Task*

This task was used to measure switching, akin to the Visual Elevator, but this time switching between the strategies of addition and subtraction. We based our design on one described by Miyake et al (2000) in whose paper this task was found to be related to the Wisconsin Card-Sorting Task. Participants were given a sheet of paper with three columns of 30 numbers each. The numbers used were all the two digit numbers, with each number from 10 to 99 used once only, and randomly mixed to form the three columns. In the first column, participants added the number three to each number in the column and wrote the answer in the space next to it. In the second column, they subtracted three from each number. In the third column, the participants were asked to alternate between adding three and subtracting three, such that they added to the first number, subtracted from the second, added to the third and so on (see Appendix C for actual test). Participants were asked to work as quickly and as accurately as they could. The time taken for each column was recorded and the mean time for the addition and subtraction columns was subtracted from the third "alternating" column to derive a time cost for having to switch between strategies. This difference was the dependent measure and was measured in seconds before being log transformed to the base e prior to analysis. We also derived an accuracy measure by counting all the errors each participant made. Where a mistake in the plus-minus alteration occurred, this switching error was counted as one error rather than counting each number that was incorrect as a direct result of the incorrect switch.

*Sustained Attention Response Task (SART)*

Robertson et al (1997) developed this task to measure sustained attention. It is a

computer-based task in which participants must make a response (in this case pressing the spacebar) every time they see a number, except when that number is three in which case they have to withhold the response. The task has two versions, one in which the numbers appear in a fixed, predictable sequence (e.g. 1, 2, 3) and one in which the numbers appear in a random sequence (e.g. 8, 3, 5). We used the random sequence as the fixed sequence is very easy for normal participants. We wrote a SART program in Superlab based on the details given in the above paper. A total of 225 single digits were presented, 25 of which were the target number three. Each number was presented for 250msec followed by a 900msec mask. The target figures were distributed across the presentations in a quasi-random fashion such that there were 25 blocks of nine digits each with the digits within each block completely randomised. Five different font sizes (symbol font) were used so that it would be necessary to process the numerical value rather than just make a perceptual judgement based on the shape of the stimulus. These font sizes were 48 point, 72 point, 94 point, 100 point and 120 point. The figures were black against a pale background. The mask consisted of a circle with a diagonal cross within it. Participants were told that they could still respond while the mask was on the screen if they had not had time to respond to the number. Before commencing the test trial each participant did a practice trial consisting of eighteen digits, two of which were targets. The dependent measure was the number of targets to which the participants responded when they should have withheld a response.

*Digit Span*

This task was used to measure the storage capacity of working memory. Participants were asked to recall ever-increasing strings of digits read out by the experimenter. The

participants had to recall the numbers in a forward direction for two trials. The experimenter read out a practice three-digit sequence before starting. The forward sequence started with three single digits up to a maximum of nine digits with two trials of each length. When all forward trials were completed, participants were then asked to recall number strings in a backwards direction. The backward sequence was similar to the forward sequence but started with two digits up to a maximum of eight. Participants were given one point for each string correctly recalled. The procedure used was that described by the WAIS-R Manual (Weschler, 1981) and testing was halted if a participant made an error in two consecutive trials of the same length. The dependent measure was the number of correctly recalled strings with a maximum possible score of 28. This score was then scaled according to the age of the participant in keeping with the procedure advised in the WAIS-R manual.

*Sentence Span*

This task was used to measure the capacity of working memory when both processing and storage of information was required. Adapted from a design by Sub et al (2002), participants were asked to rate a series of sentences as true or false and, in addition, to remember the last word of each sentence within a block. The task was presented on a computer using the Superlab program. All sentences were easily true or false and no longer than seven words. The last word of each sentence was always a singular noun between one and three syllables in length. The list of sentences used is included in Appendix C and took the form "The letter k is a vowel" to which the participant responded by pressing a key to indicate if the sentence was true or false.

Each sentence was displayed for three seconds and if no true/false response had

been given within that time a prompt appeared for one second, giving participants four seconds in total to read the sentence and indicate true or false. Participants started with a practice trial of two sentences and after that attempted two blocks of three sentences eventually increasing to two blocks of six sentences. At the end of each block, participants had to recall the last word of each sentence in that block in order if possible. Therefore the number of words to be recalled started at three and increased to six. Before each trial participants were told how many sentences would be in the following block. As with the digit span, testing was halted if participants made errors in two consecutive trials of the same length. The dependent measure was the total number of words recalled correctly up to a maximum of 36.

***Procedure and design***

The procedure for administration differed slightly from that of the previous experiment as participants were not allowed to write while attempting to solve the problems. This change was introduced to maximise the load on working memory as outlined earlier. Participants were tested individually and said aloud their answers to the experimenter when they were ready, within the two minute time limit.

Scores on these problems were compared with performance on four measures of attention and two of working memory. All participants started with the insight problems, the order of which was randomised for each participant. The other tasks were then given in the following order: non-insight problems (in random order), Visual Elevator, Digit Span, Map Search, Plus-Minus task, Sentence Span, SART.

**Results**

All correlations are Pearson's r, n = 33, one-tailed unless otherwise stated. The score

for insight problem-solving was calculated in the same way as for Experiment 6, with the number of correct solutions expressed as a proportion of the total unfamiliar problems attempted. However, on this occasion there were no previously seen problems. Overall, the mean proportion of insight problems correctly solved was .52. A similar score was also calculated for the verbal and non-verbal problems. For the verbal problems, the mean score was .63 and for the non-verbal problems it was .42. The mean scores for all the problem categories are given in Table 4.4.

**Table 4.4:** Mean scores, standard deviations and possible scoring range for all problem categories in Experiment 7

| **Task** | **Mean Score** | **SD** | **Possible Range** |
|---|---|---|---|
| Insight | .52 | .28 | 0.00 – 1.00 |
| *verbal* | .63 | .35 | 0.00 – 1.00 |
| *non-verbal* | .42 | .31 | 0.00 – 1.00 |
| Non-insight | .52 | .26 | 0.00 – 1.00 |

***Correlation of insight problem score with measures of attention***

The measures of attention included the Map Search, Visual Elevator, Plus-Minus and SART tasks. The mean scores of all the attention measures are summarised in Table 4.5.

Correct performance on the Map Search (selective attention) task was not related to correct performance on the insight problems ($r = .023$, $p = .449$) as expected. This result suggests that being able to ignore the distracting information was not related to better performance on the insight problems.

Correct performance on the Visual Elevator (attention-switching) task was associated with correct performance on the insight problems ($r = .515$, $p < .01$) as

expected. However, faster switching on the Visual Elevator task was not associated with correct performance on the insight problems ($r = .069$, $p = .352$). A similar pattern was seen in the Plus-Minus task, which involved switching between adding and subtracting. Higher scores on the insight problems were associated with fewer errors on the Plus-Minus task ($r = -.511$, $n = 32$, $p < .001$) but not with the time taken to switch between functions ($r = -.085$, $n = 32$, $p = .323$). These results suggest that good insight problem-solving is associated with accurate switching but not with speed to complete the switch.

**Table 4.5:** Mean scores, standard deviations and possible scoring ranges, where applicable, for all attention measures in Experiment 7

| Task | Mean Score | SD | Possible Range |
|---|---|---|---|
| Map search (raw) | 78.03 | 2.94 | 0 - 80 |
| Map search (scale)† | 12.64 | 3.30 | |
| Visual elevator (acc/raw) | 8.82 | 1.61 | 0 - 10 |
| Visual elevator (acc/scale)† | 11.39 | 2.66 | |
| Visual elevator (time) | 6.97 | 1.96 | - |
| Plus-minus (errors) | 1.56 | 2.11 | - |
| Plus-minus (log time) | 3.30 | .57 | - |
| SART (errors) | 6.58 | 4.67 | 0 – 25 |

Key: † Only the scale scores were used in analysis
acc/raw = accuracy raw score
acc/scale = accuracy scaled score

The last task measured the errors made in the sustained attention task, the SART. Fewer errors on the SART was not associated with more correct insight problems ($r = -.079$, $n = 31$[1], $p = .336$) as expected, but neither did good insight problem-solvers make more errors. The values of the correlation between the insight problems and the attention measures are summarised in Table 4.6 and we will refer to the verbal and non-

[1] A technical error resulted in scores for two participants on this measure being lost

verbal categories shortly.

**Table 4.6:** Correlation between insight problems and measures of attention in Experiment 7

| Task | Insight (8 problems) | Verbal (4 problems) | Non-verbal (4 problems) |
|---|---|---|---|
| Map search | .023 | -.121 | .229 |
| Visual elevator (acc) | .515** | .483† | .407† |
| Visual elevator (time) | .069 | .063 | .051 |
| Plus-minus (errors) | -.511** | -.563†† | -.265 |
| Plus-minus (time) | -.085 | -.107 | -.087 |
| SART | -.079 | -.207 | .025 |

Key: * $p < .05$, one-tailed † $p < .05$, two-tailed
** $p < .01$, one-tailed †† $p < .01$, two-tailed

***Correlation between insight problems and measures of working memory***

The working memory measures were Digit Span and Sentence Span. The means for these measures are given in Table 4.7.

**Table 4.7:** Mean scores, standard deviations and possible scoring ranges, where applicable, for both working memory measures in Experiment 7

| Task | Mean Score | SD | Possible Range |
|---|---|---|---|
| Digit Span (raw) | 18.82 | 4.48 | 0 – 28 |
| Digit Span (scale) | 12.58 | 3.33 | - |
| Sentence Span | 18.29 | 10.24 | 0 – 36 |

High performance on a measure of working memory, in terms of storage capacity only, measured by the Digit Span scaled scores, was reliably related to higher insight problem score ($r = .405$, $p < .05$). High performance on a measure of working memory,

for processing and storage, measured by the total number of words recalled on the Sentence Span task, was also related to higher scores on the insight problems ($r = .396$, $n = 31$[2], $p < .05$). The correlation values are summarised in Table 4.8. These results suggest that being able to store more information in working memory and have more capacity to process the information is associated with an increased ability to solve insight problems.

**Table 4.8:** Correlations between insight problems and measures of working memory in Experiment 7

| **Task** | **Insight** (8 problems) | **Verbal** (4 problems) | **Non-verbal** (4 problems) |
|---|---|---|---|
| Digit Span | .405* | .358* | .342* |
| Sentence Span | .396* | .456** | .250 |

Key: * $p < .05$, one-tailed
** $p < .01$, one-tailed

***Regression Analysis of Insight Problem Score***

We also analysed insight problem score as the dependent variable in a regression analysis with all the attention and working memory measures. From a stepwise regression, the most efficient model that emerged accounted for nearly 72% of the variance in insight problem score. The predictor variables and their direction were: fewer errors on the plus-minus task, higher score on the digit span, greater accuracy on the visual elevator and more time taken to switch on the plus-minus task. The co-efficients for the variables are summarised in Table 4.9. The adjusted R square was .717 and the model was significant ($F[4,23] = 18.067$, $p < .01$).

[2] Scores for participants who did not follow instructions correctly on this measure were not included

**Table 4.9:** Summary table for results of stepwise regression in Experiment 7 with insight problem score as dependent variable

| **Predictor Variable** | **Beta** | ***p*** |
|---|---|---|
| Plus-Minus Errors | -.439 | < .01 |
| Digit Span | .498 | < .01 |
| Visual Elevator (accuracy) | .420 | < .01 |
| Plus-Minus Time | .248 | < .05 |

***Differences between verbal and non-verbal insight problems***

All correlations are Pearson's r, n = 33. As can be seen from Tables 4.6 and 4.8, the verbal and non-verbal insight problems had similar patterns of correlation. However, there were some differences. On the attention measures, correct performance on the Map Search had a higher correlation with the non-verbal problems (r = .229, p > .05, two-tailed) than with the verbal problems (r = -.121, p > .05, two-tailed) but neither of these values was reliable. The Visual Elevator task had similar correlation values with both the verbal problems (r = .483, p < .05, two-tailed) and the non-verbal ones (r = .407, p < .05, two-tailed). Greater accuracy on the plus-minus task was strongly related to the verbal problems (r = -.563, n = 32, p <. 001, two-tailed) but had a weaker relationship with the non-verbal ones (r = -.265, n = 32, p > .05, two-tailed). There were fewer SART errors associated with more correct verbal problems (r = -.207, n = 31, p > .05, two-tailed) but the correlation with the non-verbal problems was close to zero (r = .025, n = 31, p > .05, two-tailed) although neither of these values was reliable.

Within the working memory measures, the Sentence Span task had a strong correlation with the verbal problems (r = .456, p < .05, two-tailed) but the non-verbal correlation was not significant (r = .250, p > .05, two-tailed) albeit in the same direction. The Digit Span had similar correlation values with the verbal (r = .358, p <

.05, two-tailed) and non-verbal problems (r = .342, p = .05, two-tailed).

The correlation between the verbal and non-verbal problems was significant (r = .457, p < .005). As in the previous experiment, participants found the non-verbal problems harder than the verbal (.43 v .62, paired t = 3.485, df = 32, p < .005, two-tailed).

***Correlation of non-insight problems with insight problems and other measures***

Higher scores on the non-insight problems were reliably associated with higher scores on the insight problems (r = .378, p < .02). The correlation between the non-insight and verbal insight problems was also reliable (r = .311, p <.05), as was the correlation with the non-verbal insight problems (r = .352, p < .03). The correlation values between the non-insight problems and the other measures are summarised in Table 4.10.

Performance on the non-insight problems was reliably related to three of the attention measures, and for each of these the relationship was different to that observed with the insight problems. More targets found on the Map Search task was reliably associated with higher scores on the non-insight problems (r = .338, p < .03). Score on this task were not related to the insight problem score.

For both switching tasks, the Visual Elevator and Plus-Minus, higher scores on the non-insight problems were associated with less time taken to switch between strategies: Visual Elevator (r = .353, p < .03), Plus-Minus (r = -.478, p < .01). The relationship with the accuracy scores was not significant on either the Visual Elevator (r = .224, p = .105) or the Plus-Minus task (r = -.121, n = 32, p = .254). This pattern was the opposite of that observed with the insight problems where accuracy was related to insight score but speed was not.

**Table 4.10:** Correlation between non-insight problems and measures of attention and working memory in Experiment 7

| Task | Non-insight score |
|---|---|
| **Attention** | |
| Map search | .338* |
| Visual elevator (accuracy) | .224 |
| Visual elevator (time) | .353* |
| Plus-minus (time) | -.478** |
| Plus-minus (errors) | -.121 |
| SART (errors) | .070 |
| **Working Memory** | |
| Sentence span | .511** |
| Digit span | .390* |

Key: *p = < .05, one-tailed
**p = < .01, one-tailed

The non-insight problem score was related to both of the working memory measures, as was the insight problem score, with greater working memory capacity associated with greater success on all problems. The correlation of the non-insight problems with the Sentence Span task was r = .511 (n = 31, p < .01) and was r = .390 (p <.02) with the Digit Span and both were reliable.

***Relationships between measures***

Table 4.11 shows the correlations found between the attention and working memory measures. We did not make specific predictions about the relationships between the measures but we had anticipated some, such as between the two working memory tasks and between the attention-switching tasks of Visual Elevator and Plus-Minus. The data confirmed the relationship between the Sentence Span and Digit Span (r = .426, n = 31, p < .05, one-tailed). However, the two switching tasks, the Visual Elevator and

Plus-Minus were not reliably related. The direction of the two accuracy scores was as expected with more correct trials on the Visual Elevator associated with fewer errors on the Plus-Minus task ($r = -.263$, $n = 32$, $p = .073$, one-tailed) but the relationship did not reach significance levels. Similarly, faster switches on the Visual Elevator were associated with faster switches on the Plus-Minus task ($r = -.208$, $n = 32$, $p = .126$, one-tailed), but not reliably so.

**Table 4.11:** Correlations between the other measures, attention and working memory in Experiment 7

| | **Visual Elevator** | | **Plus-Minus** | | **SART** | **Digit** | **Sentence** |
|---|---|---|---|---|---|---|---|
| | *Acc.* | *Time* | *Errors* | *Time* | | **Span** | **Span** |
| *Attention* | | | | | | | |
| Map Search | .238† | .409 | .171 | -.132 | .321 | .102 | -.099 |
| Visual Elevator (acc.) | | .062 | -.263 | -.254 | -.044 | .203 | .251 |
| Visual Elevator (time) | | | -.068 | -.208 | .253 | .113 | .383 |
| Plus-Minus (errors) | | | | | .162 | -.177 | -.208 |
| Plus-Minus (time) | | | | | .099 | -.392 | -.504 |
| SART (errors) | | | | | | .026 | -.220 |
| *Working Memory* | | | | | | | |
| Digit Span | .203 | .113 | -.177 | -.392† | .026 | | |
| Sentence Span | .251 | .383† | -.208 | -.504** | -.220 | .426* | |

Key: * $p < .05$, one-tailed † $p < .05$, two-tailed
** $p < .01$, two-tailed

Among the other attention measures, there was a relationship between faster switches on the Visual Elevator and more targets identified on the Map Search ($r = .409$, $n = 33$, $p < .05$, two-tailed). There was no significant correlation between the SART and any of the other attention measures. However, the correlation with the Map Search task was not far outside the significant range ($r = .321$, $n = 31$, $p = .078$, two-tailed), with more targets incorrectly responded to on the SART associated with more targets found on the

Map Search.

Between the attention and working memory measures, higher scores on the Sentence Span measure was correlated with faster times on both the Visual Elevator ($r = .383$, $n = 31$, $p < .05$, two-tailed) and Plus-Minus tasks ($r = -.504$, $n = 30$, $p < .01$, two-tailed). The other working memory measure, the Digit Span, was significantly correlated with the plus-minus time in the same direction, but weaker ($r = -.392$, $n = 32$, $p < .05$, two-tailed) and there was no significant correlation with the visual elevator time score ($r = .113$, $n = 33$, $p = .531$, two-tailed).

**Discussion**

This experiment provides four main pieces of information about the roles of attention and working memory in insight problem-solving: (1) Ability to divide attention as measured by accuracy in both the Visual Elevator and Plus-Minus tasks was associated with better insight problem-solving. However, neither good selective attention nor good sustained attention was associated with insight problem-solving. (2) Working memory capacity for both storage only, and storage plus processing is associated with both insight and non-insight problem-solving. (3) The combination of working memory storage capacity, accuracy on the attention-switching measures and time on taken on the plus-minus task accounted for 72% of the variance in insight problem scores. (4) There were different patterns of association for the insight and non-insight problems, particularly on the attention measures.

## General Discussion

In Experiment 6, we examined the role of imagery in insight problem-solving. We expected that being able to mentally simulate an image of the situation described in the

problem and to manipulate that image would be associated with better performance on a range of insight problems. The results indicated that some imagery ability may be needed to solve insight problems. People who scored higher on the Mental Rotations Task solved more insight problems correctly. People who reported that the images they created were clearer in the Questionnaire upon Mental Imagery did not answer more insight problems overall, but this ability seemed to be more related to the non-verbal problems than the verbal ones.

Experiment 6 also compared insight problem score with scores on a more verbal measure, the Remote Associates Test. Greater verbal ability was associated with higher scores on the insight problems but on closer inspection this relationship seems to have been largely related to correct performance on the verbal rather than the non-verbal problems. This finding suggests not only a distinction between insight problems requiring language skills and other insight problems but also raises questions for previous studies that have linked performance on the RAT with a mixed selection of insight problems (e.g. Schooler & Melcher, 1995). We also found that higher self-reported absent-mindedness as measured by the Cognitive Failures Questionnaire was associated with more correct insight problems.

The results relating to the CFQ in Experiment 6 suggested a possible role for attention in insight problem-solving. The results of Experiment 7 provide further support for this proposition. In this experiment, a model combining one measure of working memory and three measures of attention was able to explain 72% of the variance on the total insight problems score. The working memory association occurred for both the insight and non-insight problems. However, the pattern of association with the attention measures was quite different for the insight and non-

insight problems. We speculated that switching one's attention from one strategy to another and to salient information is central to insight problems. However, focussing on one strategy and a limited set of key information may be more central to solving non-insight problems.

Our findings from the experiments in this chapter may have important implications for existing theories. The importance of working memory and attention supports search-based theories (e.g. Chronicle et al, 2004) and the idea that how people plan their moves and select strategies is an important part of insight problem-solving. It suggests that there may not be a single 'insight skill' but that ability on these problems is a combination of several executive functions. There was also some support for the representation-based theories regarding the role of imagery, however, in the regression analysis neither imagery measure accounted for any additional variance in insight problem-score over that which was explained by the CFQ.

Our aim in this series of experiments had been to examine what skills were useful to good insight problem-solving. The results indicate that executive functions such as working memory and attention are important to insight problem-solving and that different aspects of attention may help to distinguish between the processes used for insight and non-insight problems. We also saw that different skills such as imagery and verbal skills are required for insight problems with different contents and that to be a good all-round insight problem-solvers probably requires a broad skill set.

# CHAPTER 5

Insight problems have interested researchers for nearly a hundred years. Despite the substantial body of work that has accumulated since then, we still do not fully understand how people reason with insight problems. Even though people possess the knowledge necessary to solve problems like the nine-dot problem, such problems typically have extremely low solution rates. The nine-dot problem requires the person to connect together nine dots, arranged in a square formation, using just four connected straight lines without lifting the pen from the page. Most adults are fully capable of drawing lines to join dots but why do most of them fail to solve this problem?

In the contemporary literature, there have been two prominent approaches to explain insight problem-solving. One approach (e.g. Ohlsson, 1992) emphasises the role of representation. When people misrepresent the requirements of the problem or its solution, the flawed representation fails to bring the relevant knowledge into awareness. The other approach (e.g. Ormerod et al, 2002) emphasises the processes that people employ to help them navigate the space between the problem and its solution. General shortcuts, called heuristics, are used to select a move that seems promising and likely to lead to a solution. However, with insight problems the promising move is often incorrect and so heuristics can impede rather than contribute to the solution process.

Our principal aim in this work was to further our understanding of how people reason with insight problems by assessing what manipulations can improve performance and what other skills are associated with success on insight problems. In a series of three experiments in Chapter 2, we explored the role of representation in insight problems and devised manipulations aimed at improving or overcoming a

misleading representation. How people mentally represent a problem and its solution is considered very important in a number of accounts of insight, such at the representational change theory proposed by Knoblich et al (2001). In a series of two experiments in Chapter 3, we manipulated the size of the search space participants had to navigate in two experiments. Again the search space has been nominated as an important factor in several accounts of insight including the computational model proposed by Chronicle et al (2004). To complement the testing of these theories through experimental manipulation, in two studies in Chapter 4 we explored what skills were associated with better performance on insight problems.

In this final chapter we will summarise the results of our seven experiments, discuss the implications and limitations of the work and suggest some future directions.

## Summary of Findings

### *The role of representation*

Chapter 2 focused on the representation of insight problems. As we discussed in Chapter 1, the role of problem and/or solution representation is a key component of several theories of insight. The representational change theory of Knoblich, Ohlsson and colleagues (2001) is at the forefront of this view in the current literature. It proposes that the flawed representation stands in the way of a solution, and when this representation is properly restructured, insight can occur and the path to a solution can progress. Although an exact definition of an insight problem remains controversial, several authors including Ohlsson (1992) and Weisberg (1995) have emphasised that the representation of an insight problem is often flawed, leading to misdirected efforts at a solution. Furthermore, they suggest that there is consistency in the way people's

representations of a given problem are flawed.

Chapter 2 reported three experiments that we conducted on the role of representation in insight problems. Our first task was to confirm that there was in fact some consistency in the kind of errors people make when working with a variety of insight problems (nine-dot, candle-box, Duncker's radiation problem, six-matches, ping-pong, camping-trip, window-cleaner and woods-walk). We approached this task by asking participants to make their representations explicit, in the form of a diagram or an explanation to a friend. This methodology confirmed that there was in fact a high degree of consistency in the kind of errors made on a range of problems and, furthermore, that making the expected error was associated with failing to solve the problem. In this experiment we also found that asking participants to represent the problem before attempting a solution did not help them reach a solution. However, we found that participants who offered a correct solution did so during the first task even when that first task was to represent the problem.

We had included a mixture of what we termed single-step problems and multiple-step problems. The ping-pong problem is an example of a single-step problem, and the nine-dot an example of a multiple-step problem (see Appendix A). We expected multiple-step problems to be more difficult because a single insight was not sufficient for a complete solution, in contrast with the single-step problems. In this experiment and in those subsequent, the distinction was found to be valid with solution rates on single-step problems exceeding those on multiple-step problems.

Having verified the common errors in representation for a range of problems in the first study, we moved on to examine the effects of changing participants' mental representations on a subset of four of these problems, two single-step and two multiple-

step (ping-pong, window-cleaner, nine-dot and six-matches). If representation is indeed a key component in the insight problem-solving process, then starting participants off with a better representation should have a positive effect on the frequency with which they reach a solution. In the second experiment of Chapter 2, we tested our prediction that how a problem is presented to the participant in terms of its exact wording would effect its representation and its subsequent solution rate. For example, in the change to the ping-pong ball problem we asked participants how they could throw a basketball instead.

We also predicted that the multiple-step problems would still be more difficult as a change in representation would not lead directly to a solution by itself. A comparison of performance on problems with the standard wording and those with an improved wording showed that the solution rates nearly doubled for the single-step problems. Only one participant solved one of the multiple-step problems. The poor performance on these problems could have been because we were not successfully targeting the participants' misrepresentations of these problems or because two minutes was not enough time. A review of participants' scripts suggested that for all problems, participants who made the expected error in misrepresentation usually failed to solve the problem, confirming our findings in the previous experiment that misrepresentation is an important factor in insight.

A second element of this experiment was the request to participants to suggest a hint to a friend attempting to solve the problems, after they themselves had seen the correct solution. The suggested hints indicated that participants considered a representation of how the solution would look to be important or more helpful than instructions of what to do. For example, clues alluding to the triangular shape of the

nine-dot problem were more common than the instruction to extend lines.

The third and last experiment in this chapter focused on representation through the use of physical props. We reasoned that participants may have difficulty constructing a working mental representation of the situation described in a given insight problem. For example, they may find it difficult to accurately simulate what would happen to a ping-pong ball thrown in a particular way. Therefore, we gave some participants physical props corresponding to the items described in four insight problems (e.g. an actual ping-pong ball) and compared their performance to participants who attempted the same problems using only pen-and-paper. The problems were the same ones used in the previous experiment (see details of the props used in Appendix A).

In addition, we gave participants extra time (10 minutes) and allowed them more than one attempt. The results showed that the provision of physical props (and extra time) made a difference to participants' ability to solve the problems. The extra time was particularly useful for the multiple-step problems across both conditions. The nine-dot problem was the only problem that showed no benefit from the physical props.

The props may have helped participants to identify mistakes in their representations. For example, in the window-cleaner problem participants must explain how a man falls from a sixty-foot ladder without injury. There is a tendency to assume that the man falls from the top of a sixty-foot ladder. However using the model ladder and figure may have helped participants to realise that the man did not start off at the top of the ladder. He had to start at the bottom and climb up, which may have helped participants to realise that the man could have fallen off at any point, not just from sixty-foot up. Similarly, when participants try to throw the ping-pong ball with a back-

spin or fit four equilateral triangles into a two-dimensional array, they can see that such efforts will not work and that they must try something else.

All three experiments support the notion that how a problem is represented is influential in the solution process. In the first two experiments, there was a clear link between misrepresenting the problem in the expected way and failing to solve that problem. In the second and third experiments, simple manipulations that did not alter the requirements of the solution but altered the way participants thought about the problem boosted solution rates on a range of problems. We have also seen, however, that efforts to manipulate the representation seldom result in 100% solution rates and that not all problems benefit equally well from similar manipulations. It seems likely that another factor, apart from the representation error, is contributing to the difficulty of some problems, especially the nine-dot problem. A candidate for this other factor, the search space, was investigated in Chapter 3.

***Searching for a solution***

In Chapter 3 we examined how searching through a large problem space could contribute to the difficulty of insight problems. The computational model of Chronicle et al (2004) emphasises the role played by heuristics in insight problem-solving. When faced with a choice of possible starting moves, they have shown that people tend to choose a move that seems to make the most progress towards the solution (a locally-rational strategy). Problems arise when the promising move is not the correct first move needed to reach a solution. Only when participants have exhausted the early, deceptively promising moves (criterion-failure) do they extend their search to include other moves. Good evidence in support of this theory has been shown, for example with the nine-dot problem (Chronicle et al, 2001) and various coin problems (Ormerod

et al, 2002; Chronicle et al, 2004). So far, however, this research has concentrated on problems that require physical moves. If this theory is to provide a general account of insight problem solving then it also needs to be tested with problems that are considered to require insight but do not use physical moves.

The chapter included two experiments on the theme of examining the effect of restrictions on the search space. We predicted that if having to select moves from an extensive range of possibilities is a factor in the difficulty of insight problems, then limiting the number of moves that need to be considered should facilitate performance. In Experiment 4, we chose four problems that required the physical movement of pieces, two that required multiple moves (T-puzzle and necklace problems) and two that required only a single move (matchstick equation and horse-and-rider problems). In the example of the T-puzzle, the participant has to arrange four pieces into the shape of a 'T'. There are many possible combinations but only one solution. We limited the number of possible combinations by giving participants an actual size template that they could use to test piece-placements (the other problems and the manipulations are given in Appendix B). The results showed that our manipulations to limit the number of possibilities to be considered had a positive effect for all four problems. Three of the problems had reliably more correct solutions when the options were limited, and the fourth problem (horse-and-rider) showed reliably faster times to solution in this condition.

We had also expected that the multiple-move problems would be more difficult than the single-move problems because of their more extensive problem space. The results supported this prediction, with participants who received the unrestricted problems displaying typical floor-level performance on both multiple-step problems.

The next stage of our enquiry, in Experiment 5, was to try to replicate these results with problems that did not involve physical moves. We used negative hints, that is telling participants what not to do, which were aimed at diverting participants away from a frequently adopted, but incorrect, solution strategy as a means of limiting the possibilities they had to consider. For example, a frequent but incorrect solution that is offered for the ping-pong ball problem is to "put a backspin on it". Therefore, the negative hint we gave participants for this problem was "there is no need to put a backspin on the ball". We used four problems in total: ping-pong, window-cleaner, camping-trip and marrying-man (further details of these problems are given in Appendix B).

The results from these problems suggested that search space was not as important in determining difficulty as it had been for the physical problems. Two of the four problems (window-cleaner and camping-trip) showed some advantage for receiving a hint as measured by the frequency of correct solutions. However, the restrictions did not have the same comprehensive effect that we had observed with the move problems in Experiment 4.

A review of the solutions offered did suggest that participants who received a hint were searching in the right space more often than those who received no hint. We categorised the solutions offered as showing evidence of a good quality representation or not. Sometimes participants seemed to have a good representation of the problem even if they did not get the correct solution. For example, for the ping-pong ball problem some participants suggested rolling the ball up a slope. This answer indicates that they had correctly represented gravity as the force that causes the ball to return but as the problem specifically says to 'throw' the ball rather than 'roll' it, this answer

could not be considered to be correct. Therefore we have an instance where there is a good representation of the problem that did not result in a correct solution. When we examined the frequency of good quality representations rather than just correct solutions, the advantage of receiving a hint was more evident with three of the four problems showing such an advantage.

The results of these last two experiments suggest that the relative influence of representation versus search space may vary for different problems. In problems where there is the physical movement of pieces, one could in principle construct every possible combination of moves and eliminate them one by one. Such problems may be what Weisberg (1995) described as 'hybrids', problems which may be solved through an insight which allows a kind of shortcut to the solution, but which may also be solved through a process of trial-and-error alone with no necessity for insight. For example, in the T-puzzle (see Appendix B) a person may have the insight that the 'notch' forms a corner and does not need to be filled and may adjust his or her attempts accordingly. However, it is also possible to test every combination of pieces through a process of trial-and-error until one eventually discovers the right combination.

For other problems, there is an additional requirement to generate all the possibilities, or at least possible search spaces, before a search can commence. For example, with the ping-pong ball problem there are a number of paths along which the solution might be located. The solution might be something to do with the physical properties of the ball (e.g. lightweight) or a special technique for throwing the ball (e.g. backspin) or the force that operates on the ball (e.g. gravity). The last space contains the correct solution but if a person has not considered the force operating on the ball, then it may not occur to that person simply through trial and error. However, assisting

participants to avoid searching through unhelpful possibilities may speed them towards generating and considering the correct possibility. This type of assistance may not, however, be as efficient as targeting participants' representation of the problem directly because eliminating other possibilities requires more work than highlighting the correct option straight away. Therefore, for some problems manipulations such as substituting a basketball for the ping-pong ball may work better than ruling out one throwing option.

Despite the negative hints being less helpful than anticipated on problems without actual moves, the review of the participants' representations did suggest that for all but one of these problems (marrying-man problem), the negative hint had succeeded in directing them towards the right space. The reason why that problem did not benefit may have been as a result of the hint misdirecting participants towards a distinction between cultural and legal permission. To direct participants away from the idea that the man was a Mormon, or from some other culture that allowed polygamy, we told them that the man's culture did not permit multiple wives. However, this hint may have suggested to participants that even if the culture did not allow polygamy, there might have been a legal loophole. Alternatively, it may be that there were not as many possibilities occurring to participants, that is they either thought the man was the groom or the priest and so if the priest option did not occur to them, they stuck to the idea that the man was the groom.

***Representation and search space***

Collectively, the results of the first five experiments indicate that both representation and search space play a role in how people reason with insight problems. As can be

seen from Table 5.1, solution rates on the window-cleaner problem were increased following both manipulations aimed at the representation and those aimed at restricting the search space. The ping-pong ball problem also benefited from both types of manipulation but only when measured in terms of good representations for the latter. What is also evident from Table 5.1, however, is the difficulty in finding a single manipulation that affects all problems similarly. This observation has important implications for future studies of insight. Researchers cannot necessarily assume that what applies to one insight problem will apply equally to all other problems presumed to require insight.

**Table 5.1:** Summary of percentage increase in solution rates from control to experimental conditions for each problem in each experiment in which it was used

| | **E1** | **E2** | **E3** | **E4** | **E5(Neg. Hint)** | |
|---|---|---|---|---|---|---|
| | Represent | Reworded | Prop | Space | Before | After |
| **Single step** | | | | | | |
| Ping pong | 22 | 24 | 0 | - | 2 | -2 |
| Window cleaner | -25 | 24 | 21 | - | 28 | 34 |
| Camping trip | 13 | - | - | - | 4 | 24 |
| Woods walk | 8 | - | - | - | - | - |
| Marrying-man | - | - | - | - | 11 | 0 |
| Horse-rider | - | - | - | 5 | - | - |
| Equation | - | - | - | 27 | - | - |
| **Multiple step** | | | | | | |
| Nine-dot | 25 | 0 | -13 | - | - | - |
| Six matches | 0 | 7 | 43 | - | - | - |
| Candle | 26 | - | - | - | - | - |
| Radiation | 11 | - | - | - | - | - |
| T-puzzle | - | - | - | 48 | - | - |
| Chains | - | - | - | 40 | - | - |

Reconciling the relative roles of representation and search space may involve acknowledging that both are factors in determining the difficulty of a problem and

instead trying to quantify their individual contributions to different problems. We speculate that the effect of manipulations aimed at either representation or search space differ in their effects on different problems because one or other may be the more influential in determining difficulty. For example in problems such as the nine-dot or T-puzzle, there is a large space containing all the possible combinations of moves. The first difficulty facing participants is choosing a starting point. In contrast, for problems such as the marrying-man it can initially be difficult to find anywhere to start because the problem seems impossible. Therefore the difficulty with some problems is there are too many possible starting options, but for others there are not enough. This distinction may define those problems that will benefit more from manipulations to reduce choice or to expand it by opening up other options following changes to the representation.

One of the problems where the representation versus search space debate gets interesting is the matchstick equation problem, where participants have to balance a false equation, in roman numerals and constructed of matches, by moving just one of the matches. This type of problem has been exclusively used so far in conjunction with the representational change theory (e.g. Knoblich et al, 2001). Yet we found that our efforts to reduce the number of possible moves to be considered significantly benefited participants who attempted this problem. Proponents of the representational change theory who have used these equations may argue that we changed the representation by confining moves to starting with the sign matches. We did not, however, confine participants to moving the match to make up another sign and they could in fact have moved it anywhere to several locations within the equation. Indeed, in the control condition some participants suggested moving a match to change the 'equals' sign into the sign for 'not equals', ≠ or ≡, suggesting that they were not specifically inhibited

from trying to alter the signs as opposed to the numbers. Just because problems may need some changes to the representation does not exclude the possibility that having to search through a space of possibilities is also a factor.

Similarly, representation may also be a factor in problems that have implicated a role for heuristics to limit search space. For example, Chronicle et al (2004) used a six-coin problem to demonstrate their computational model concerning the effects removing the availability of a seemingly promising, but ultimately incorrect, move. They acknowledged that how the participants represented the solution influenced their success at the problem. Participants who were working towards a solution consisting of two groups of two coins solved the problem more often than those who were aiming for a ring-shaped solution. In Experiment 2, a description of how the finished solution would look was the most common hint suggested by participants for the nine-dot and six-matches problems. It may be that the representation of what the solution will look like influences a participant's judgement on what is a good first move. Ormerod et al (2002) observed that in the nine-dot problem, an extended line cancelling the three top dots was not as helpful as a diagonal line cancelling the centre dots. This failure of the hint may have been because the former line did not obviously contradict an incorrect representation of the solution as a kind of square-shape, even though the extended line did break the square formed by the dots themselves. Furthermore, the hints given by the participants for the nine-dot problem in Experiment 2 suggested that they were surprised by the triangular shape of the solution. Although it is worth remembering that participants did not seem to be able to use a hint of a shaded arrow-shape in the Ormerod et al study mentioned above, it was possibly because the arrow shape could not be reconciled with participants' own expectations for a box-shaped solution.

Our conclusion is that a theory that will account for all problems presumed to require insight would have to include an account of search space and mental representations. The fourth chapter reported two further experiments to help us understand the processes and skills that people employ when reasoning about insight problems.

***Individual differences***

The results from our first five experiments suggest that it is difficult to pinpoint one key manipulation that works equally well for all insight problems. We therefore extended our line of enquiry towards identifying what skills are necessary for solving insight problems, keeping in mind that it may be possible to identify a set of skills that predict a good all round insight problem-solver. We also wanted to investigate whether there was a distinction between problems that ask participants to do something, such as manipulate objects or rearrange pieces, for example the horse-and-rider problem (non-verbal), and those whose difficulty rests on the wording of the problem is interpreted, for example the marrying-man problem (verbal). We have previously identified differences between multiple-step and single-step problems. However, it is also true that most multiple-step problems are non-verbal rather than verbal. In a study of individual differences, the traditional multiple-step problems like the nine-dot are too difficult to afford comparisons with scores on other measures. Therefore we selected four single-step problems, which normally have higher solution rates than multiple-step problems. We selected four problems that required participants to think about how they could manipulate objects to reach a solution (horse-and-rider, ping-pong, inverted-pyramid and matchstick equation). These problems were used in addition to four

problems that exploited how the wording of the problem was interpreted (camping-trip, marrying-man, woods-walk and window-cleaner).

Our first experiment in Chapter 4 concentrated on how imagery might be used in insight problem-solving. We included two measures of imagery ability. The first measure was an objective test of visual imagery, the Mental Rotations Task (MRT) from a version by Peters (1995). This task requires participants to mentally rotate an abstract block figure and then identify the new correct image from a choice of images (see Appendix C). Comparing the scores on this task to scores on the set of insight problems indicated that imagery ability was associated with performance on insight problems, although the correlation was only significant with outliers removed ($r = .385$).

The subjective measure of imagery ability was the self-report Questionnaire upon Mental Imagery (QMI), revised by Sheehan (1967). In this questionnaire participants rate how clearly they can imagine items such as the 'smell of an orange' from 1 (a perfectly clear image) to 7 (no image at all). Scores on this measure were not related to overall insight score ($r = .111$) but they did suggest a distinction between verbal and non-verbal problems. Higher scores on the verbal set were reliably associated with lower reported imagery ($r = .346$) but the correlation with the non-verbal problems was in the opposite direction, although not reliably so ($r = -.256$).

In this experiment we also used the Remote Associates Test (RAT), adapted from the version by Bowden & Jung-Beeman (2003), to facilitate a comparison between imagery and verbal ability. This task reflects participants' ability to generate remote associates for words and also the different possible meanings for words by asking participants to find a fourth word that will link three others (see Appendix C).

Greater verbal dexterity was reliably associated with higher scores on the insight problems (r = .327). However, the association between the RAT scores and the verbal problems was high and positive (r = .508) whereas between the RAT and non-verbal problems it was low and negative (r = -.108).

The last item used in Experiment 6 was the Cognitive Failures Questionnaire (CFQ) developed by Broadbent et al (1982). It asks participants to report how often they have experienced everyday lapses such as forgetting why they went from one part of the house to another, which together can be thought of as absent-mindedness. We used this measure in an attempt to gauge whether attention, particularly not focusing attention, was related to insight problem-solving ability. We expected that people who had higher scores on the CFQ (indicating greater self-reported absent-mindedness) would have higher scores on the insight problems. The result confirmed our prediction about the relationship between the two scores, the CFQ had a high and reliable correlation with total insight problem score (r = .446). Furthermore in a stepwise regression, CFQ score alone was the best model for predicting the total insight score and explained 17% of the variance.

This experiment indicated that insight problems draw on imagery and verbal ability but that, again, the individual requirements of each problem will be somewhat different. To perform well on a wide range of insight problems, a person would probably have to be competent in both skills. Perhaps surprisingly, scores on the verbal insight problems were not reliably associated with scores on the nonverbal problems. This dissociation may be due to the non-verbal problems being harder than the verbal. Whatever the reason, the data suggested that neither imagery nor verbal skills would completely account for the variance in insight problem scores.

In Experiment 7, we sought to investigate some aspects of executive functioning in more detail, specifically attention and working memory. We measured three aspects of attention. These were selective attention (Map Search), sustained attention (Sustained Attention Response Task, SART), and attention switching (Visual Elevator, Plus-Minus Task). The Map-Search task requires participants to locate all the symbols for a particular facility on a detailed map. In the SART numbers are presented at speed on a computer screen and the participants must press a button every time they see a number except for the target number when they try to withhold the response. The attention-switching task requires changing quickly from one function to another, either from counting up to counting down (Visual Elevator) or from addition to subtraction (Plus-Minus Task).

The results from the CFQ in Experiment 6 suggested that participants who did well on insight problems would do less well on attention measures that required ignoring distracters, as in the selective attention task. Higher scores on the CFQ have also previously been associated with more errors on the SART (Robertson et al, 1997), so we expected a similar trend with insight problem scores. We expected that good insight problem-solvers would not do well on these measures because if attending to more and potentially distracting information is useful in insight problem-solving then such participants should not perform well on tasks requiring the focusing of attention. In contrast, we hypothesised that high scorers on the insight problems may do better on the measures of attention-switching (Visual Elevator, Plus-Minus). Success on insight problem-solving might involve being able to switch between different alternatives. Research in other areas of reasoning suggests that people do not tend to keep more than one possibility in mind unless it becomes necessary for them to do so, when reasoning

with counterfactuals for example (Johnson-Laird & Byrne, 2002). Similarly, Simon (1981) proposed that rationality in decision-making is bounded by internal constraints (e.g. processing capacity) and external constraints (e.g. time demands) and that in general we try to do as well as possible within these constraints. They are facilitated in this goal by heuristics such as recognition-based inferences and satisficing.

The requirement to hold a number of alternatives in mind, we expected, would also be associated with greater working memory capacity. We therefore included two working memory measures, capacity for storing information alone (Digit Span) and capacity for the storage plus processing of information (Sentence Span). In keeping with the suggestion of Gilhooly and Murphy (2004) that the role of working memory in insight might be more observable if no memory aids were allowed during the problem-solving process, we presented the same problems as used in Experiment 6 but without the availability of pen-and-paper. A subset of four non-insight problems was also included to allow a comparison between the patterns of association for both insight and non-insight problems (a syllogism, a move problem, an algebra equation and another maths type problem, see Appendix C for full details).

The results indicated that the insight and non-insight problems had different relationships with the attention measures. Being able to ignore the distracting information on the selective attention task was relevant for the non-insight problems ($r = .338$) but not the insight problems ($r = .023$). On both of the attention-switching tasks, higher insight problem score was associated with greater accuracy (Visual Elevator: $r = .515$; Plus-Minus: $r = -.511$) while non-insight score was not. However, for the same tasks higher score on the non-insight problems was associated with faster times (Visual Elevator: $r = .353$; Plus-Minus: $r = -.478$) while the insight score was not.

The number of errors on the SART, measuring sustained attention, was not related to scores on either problem set.

Higher scores on both of the working memory measures were associated with better scores on the insight problem-solving tasks, as expected, and on the non-insight problem-solving task as well. The Digit Span task measuring storage capacity alone had a correlation of .405 with the insight problems and .390 with the non-insight problems. The Sentence Span measuring capacity for storage plus processing had a correlation of .396 with insight score and .511 with non-insight score.

The results of the regression analysis with insight score as the dependent variable showed that a model composing a selection of the attention and memory measures could account for 72% of the variance in the insight problem-solving score. The measures included in the model were Digit Span, accuracy scores on the two attention-switching tasks and time taken on the attention-switching, Plus-Minus task (although this last was the least of the contributors). These findings support suggestions that insight problem-solving ability can be explained by a combination of cognitive processes, rather than a special, as yet unidentified, insight ability (e.g. Chronicle et al, 2004). However, as already described the relationship between the attention measures was quite different for the insight score than the non-insight score so how the processes of attention are employed might be quite different. A difference in how information is selected for further processing might distinguish between people who are good at solving insight problems and those who are not. Perhaps the good solvers attend to more information about a problem rather than selecting just the seemingly important information with the result that they are better able to change strategies when failure with one strategy becomes apparent.

## Implications

In this section we will discuss in further detail the implications of our findings. We have identified three main areas where our findings could have important implications: in relation to theories of insight, re-assessing people's ability to solve insight problems and the search for insight. Throughout we also consider the implications for the study of insight in general.

### *Implications for theories of insight*

Our findings support roles for both representation and searching problem space. Our first experiment established that the expected errors for a number of insight problems did actually occur and that failure to solve was strongly related to these errors in representation. The representation view (e.g. Ohlsson, 1992) proposes that the representation of a problem influences what knowledge is accessed. If that representation is flawed then the likelihood of the necessary information being accessed is reduced, making the problem difficult. It follows from this proposal that if the problem is better represented then the relevant knowledge will be accessed and the problem is more likely to be solved. In subsequent experiments, manipulations targeted at improving the representation of problems resulted in improved solution rates for some problems (ping-pong, window-cleaner and six-matches) but not all of them. The nine-dot problem remained particularly resistant to manipulations although extra time did help across both conditions in Experiment 3. Conversely, the hints that the participants suggested in Experiment 2 indicated that the representation of how the finished solution would look could be a factor in the nine-dot and six-matches problems.

To even attempt a problem, participants must have some understanding of what the problem requires. The question then becomes not so much whether there is a role for representation in insight problem-solving but rather whether a mistake in the representation influences the difficulty of the problem. The experiments in Chapter 2 did find evidence supporting the idea that errors can arise in the representation of insight problems and that these errors can be linked to failure to solve the problems.

However, the manipulations seldom raised solution rates to 100% (although this rate did occur for the window-cleaner problem in the props condition of Experiment 3). So even though the results support the idea that errors in representation are influential in problems, it also indicates that some other factor has a role to play. We suggested that the size of the search space could also influence the difficulty of a problem. The size of the search space can affect the difficulty of a problem in several ways, for example if a heuristic or shortcut of some kind is used incorrectly, it may make the problem more difficult. The role of search space was examined in Chapter 2. We investigated whether correct solution rates for insight problems would improve when the number of possibilities participants had to consider was restricted. This kind of manipulation worked better for problems that involved moving pieces such as the horse-and-rider problem. In these problems it was theoretically possible to plot all possible combinations of moves. It has been these kinds of problems that have been most commonly used in the literature on heuristics and move-planning (e.g. Chronicle et al., 2004). The results of search limitations were not as clear-cut for problems where it was not possible to plan every possible combination. For some problems the first step is to find a problem space to search because the problem initially seems impossible, or what initially seems like the appropriate problem-space is not the right

space. For example, in the marrying-man problem the question of how one can legally marry twenty women initially seems a question of how a man can be the groom in twenty different weddings but this is not the case.

We speculate that for some problems the search for a solution is heavily influenced by how the problem and its solution are represented such as the ping-pong ball problem. For other problems the search is dictated by shortcuts that are employed to make the search of a large space more manageable such as the nine-dot problem. However, with these insight problems what seems to be a useful shortcut does not bring the person any closer to a solution. What distinguishes problems for which the influence of the size of space or the representation is greater? The answer may be how clear the misrepresentation is likely to be and hence how credible it seems to the participant. For example, in the nine-dot problem, even if participants assume that lines must stay in the square this representation is incomplete and does not readily lead to a solution, correct or otherwise. It is more a characteristic of a potential solution than a specific solution representation. Hence, evaluating possible initial moves towards a still rather ill-defined goal becomes a more practical way to get started. There may be a certain amount of 'fishing' for a solution path. For the T problem, there is no specific misrepresentation to be adjusted: the problem asks for a 'T' by arranging the four pieces, which is what needs to be done to reach a solution. The difficulty in this problem arises because what looks like a good reference point to start off, filling the 'empty' notch, is not a helpful way to start.

In other problems such as the ping-pong ball problem the misrepresentation of finding a way to get the ball travelling forwards and then backwards is well-defined and credible. There is initially no reason to doubt this representation and so a search for a

solution can commence within this problem space. It is only after failed attempts that doubts may arise about the initial representation. Similarly for the window-cleaner problem, the misrepresentation is that of finding a way a man can fall sixty-feet without injuring himself. The fact that the height of the ladder is specified in the problem suggests that it is relevant and is in keeping with the idea of how a man can fall a great distance and not hurt himself. Hence the participant may at first believe that they have represented the problem correctly and start the search for a solution within this space. For problems where the representation is not so complete and where a sequence of moves is necessary, such as in the nine-dot, the space is often too big and must be narrowed down by other methods. Thus from this distinction of complete (but incorrect) and incomplete representations it may be possible to predict what kind of manipulations are going to be most helpful to participants in solving particular problems: manipulations aimed at improving the representation, or manipulations aimed at improving the processes that direct the search of a large search space.

Such a distinction maps to our distinction between single-step and multiple-step problems. Many, though not all, of the single-step problems are heavily influenced by a flawed representation. The mistaken representation is in keeping with real world experience, a window-cleaner, throwing a ball, a man marrying twenty women. Once the representation is corrected then the search space is rather small because the number of steps between problem and solution is short. The multiple-step problems on the other hand are often more specific and not as related to real events, that is to join dots or move coins, and the search may be directed more by heuristics. The number of steps between problem and solution is much greater because there are so many more possibilities.

If the difficulty of insight problems is dictated by a combination of representation and search than any comprehensive theory will have to include an account of both aspects, otherwise it will not be able to predict performance or the effects of manipulations on all insight problems. Furthermore, there are implications for generalising the results of one insight problem to all others. If the potential influence of narrowing search space and misrepresentation varies from problem to problem then detailing the particular processes with regard to just one problem may not help to fully explain another problem. Single problem studies allow for the identification of specific processes in comparison to multiple problem studies where the emphasis is on identifying common trends. In the future, it may be helpful to combine these approaches to achieve a detailed understanding of the processes involved in a range of problems.

***Implications for assessing performance on insight problems***

In our second and third experiments, we found that relatively simple manipulations that did not negate the requirement for insight could improve people's performance on some problems. For example, in Experiment 3 just the provision of extra time and the freedom to make more than one attempt, boosted performance on the nine-dot problem to a high of 38% from a low of 0% in the previous experiment. Also in Experiment 3, important increases in solution rates were observed for two other problems where physical props were provided, such the six-matches problem, on which performance is typically very low.

Where performance on problems is low, it may be partially due to the fact that the problem-solving context is somewhat alien. Therefore, simulating a more natural

problem-solving environment by providing props where appropriate, allowing extra time and repeated attempts might give a more accurate picture of how individuals might perform in a real world context. In everyday reasoning, people seldom impose a very short time-limit on themselves or determine that they must produce an entire and completely accurate solution on their first attempt. Asking people to think without recourse to aids or other resources may also place an unnatural burden on them. To get a more realistic reflection of how well people can perform with ill-defined problems, it may be necessary to afford them a more ecologically valid environment in which to work.

***Implications for the search for insight***

In recent years the debate concerning insight as a special process has been leaning away from this notion, particularly given the evidence from experiments that have demonstrated the applicability of general heuristics such as hill-climbing to a selection of insight problems (Chronicle et al, 2004). Our findings in Experiment 7 are consistent with this view to the extent that over 70% of the variance in insight problem-solving performance could be explained by performance on measures of working memory and attention. These findings argue against the notion that the ability to solve insight problems is a single identifiable skill.

In the same study, however, we also saw that insight problems did not relate to the attention measures in the same way as the non-insight problems. It may be found in the future, therefore, that although there is no special skill for insight, the way in which the more general processes such as attention and memory are employed by good insight problem-solvers may differ from the way they are used by people who are not.

## Future Work

For many years, the focus of research on insight problems has been on explaining failure. We have tried to identify ways in which performance on insight problems can be improved. It may be time to consider not only why people fail to solve these sorts of problems but also why people are successful. Our work here has gone some way towards this goal. For example we found that clarifying the problem through minimal wording changes can help people achieve the insight necessary to solve some problems. Likewise, basic props can also help people reason about insight problems. Future work may build on these findings to identify interventions that would reliably help people to develop their ability to reason with ill-defined problems. The potential for assisting human thinking exists. Much of everyday human reasoning involves ill-defined problems, from planning a dinner menu to problems of design and engineering, and helping this kind of reasoning to be more effective may be an important goal for future insight problem research.

## Conclusion

Even at this stage of insight problem-solving research, researchers are not in complete agreement as to what exactly constitutes an insight problem. We propose that previous experience of solving problems may lead people to hold certain expectations of problems. Problems that are often labelled as 'insight' typically violate some of these expectations and the criteria that people normally use to judge the appropriateness of a solution strategy do not serve them well. In Chapter 2, we showed that how a problem is represented can lead to difficulties in problems that people should be capable of solving. People may expect that a problem will include all, and only, relevant

information and hence they give undue weight to information such as the height of the ladder in the window-cleaner problem, for example. Similarly, they may expect that relevant information such as the need to use three dimensions in the six matches problem will be made explicit.

In Chapter 3, we showed that making a large search space more manageable helped people to find a solution. Navigating the search space of insight problems may be problematic because heuristics that are normally useful, such as selecting the move that makes most progress towards the goal, are not helpful with these problems. Again, there may be an expectation that moves or options in insight problems can be evaluated in a similar way to previously encountered problems.

The first step to consistently better insight problem-solving might be an awareness that the normal expectations do not always apply. In addition a person would need to be able to recognise when the usual strategies are not working and switch to a consideration of other options. The results described in Chapter 4 implied that more accurate attention-switching and greater working memory capacity, perhaps to hold in mind more alternatives, were useful in insight problem-solving ability.

## REFERENCES


Adeyemo, S. A. (1994). Individual differences in thinking and problem solving. *Personality and Individual Differences*, 17, 1, 117-124.

Ansburg, P. I. (2000). Individual differences in problem solving via insight. *Current Psychology: Developmental, Learning, Personality, Social*,19, 2, 143-146.

Ansburg, P.I. and Dominowski, R.L. (2000). Promoting insightful problem-solving. *Journal of Creative Behaviour*, 34, 1, 30-60.

Ansburg, P.I. and Hill, K. (2003). Creative and analytic thinkers differ in their use of attentional resources. *Personality and Individual Differences*, 34, 1141-1152.

Antonietti, A. and Gioletta, M. A. (1995). Individual differences in analogical problem solving. *Personality and Individual Differences, 18, 5, 611-619*

Asimov, I. (1993). *The Giant Book of Facts and Trivia.* London: Magpie Books

Astur, R. S., Troop, J., Sava, S., Constable, R. T. and Markus, E. J. (2004). Sex differences and correlations in a virtual Morris water task, a virtual radial arm maze, and mental rotation. *Behavioural Brain Research, 152, 1-2, 103-115.*

Blackwell, J. (n.d.) 1933: The genius next door. Retrieved August 26th 2004 from http://www.capitalcentury.com/1933.html

Boomsma, D. I. (1998). Genetic analysis of cognitive failures (CFQ): A study of Dutch adolescent twins and their parents. *European Journal of Personality, 12, 5, 321-330.*

Bowden, E. M. and Jung-Beeman, M. (2003). Normative data for 144 compound remote associate problems. *Behaviour Research Methods, Instruments and Computers, 35, 4, 634-639*

Borella, E., Jouffray, C. and de Ribaupierre, A. (2003, September). *Individual*

*differences in working memory and fluid intelligence: A lifespan perspective.* Poster session presented at the Thirteenth Conference of the European Society of Cognitive Psychology, Granada, Spain.

Broadbent, D. E., Cooper, P. F., Fitzgerald, P. and Parkes, K. R. (1982). The Cognitive Failures Questionnaire and its correlates. *British Journal of Clinical Psychology, 21, 1 –16.*

Capon, A., Handley, S. and Dennis, I. (2003). Working memory and reasoning: An individual differences perspective. *Thinking and Reasoning*, 9, 3, 177-288.

Chan, R. C. K. (2000). Attentional deficits in patients with closed head injury: a further study to the discriminative validity of the test of everyday attention. Brain Injury, 14, 3, 227-236.

Choudhury, N. and Gorman, K. S. (2000). The relationship between sustained attention and cognitive performance in 17-24-month old toddlers. *Infant and child Development, 9, 3, 127-146.*

Chronicle, E. P., Ormerod, T. C. and MacGregor, J.N. (2001). When insight just won't come: The failure of visual cues in the nine-dot problem. *Quarterly Journal of Experimental Psychology*, 54A, 903-919.

Chronicle, E.P., MacGregor, J. N. and Ormerod, T. C. (2004). What makes an insight problem? The role of heuristics, goal conception and solution re-coding in knowledge-lean problems. *Journal of Experimental Psychology: Learning, Memory and Cognition, 30, 1, 14 - 27*

Conway, A. R. A., Kane, M. J. and Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences, 7, 12, 557-552.*

Csikszentmihalyi, M. and Sawyer, K. (1995). Creative insight: The social dimension of a solitary moment. In R.J. Sternberg and J.E. Davidson (Eds.), *The Nature of Insight*. Cambridge, Mass.: MIT Press.

Davidson, J.E. (1995). The suddenness of insight. In R.J. Sternberg and J.E. Davidson (Eds.), *The Nature of Insight*. Cambridge, Mass.: MIT Press.

Dominowski, R.L. and Dallob, P.I. (1995). Insight and problem solving. In R.J. Sternberg and J.E. Davidson (Eds.), *The Nature of Insight.* Cambridge, Mass.: MIT Press.

Dominoski, R. L. (1995). Productive problem solving. In S. M. Smith, T. B. Ward and R. A. Finke (Eds.), *The Creative Cognition Approach*. Cambridge, MA: MIT Press

Dunbar, K. (1995). How scientists really reason: Scientific reasoning in real-world laboratories. In R.J. Sternberg and J.E. Davidson (Eds.), *The Nature of Insight.* Cambridge, Mass.: MIT Press.

Duncker, K. (1945). On problem solving. *Psychological Monographs, 58 (Whole no. 270)*

Eysenck, M. W. and Keane, M. T. (2000). *Cognitive Psychology: A Student's Handbook.* (4th Ed.). Hove and Hillsdale: Erlbaum.

Farrin, L., Hull, L., Unwin, C., Wykes, T. and David, A. (2003). Effects of depressed mood on objective and subjective measures of attention. *Journal of Neuropsychiatry and Clinical Neurosciences, 15, 1, 98-104.*

Finke, R. A., Pinker, S. and Farah, M. J. (1989). Reinterpreting visual patterns in mental imagery. *Cognitive Science, 13, 1, 51-78*

Garavan, H., Ross, T. J., Murphy, K., Roche, R. A. P., and Stein, E. A. (2002).

Dissociable executive functions in the behavioural control: Inhibition, error detection and correction. *NeuroImage, 17, 1820-1829*

Garcia-Madruga, J. A., Gutierrez, F., Carriedo, N., Luzon, J. M. and Vila, J. O. (2004). Working memeory and propositional reasoning: Searching for new working memory tests. (In submission)

Gazzaniga, M. S., Ivry, R.B. & Mangun, G. R. (1998). *Cognitive Neuroscience: The Biology of the Mind.* New York: Norton

Gick, M. L. and Lockhart, R. S. (1995). Cognitive and affective components of insight. In R.J. Sternberg and J.E. Davidson (Eds.), *The Nature of Insight.* Cambridge, Mass.: MIT Press.

Gilhooly, K. and Murphy, P. (2004). Differentiating insight from non-insight problems. *In submission.*

Glucksberg, S. and Weisberg. R. W. (1966). Verbal behaviour and problem solving: Some effects of labelling in a functional fixedness problem. *Journal of Experimental Psychology, 71, 5, 659-664.*

Grant, E. R. and Spivey, M. J. (2003). Eye movements and problem-solving: Guiding attention guides thought. *Psychological Science, 14, 5, 462-466.*

Hatakeyama, T. (1997). Adults and children with high imagery show more pronounced perceptual priming effect. *Perceptual and Motor Skills, 84, 3, 1315-1329.*

Johnson-Laird, P. N. and Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics and inference. *Psychological Review, 109, 4, 1-33*

Jung-Beeman, M., Bowden, E. M., Haberman, J., Frymiare, J. L., Arambel-Liu, S., Greenblatt, R., Reber, P. J. and Kounios, J. (2004). Neural activity when

people solve verbal problems with insight. *PloS Biology, 2, 4, 500-510.* Retrieved from http://www.plosbiology.org/archive/15457885/2/4/pdf/10.1371_journal.pbio.0020097-L.pdf on 5th May, 2004

Kaplan, C. A. and Simon, H. A. (1990). In search of insight. *Cognitive Psychology, 22, 3, 374-419*

Keane, M. T. (1989). Modelling "insight" in practical construction problems. *Irish Journal of Psychology, 11, 202-215*

Kershaw, T.C. and Ohlsson, S. (2001). Training for insight: The case of the nine-dot problem. In J.D. Moore and K. Stenning (Eds.), *Proceedings of the 23rd Annual Conference of the Cognitive Science Society.* Mahwah, NJ: Erlbaum.

Kershaw, T.C. and Ohlsson, S. (2004). Multiple causes of difficulty in insight: The case of the nine-dot problem. *Journal of Experimental Psychology: Learning, Memory and Cognition, 30, 1, 3 – 13*

Kosslyn, S., Brunn, J., Cave, K. and Wallach, R. (1984). Individual differences in mental imagery ability: A computational analysis. *Cognition, 18, 195-243.*

Knoblich, G., Ohlsson, S., Haider, H. and Rheinus, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 25, 6, 1534-1555.

Knoblich, G., Ohlsson, S. & Raney, G. E. (2001). An eye movement study of insight problem-solving. *Memory and Cognition*, 29, 7, 1000-1009.

Knoblich, G. and Wartenburg, F. (1998). Unnoticed hints facilitate representational change in problem solving. *Zeitschrift für Psychologie*, 206, 3, 207-234.

Kwekkeboom, K., Huseby-Moore, K. and Ward, S. (1998). Imaging ability and effective use of guided imagery. *Research in Nursing and Health, 21, 3, 189-*

*198.*

Landry, P. (2002). Adam Smith. Retrieved August 26th from http://www.Bluepete.com/Literature/Biographies/Philosophy/Smith.htm

Lavric, A., Forstmeier, S and Rippon, G. (2000). Differences in working memory involvement in analytical and creative tasks: An ERP study. *NeuroReport*, 11, 8, 1613-1618.

Lockhart R. S., Lamon, M. and Gick, M.L. (1988). Conceptual transfer in simple insight problems. *Memory and Cognition*, 16, 1, 36-44.

Lung, C-T. and Dominowski, R. L. (1985). Effects of strategy instructions and practice on nine-dot problem-solving. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 11, 4, 804-811.

MacGregor J. N., Ormerod T. C. and Chronicle, E. P. (2001). Information processing and insight: A process model of performance on the nine-dot and related problems. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 27, 1, 176-201.

Maier, N. R. F. (1930). Reasoning in humans: I. On direction. *Journal of Comparative Psychology, 10, 115-143*

Malinowski, J. C. (2001). Mental rotation and real-world wayfinding. *Perceptual and Motor Skills, 92, 1, 19-30.*

Markovits, H., Doyon, C. and Simoneau, M. (2002). Individual differences in working memory and conditional reasoning with concrete and abstract content. *Thinking and Reasoning*, 8, 2, 97-164.

Mayer, R. E. (1995). The search for insight: Grappling with Gestalt psychology's unanswered questions. In R.J. Sternberg and J.E. Davidson (Eds.), *The Nature*

*of Insight*. Cambridge, Mass.: MIT Press.

McAvinue, L. and Robertson, I. (2003, April). *Individual differences in visual and motor imagery.* Poster session presented at the Ninth European Workshop on Imagery and Cognition, Pavia, Italy.

Mednick, S. A. (1962). The associative basis of the creative process. Psychological Review, 69, 3, 220-232.

Metcalfe, J. and Wiebe, D. (1987). Intuition in insight and non-insight problem solving. *Memory and Cognition*, 15, 3, 238-246.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H. and Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49-100.

Moores, E. and Andrade, J. (2000). Ability of dyslexic and control teenagers to sustain attention and inhibit responses. *European Journal of Cognitive Psychology, 12, 4, 520 – 540.*

Newell, A. and Simon, H. A. (1972). *Human Problem Solving*. NJ: Prentice-Hall

Newstead, S. E., Handley, S. J., Harley, C., Wright, H. and Farrelly, D. (2004). Individual differences in deductive reasoning. *The Quarterly Journal of Experimental Psychology, 57, 1, 33 - 59*

Novick, L. R. and Sherman, S.J. (2003). On the nature of insight solutions: Evidence form skill differences in anagram solution. *Quarterly Journal of Experimental Psychology, 56A, 2, 351-382*

Ohlsson, S. (1992). Information-processing explanations of insight and related phenomena. In M.T. Keane and K.J. Gilhooly (Eds.), *Advances in the*

*Psychology of Thinking*. London: Harvester-Wheatsheaf.

Okada, H., Matsuoka, K. and Hatakeyama, T. (2000). Dream-recall frequency and waking imagery. *Perceptual and Motor Skills, 91, 3, 759-766.*

Ormerod, T.C., MacGregor, J.N. and Chronicle, E.P. (2002). Dynamics and constraints in insight problem-solving. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 28, 4, 791-799.

Passolunghi, M. C., Cornoldi, C. and De Liberto, S. (1999). Working memory and intrusions of irrelevant information in a group of specific poor problem-solvers. *Memory and Cognition, 27, 5, 779-790*

Perfetto, G.A., Bransford, J.D. and Franks, J.J. (1983). Constraints on access in a problem-solving context. *Memory and Cognition*, 11, 1, 24-31.

Perkins, D. N. (2000). *The Eureka Effect: The Art and Logic of Breakthrough Thinking*. New York; London: W. W. Norton & Co.

Pylyshyn, Z. (2003). Return of the mental image: Are there really pictures in the brain? *Trends in the Cognitive Sciences, 7, 3, 113-118*

Reber, A. S. and Reber, E. (2001). *The Penguin Dictionary of Psychology* (3rd Ed.). London: Penguin Books

Richmond, W. (2003). Design and technology, inventors and inventions: Trevor Bayliss. Retrieved August 20th from http://www.design-technology.info/inventors/page8.htm

Robertson, I. (2002). *The Mind's Eye*. London: Bantam Books

Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T. and Yiend, J. (1997). 'Oops!': Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologist*, 35, 6, 747-758.

Robertson, I. H., Ward, T., Ridgeway, V. and Nimmo-Smith, I. (1994). The Test of Everyday Attention. Bury St Edmunds, England: Thames Valley Test Company

Robertson, S. I. (2001). *Problem Solving*. Hove, UK: Psychology Press.

Scheerer, M. (1963). Problem-solving. *Scientific American*, 208, 118-128.

Schooler, J. W. and Melcher, J. (1995). The ineffability of insight. In S. M. Smith, T. B. Ward and R. A. Finke (Eds.), *The Creative Cognition Approach*. Cambridge, MA: MIT Press

Schooler, J. W., Ohlsson, S. and Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology: General,* 122, 2, 166-183.

Segal, E. (2004). Incubation in insight problem solving. *Creativity Research Journal, 16, 1, 141-148.*

Seifert, C.M., Meyer, D.E., Davidson, N., Patalano, A.L. and Yaniv, I. (1995). Demystification of cognitive insight: Opportunistic assimilation and the prepared-mind perspective. In R.J. Sternberg and J.E. Davidson (Eds.), *The Nature of Insight.* Cambridge, Mass.: MIT Press.

Sheehan, P.W. (1967). A shortened form of Betts' Questionnaire upon Mental Imagery. *Journal of Clinical Psychology, 23, 3, 386-389*

Shepard, R. N. and Metzler, J. (1971). Mental rotation of 3-D objects. *Science, 171 (3972), 701-703*

Shrady, M. (1972). *Moments of insight: the emergence of great ideas in the lives of creative men.* New York; London: Harper and Row.

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

Smith, S. M. (1995). Getting into and out of mental ruts: A theory of fixation, incubation and insight. In R.J. Sternberg and J.E. Davidson (Eds.), *The Nature of Insight*. Cambridge, Mass.: MIT Press.

Sternberg, R. J. and Davidson, J. E. (1989). A four-prong model for intellectual skills development. *Journal of Research and Development in Education, 22, 22-28*

Sternberg, R.J. and Davidson, J.E. (1995). *The Nature of Insight.* Cambridge, Mass.: MIT Press.

Sterr, A. M. (2004). Attention performance in young adults with learning disabilities. *Learning and Individual Differences, 14, 2, 125-133.*

Süß, H-M. Oberauer, K. Wittmann, W. W., Wilhelm, O. and Schulze, R. (2002). Working memory capacity explains reasoning ability – and a little bit more. *Intelligence*, 30, 3, 261-288.

Suzuki, H. and Hiraki, K. (1997). Constraints and their relaxation in the processes of insight. *ETL Technical Report*, TR-97-13.

Suzuki, H., Abe, K., Hiraki, K. and Miyazaki, M. (2001). Cue-readiness in insight problem-solving. In J.D. Moore and K. Stenning (Eds.), *Proceedings of the 23rd Annual Conference of the Cognitive Science Society.* Mahwah, NJ: Erlbaum.

Suzuki, H., Miyazaki, M. and Hiraki, K. (1999). Goal constraints in insight problem-solving. *Proceedings of the Second International Conference on Cognitive Science*, 159-164.

Tsai, L. S. (1987). Influence of various tridimensional clues on solving a tetrahedron problem. *Perceptual and Motor Skills, 65, 2, 672-674*

Vom Hofe, A., Mainemarre, G. and Vannier, L-C. (1998). Sensitivity to everyday failures and cognitive inhibition: Are they related? *European Review of Applied*

*Psychology, 48, 1, 49-55*

Wallace, J. C., Vodanovich, S. J. and Restino, B. M. (2003). Predicting cognitive failures from boredom processes and daytime sleepiness scores: an investigation within military and undergraduate samples. *Personality and Individual Differences, 34, 4, 635-644.*

Weisberg, R.W. (1995). Prolegomena to theories of insight in problem-solving: A taxonomy of problems. In R.J. Sternberg and J.E. Davidson (Eds.), *The Nature of Insight.* Cambridge, Mass.: MIT Press.

Weisberg, R. W. and Alba, J. W. (1981). An examination of the alleged role of "fixation" in the solution of several "insight" problems. *Journal of Experimental Psychology: General, 110, 2, 169-192.*

Wertheimer, M. (1971, originally published 1945). *Productive Thinking.* (Enlarged edition, Michael Wertheimer, ed.). New York; London: Harper Row

Weschler, D. (1981). *Weschler Adult Intelligence Scale – Revised.* New York: Psychological Corporation

Wickelgren, W. A. (1974). *How to solve problems.* San Francisco: Freeman

Wiley, J. (1998). Expertise and mental set: The effects of domain knowledge in creative problem-solving. *Memory and Cognition, 26, 4, 716-730*

Willingham, D. T. (2004). *Cognition: The Thinking Animal* (2nd Ed.). New Jersey: Pearson Prentice Hall

# APPENDIX A

## Instructions for Experiment 1

*[instructions for Represent First shown, order of tasks reversed for Solution First participants]*

Thank you for agreeing to participate in this study. The tasks examine everyday thinking – they are not tests of intelligence, but there are right and wrong answers so please take your time to understand the problems and to choose your answer. Your answers will be treated as confidential and anonymous, and you may withdraw your participation at any time.

You will be given different sorts of problems to solve. *Example: Farmer Giles has 30 sheep in a pen. One day he fails to close the gate properly and all but 9 escape. How many sheep does Farmer Giles have left in his pen? (Answer: 9)*

Each problem will be presented on the screen at the front of the room. You will be given sufficient time to read it, and then it will be removed and it will not be shown again. Please read each problem carefully. Please do not take notes during the presentation of the problem.

**The first task for each problem is to communicate your understanding of it.** This will be recorded on the COLOURED page. For some of the problems you will be asked to explain the problem in writing as if you were giving it to a friend, giving as much detail as possible but without revealing the answer directly. Specific instructions are on the coloured page for each problem. You will have **1 minute** to finish this task.

If the solution occurs to you at this stage, you may record it on the coloured sheet in the space provided.

**Your second task will be to try and solve the problem.** Please write your solution attempts on the WHITE sheet of paper labelled with the problem name. It is important that you show all your rough work. When you reach a solution, please record it in the space provided on the WHITE sheet of paper. You will have **1 minute 30 seconds (1m 30s)** to work on each problem. When the time has elapsed, the experimenter will tell you. Please stop all attempts to solve the problem promptly at that time. If you reach a solution before the time runs out, use the remaining time to rest before the next problem is shown.

You will be given eight problems to solve. Usually they take about 20 – 30 minutes to complete all eight of them. You may ask to take a break at any time. Please do the problems in the order that they are given to you, do not skip any, and do not go back to previous problems to change an answer once you have completed it. If you have any questions, the experimenter will be pleased to answer them.

After you have completed the eight problems, you will be given an explanation of the main aims of the study, and the solutions to the problems will be presented to you. If you know any of the future participants in this study, please do not discuss the problems or solutions with them until after they have participated. Thank you again for your participation.

Summary of instructions:

1. Read the problem carefully as it will only be shown once.

2. On the coloured sheet, follow the instructions to either draw a diagram of the situation OR explain it in writing as if to a friend (1 min).

Turn to the white sheet and attempt to solve the problem (1 m 30 s).

**The problems used in Experiment 1, their solutions, conjectured errors, and allowed reading times.**

***Single-Step Problems***

i. Ping-pong problem. Describe how to throw a ping-pong ball so that it will go a short distance, come to a dead stop, and then reverse itself. You are not allowed to bounce the ball against any object or attach anything to it.

*Solution:* Throw the ball straight up in the air.

*Error:* Represent the ball travelling along the horizontal plane and not the vertical plane.

*Reading Time:* 30 secs

*Source:* Adapted from Ansburg and Dominoski (2000)

ii. Window cleaner problem. A man was washing windows on a high-rise office building when he slipped and fell off a sixty-foot ladder onto the concrete pavement below. Despite having no safety equipment or anything to break his fall, he was not injured in any way. How was this possible?

*Solution:* The man was only at the bottom of the ladder when he fell.

*Error:* Represent the man at the top of the ladder when he fell and not lower down.

Reading Time: 35 secs

*Source:* Adapted from Ansburg and Dominowski (2000)

iii. Camping trip problem. While out on a camping trip, Miss Jones woke one morning and felt something in the pocket of her shorts. It had a head and tail but no legs. When Miss Jones got up, she could feel it move inside her pocket. Miss Jones,

however, was not concerned and allowed the thing to remain in her pocket all day. What was it?

*Solution:* The 'thing' is a coin.

*Error:* Represent the 'thing' as animate and not inanimate.

*Reading Time:* 30 secs

*Source:* Adapted from Ansburg and Dominowski (2000)

iv. Woods walk problem. Two men get lost while walking in the woods. One starts walking northwards, whilst the other heads south. They bump into each other a quarter of an hour later. Presuming that neither man changed direction during that time, explain how such a situation could have arisen.

*Solution:* The men were not walking together at the outset.

*Error:* Represent the men walking together at the outset.

*Reading Time:* 35 secs

***Multiple-Step Problems***

v. Nine-dot problem. In a computer game, there are nine aliens to be captured. The aliens are standing in a formation of three rows of three across with equal spaces between them, in other words a 3 x 3 matrix. To capture an alien, you must walk through him. Show how you can trace a route that goes through all nine aliens using just four connected straight paths. Do not retrace any part of your route or raise your pen from the page once you have started on the first path. You may use simple dots to represent the aliens

*Solution:* The solution requires the extension of some lines beyond the matrix formed

by the dots. The most common solution is roughly arrow-shaped, although other valid solutions with different shapes exist.

*Error:* Represent the lines within the 3 x 3 matrix and not outside it.

*Reading Time:* 65 secs

*Source:* Adapted from Scheerer (1963)

vi. Six matches problem. A teacher challenges her pupils to form four equilateral triangles using just six matches of equal length. They are also told that each complete match must form one complete side of a triangle. How can it be done?

*Solution:* Three-dimensional figure with one triangle forming the base, and the other triangles forming the three sides.

*Error:* Represent the triangles in two dimensions, and not three-dimensions.

*Reading Time:* 40 secs

*Source:* Adapted from Scheerer (1963)

vii. Candle problem. In a room there is a table pressed against a wall. On the table are a candle, a box of tacks and a packet of matches. Your task is to attach the candle to the wall above the table in such a way that the wax from the lighted candle will not drip onto the table or the floor. Sketch or describe how this can be done.

*Solution:* Use the box containing the tacks as a candle-holder. Use some of the tacks to attach it to the wall, light candle and melt some of the wax into the box, secure candle in box.

*Error:* Represent the box as a container for the tacks only, not as a platform for the candle.

*Reading Time:* 45 secs

*Source:* Adapted from Duncker (1945)

viii. Radiation problem. A doctor has a patient with a malignant tumour in her stomach. It is impossible to operate but, unless the tumour is destroyed, the patient will die. There is a kind of ray that, if directed at the tumour at a sufficiently high intensity, will destroy the tumour. Unfortunately, at this intensity the healthy tissue through which the ray passes will also be destroyed. At lower intensities, the ray is harmless to the healthy tissue but will not affect the tumour either. What type of procedure might be used to destroy the tumour with such rays and at the same time avoid destroying the healthy tissue?

*Solution:* Multiple low intensity rays converge at the tumour to form a laser of sufficient intensity at that point only.

*Error:* Represent one high intensity ray at a time getting to tumour, not low intensity rays.

*Reading Time:* 70 secs

*Source:* Adapted from Duncker (1945)

**Instructions for Experiment 2**

Thank you for agreeing to participate in this study. The tasks examine everyday thinking – they are not tests of intelligence, but there are right and wrong answers so please take your time to understand the problems and to choose your answer. Your answers will be treated as confidential and anonymous, and you may withdraw your participation at any time.

In this experiment you will be given four problems to solve. For each problem, you will have **2 minutes** to read the problem and attempt a solution on the given sheet. After 2 minutes has elapsed, this sheet will be taken up and replaced with a second sheet that will give you the solution to the problem and ask you to complete the following task:

"Imagine a friend is a contestant on a TV game show. They are given this problem and allowed to phone one friend for help. You are the friend they call. What hint would you give to help them as much as possible without merely stating the solution, as this would result in the contestant's disqualification?"

You will be given **1 minute** to complete this task. The experimenter will tell you when the minute is up and will take up the solution sheet before giving you your next problem sheet. If you complete either the solution or the task before the time limit is up, please use the time to rest and wait for the experimenter to hand out the next page.

Example: "Farmer Giles has 29 sheep in a pen. One night he leaves the gate open and

all but 9 escape. How many sheep does Farmer Giles have left?" The answer to the problem is 9. A suitable hint would be "The answer is in the problem, no calculations are needed."

In summary, read each problem as it is presented and attempt a solution (2 minutes) then on the next sheet, read the solution and give a hint to a friend (1 additional minute).

During the experiment it is important that you show all your rough work. If you have any questions, please do not hesitate to ask the experimenter.

**The standard wording for the problems used in Experiment 2**

(the improved wording version omitted the words in parentheses and included instead the words in italics)

***Single step problems***

i. Ping pong problem. Describe how to throw a [ping-pong ball] basket-ball so that it will go a short distance, come to a dead stop, and then reverse itself. You are not allowed to bounce the ball against any object or attach anything to it.

ii. Window cleaner problem. A man was washing windows on [a high-rise] an office building when he slipped and fell off a [sixty-foot] ladder onto the concrete pavement below. Despite having no safety equipment or anything to break his fall, he was not injured in any way. How was this possible?

***Multiple step problems***

iii. Nine dot problem. In a computer game, there are nine aliens to be captured. The aliens are standing [Nine dots are arranged] in a formation of three rows of three across with equal spaces between them, in other words a 3 x3 matrix. To capture an alien, you must walk through him. Show how you can trace a route that goes through all nine aliens [Show how you can join all nine dots] using just four connected straight lines. Do not retrace any part of any line or raise your pen from the page once you have started on the first line. You may use simple dots to represent the aliens.

iv. Six matches problem. Given six matches of equal length, show how [they can be used to form] a shape could be formed that is made up of four equilateral triangles.

Each side of every triangle must consist of just one unbroken match.

(NB: Equilateral means that all sides of the triangle are equal in length).

**Instructions for Experiment 3**

Thank you for agreeing to participate in this study. The tasks examine everyday thinking – they are not tests of intelligence, but there are right and wrong answers so please take your time to understand the problems and to choose your answer. Your answers will be treated as confidential and anonymous, and you may withdraw your participation at any time.

You will be asked to solve four problems. For each problem, you will be given the problem written on a sheet.

[*For the pen-and-paper condition*] Read the problem carefully and record all your rough work in the answer booklet while you are considering the problem.

[*For the physical props condition*] Read the problem carefully. You will then be given some items that represent elements of the problem that you can manipulate while considering the problem. You will not be allowed to write anything down on paper while you are working.

The problem page will remain available to you for the duration. If any of the problems are familiar to you, please alert the experimenter.

Your time to reach a solution will be recorded by the experimenter. When you think you have found a solution, inform the experimenter and the clock will be stopped. If your solution is correct, you will move on to the next problem. If, however, you are not correct the experimenter will tell you to keep working and the clock will be started again. You will have a maximum of **10 minutes** to work on each problem. If, after this time, you have not reached a solution, [the materials will be taken away and] the

experimenter will give you a new problem.

You will be given the solutions at the end of the session. If you have any questions the experimenter will be happy to answer them.

**Problems used in Experiment 3**

(including description of materials given to participants in physical props condition shown in italics)

***Single step problems***

i. Ping pong problem. Describe how to throw a ping-pong ball so that it will go a short distance, come to a dead stop, and then reverse itself. You are not allowed to bounce the ball against any object or attach anything to it.

*Instructions: You will be provided with a ping-pong ball*

*Materials given:* Standard white ping-pong ball

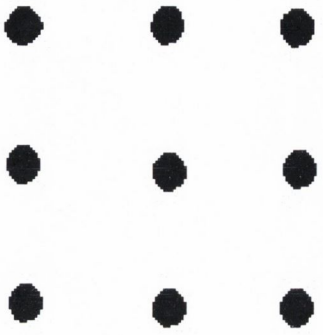
ii. Window cleaner problem. A man was washing windows on a high-rise office building when he slipped and fell off a sixty-foot ladder onto the concrete pavement below. Despite having no safety equipment or anything to break his fall, he was not injured in any way. How was this possible?

*Instructions: You will be given a model-size ladder and figure*

*Materials given:* a model wooden ladder (measuring 595mm long and 40mm wide) and a small figure of a man (measuring 70mm in height).

***Multiple step problems***

iii. Nine dot problem. Nine dots are arranged in a formation of three rows of three across with equal spaces between them, as illustrated below. Show how you can join all nine dots using just four connected straight lines. Do not retrace any part of any line or raise your pen from the page once you have started on the first line.



*Instructions: You will be given a board with pins in it. The nine red pins in the centre represent the nine dots of the puzzle as shown above. The rest of the board represents the rest of the page. Use one length of string to symbolise the four connected lines by wrapping it around the pins as you would draw the lines on the page.*

*Materials given:* corkboard with nine red pins arranged in the matrix formation, some extra pins in a cup, a ball of string and scissors

iv. Six matches problem. Given six rods of equal length and some blue-tack, show how they can be used to form four equilateral triangles. Each side of every triangle must consist of just one unbroken rod.(NB: Equilateral means that all sides of the triangle are equal in length).

*Instructions: You will be given six rods and some blue-tack*

*Materials given:* six wooden rods (measuring 8mm in diameter and 250mm long) and a packet of blue-tack.

# APPENDIX B

**Instructions for Experiment 4 (both conditions)**

This task requires you to solve the problems presented on the following pages. The problems do not test intelligence but there are correct solutions to be found. You will have up to 10 minutes to work on each problem. As soon as you think you know the answer, alert the experimenter and they will tell you if your answer is correct. If the answer is not correct, you can continue to work on it while there is time remaining. Alternatively, if you reach a stage where you think you cannot solve the problem and would like to move on, please inform the experimenter.

The experimenter will time you and may observe your progress. You may ask questions to clarify the task but the experimenter cannot help you any more than that. When you are ready, the experimenter will give you the first problem and any materials necessary. Please refrain from writing during the experiment. The solutions to both problems will be demonstrated at the end. You are free to withdraw your participation at any time.

**Problems and hints used in Experiment 4**

i. The T-puzzle used 4 segments that could be used to form a 'T' shape. The finished T measured 160mm wide and 180mm wide. Each bar of the T was 60mm wide. The puzzle pieces were cut from rigid foam 5mm thick in the manner illustrated in Figure B.1 over the page. The upper side was painted silver to differentiate it from the black underneath. The pieces were handed to participants in a random pile.

*Instructions given to participants:* Arrange the four pieces of the puzzle to form a T-shape as illustrated below.



*(not actual size)*

*Note to participants in experimental group:* Note: The T cut-out is an actual size solution

*Solution:* The solution is provided in Figure B.1.

*Physical hint:* An actual size cardboard cut-out of the finished puzzle that could be used as a template was provided.

*Source:* Adapted from Suzuki & Hiraki (1997)

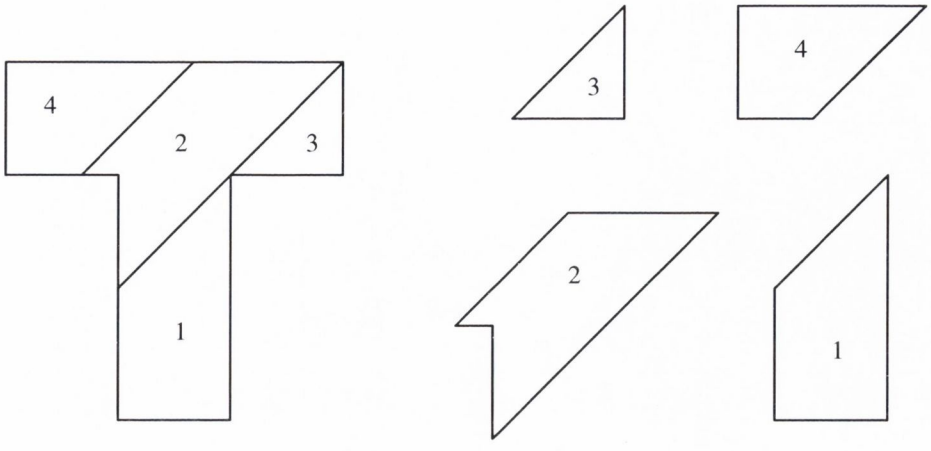



**Figure B.1:** The T-Puzzle segments and solution.

ii. The horse-and-rider puzzles used two light cardboard panels one of which had an illustration of two horses and the other had two riders sitting on part of a horse. The latter can be placed on the former so that the riders are properly astride the horses. The horse panel was 127mm square and the riders panel measured 127mm by 32mm.

*Instructions to participants:* Place panel B on panel A in such a way that both riders are properly astride their horses. (Instruction was accompanied by illustration of panels as well as actual panels to manipulate)

*Note to participants in experimental group:* Use the cut-out sections provided

*Solution:* Rotate the riders panel through 180 degrees as illustrated in Figure B.2.

*Physical hint:* The panels were pinned through the centre so that the riders panel could only be rotated into position over the horse panel.

*Source:* Adapted from Weisberg & Alba (1981)

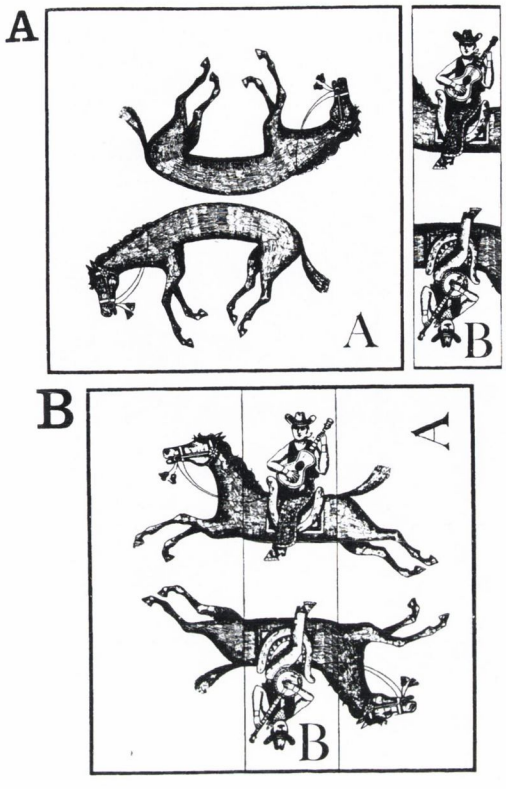



**Figure B.2:** Horse and Rider problem. (A) Panel Presentation and (B) Solution. Shown smaller than actual size.

*Note.* From "An examination of the alleged role of "fixation" in the solution of several "insight" problems by R.W. Weisberg and J. W. Alba, 1981, *Journal of Experimental Psychology: General,* 110, 2, p.185. Copyright 1981 by the APA. Reprinted with permission.

iii. The cheap necklace problem used 12 paper chain-links, divided into four sections of three links, as well as 15 cent worth of 1 and 2 cent coins. The task was to join all the links together to form a chain like a necklace. It costs 2 cents to open a link and 3 cents to close a link and you have a maximum of 15 cent to spend. The paper links could be opened and closed via blue-tack. A representative figure is given in B.3. (NB: This

diagram was not given to participants).

*Instructions to participants:* You are given 4 separate pieces of chain that are each 3 links long. It costs 2 cents to open a link and 3 cents to close a link. All links are closed at the beginning of the problem. Your goal is to connect all 12 links into a single circle (similar to a necklace). The total cost must be no more than 15 cents.

*Note to participants in experimental group:* You can only afford to open and close three links represented by the coloured links in the completed chain.

*Solution:* Open all the links in one section (2c x 3 = 6c). Use the open links as joining links between the sections and close each one (3c x 3 = 9c).

*Physical hint:* A completed necklace-like chain was also given to participants. Each of the three sections consisting of three white links that were joined with a blue link (NB: all the links given to participants to work with were white).

*Source:* Adapted from Weisberg (1995)

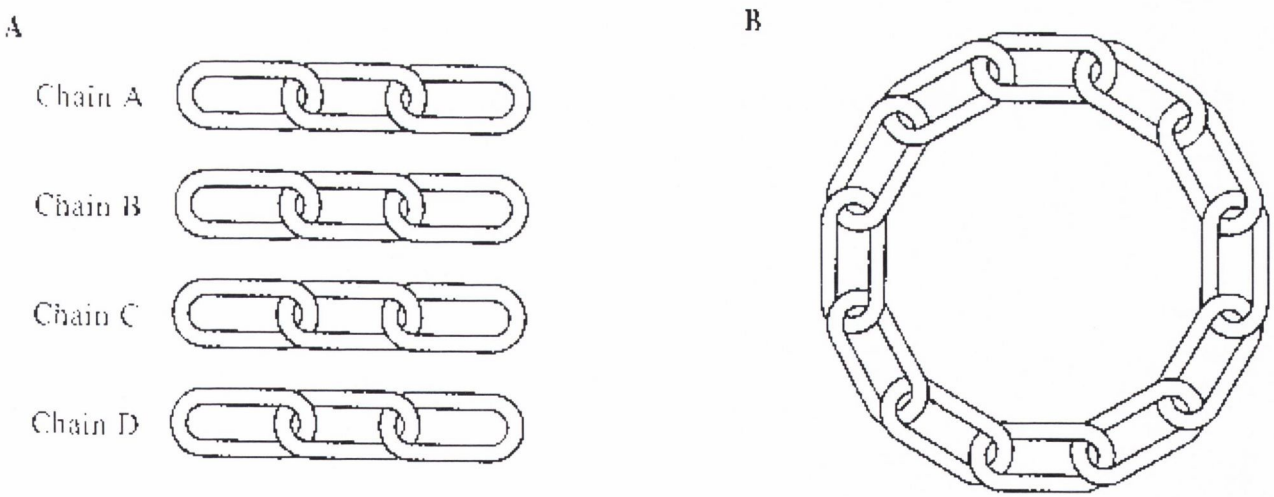



**Figure B.3:** The necklace problem. (A) Starting array (B) Solution

iv. The matchstick equation used matches to form an unbalanced equation reading VII

= V – II. Participants were asked to balance the equation by moving just one match and without taking any away.

*Instructions to participants:* A false equation using roman numerals is constructed of matchsticks. Show how the equation can be made true by moving only one of the matches and without taking any away.

*Note to participants in experimental group:* The match you need to move is one of the coloured ones.

*Solution:* VII – V = II

*Physical hint:* The three matches making up the equals and minus signs were coloured pink and participants were instructed that the key match was one of these matches.

*Source:* Adapted from Knoblich, Ohlsson and Raney (2001)

**Instructions for Experiment 5**

(Hint-After instructions shown below, instructions for other conditions were adjusted accordingly and are shown in italics)

Thank you for agreeing to participate in this study. You will receive five problems. For each problem, you will be shown the problem first and then you will be given some extra information about it. (Hint Before: *You will be given some extra information about the problem first, and then the actual problem*; No-Hint: *You will be shown the problem first and given a chance to read it, then it will be shown to you again and you will have a chance to solve it*). You will be timed on how long you spend reading each screen so press the space bar as soon as you have read and understood the information. All the information will then be shown on one screen and you will be given 3 minutes to think about a solution.

As soon as you know the solution, press the spacebar and say aloud your answer to the experimenter who will record it. You will only be able to offer one solution. We will start with a practice problem.

After you have given your answer, the next screen will tell you the correct solution. You can indicate if you reached the correct solution by ticking one of the options offered to you on that screen.

At each stage you will be instructed which key on the computer you need to press. Please work as speedily and as accurately as you can. If you have any questions, please do not hesitate to ask the experimenter. When you are ready, press the space bar to start the practice problem.

**Problems and hints used in Experiment 5**

i. Ping-pong: Describe how to throw a ping-pong ball so that it will go a short distance, come to a dead stop, and then reverse itself. You are not allowed to bounce the ball against any object or attach anything to it.

*Solution:* Throw the ball straight up in the air

*Hint:* There is a no need to put a spin on the ball.

*Source:* Adapted from Ansburg and Dominoski (2000)

ii. Window-cleaner: A man was washing windows on a high-rise office building when he slipped and fell off a sixty-foot ladder onto the concrete pavement below. Despite having no safety equipment or anything to break his fall, he was not injured in any way. How was this possible?

*Solution:* The man was only at the bottom of the ladder when he fell

*Hint:* The man is not a stuntman.

*Source:* Adapted from Ansburg and Dominoski (2000)

iii. Camping-trip: While out on a camping trip, Miss Jones woke one morning and felt something in the pocket of her shorts. It had a head and tail but no legs. When Miss Jones got up, she could feel it move inside her pocket. Miss Jones, however, was not concerned and allowed the thing to remain in her pocket all day. What was it?

*Solution:* The 'thing' is a coin.

*Hint:* The 'thing' is not alive.

*Source:* Adapted from Ansburg and Dominoski (2000)

iv. Marrying-man: A man who lived in a small town married 20 women. The women

are all still living and he has never divorced, yet he broke no law. Can you explain?

*Solution:* The man is the priest who conducted the ceremony

*Hint:* The man's culture does not permit multiple wives.

*Source:* Adapted from Weisberg (1995)

Practice problem: Farmer Giles had thirty sheep in his field. One day, somebody left the gate into the field open and all but nine sheep escaped. How many sheep did Farmer Giles have left in his field?

*Solution:* Nine sheep are left

*Hint:* Do not subtract 9 from 30

# APPENDIX C

**Insight problems used in Experiments 6 and 7**

*Non-verbal Problems*

i. Horse and Rider (see figure C.1): Panel B can be placed on Panel A below in such a way that both riders appear to be properly astride their horses. Indicate the correct positioning of Panel B on Panel A by [drawing on Panel A below. A simple rectangular outline will be sufficient.] *pointing to the panel below* (wording variation for Experiment 7 in italics)

ii. Ping-pong: Describe how to throw a ping-pong ball so that it will go a short distance, come to a dead stop, and then reverse itself. You are not allowed to bounce the ball against any object or attach anything to it.

iii. Matchstick Equation: A false equation using roman numerals is constructed of matchsticks as illustrated below. Show how the equation can be made true by moving only one of the matches and without taking any away. (see Figure C.2)

iv. Steel Pyramid: A giant inverted steel pyramid is perfectly balanced on its point. Any movement of the pyramid will cause it to topple over. Underneath the pyramid is a 100 euro note. How would you remove the note without disturbing the pyramid?

*Verbal Problems*

v. Window-cleaner: A man was washing windows on a high-rise office building when he slipped and fell off a sixty-foot ladder onto the concrete pavement

below. Despite having no safety equipment or anything to break his fall, he was not injured in any way. How was this possible?

vi. Camping-trip: While out on a camping trip, Ann woke one morning and felt something in the pocket of her shorts. It had a head and tail but no legs. When Ann got up, she could feel it move inside her pocket. Ann, however, was not concerned and allowed the thing to remain in her pocket all day. What was it?

vii. Marrying-man: A man who lived in a small town married 20 women. The women are all still living and he has never divorced, yet he broke no law. Can you explain?

viii. Walk-in-woods: Two men get lost while walking in the woods. One starts walking northwards, whilst the other heads south. They bump into each other a quarter of an hour later. Presuming that neither man changed direction during that time, explain how such a situation could have arisen.

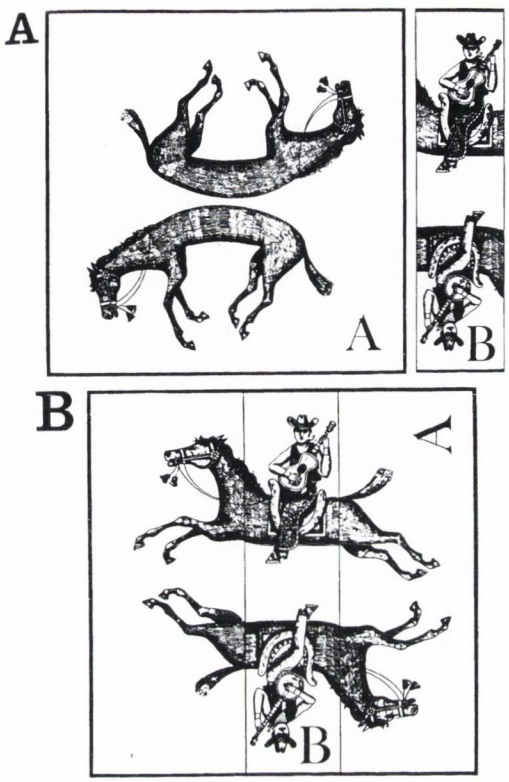



**Figure C.1:** Horse and Rider problem. (A) Panel Presentation and (B) Solution. Adapted from Weisberg and Alba (1981). Shown smaller than actual size.

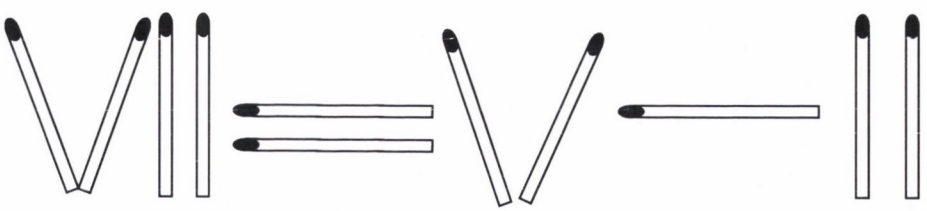


**Figure C.2:** Illustration used for matchstick equation problem. Shown slightly smaller than actual size.

**Non-insight problems used in Experiment 7**

*i. Syllogism*

Complete the conclusion:

All men wear skirts.

Some men are farmers.

Therefore, _____

*ii. Move problem*

There are four coins – two heavier coins of equal weight and two lighter coins of equal weight. The coins cannot be distinguished by looking at them or holding them. How can you tell which coins are which using just two weighings on a balance scale?

*iii. Algebra problem*

Find the value for X:

$^1/_5 X + 10 = 25$

*iv. Arithmetic problem*

What number gives the same result when it is added to 1.5 as when it is multiplied by 1.5?

**Instructions for Insight Problems in Experiment 6**

In the following booklet there are eight problems to be solved. Each problem is presented on a separate page. The problems are similar in form to the example given below.

*Problem: Farmer Giles had thirty sheep in his field. One day, somebody left the gate into the field open and all but nine sheep escaped. How many sheep did Farmer Giles have left in his field?* The solution to this problem is that *nine sheep are left.*

You will have 2 minutes to work on each problem and record your answer on the problem sheet. The experimenter will time each problem and tell you when to move on. Please do the problems in the order they are presented to you and do not skip ahead to later problems or return to any that you have already attempted. If you solve a problem before the time limit has expired please wait until the time has elapsed before moving on to the next problem.

If you have seen any of the problems or their solutions before it is important that you indicate this by ticking the relevant boxes on the sheet at the back of the booklet.

If you have any questions, please do not hesitate to ask the experimenter

**Instructions for Problems in Experiment 7**

You will receive 12 problems in total, one per page, with a short break after the first eight. When you think you know the solution to a problem, call out your answer to the experimenter who will record it. You will have two minutes per problem to think about a solution but you are only allowed one attempt. You are not allowed to right anything while considering the problem.

If you have seen any of the problems or their solutions before it is important that you indicate this to the experimenter.

If you have any questions, please do not hesitate to ask.

**Other Measures**

***In Experiment 6, we used:***

- The Mental Rotations Task (Note: The author has given permission for the inclusion of only the practice sheets in this work. These are given on the following pages)
- The Cognitive Failures Questionnaire
- The Remote Associates Test
- The Questionnaire upon Mental Imagery

***In Experiment 7, we used:***

- Visual Elevator
- Digit Span
- Plus-Minus Task
- Map Search
- Sentence Span
- Sustained Attention Response Task

Mental Rotations Test – Version A – Practice Sheet 1
*(Note: for illustration purposes only, this test is copyrighted)*

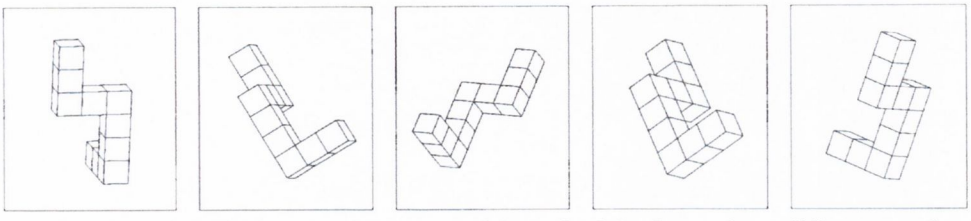


# MENTAL ROTATIONS TEST (MRT-A)

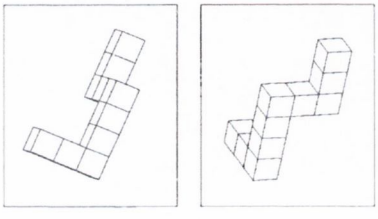
This test is composed of the figures provided by Shepard and Metzler (1978), and is, essentially, an Autocad-redrawn version of the Vandenberg & Kuse MRT test.

©Michael Peters, PhD, July 1995

Please look at these five figures



Note that these are all pictures of the same object which is shown from different angles. Try to imagine moving the object (or yourself with respect to the object), as you look from one drawing to the next.

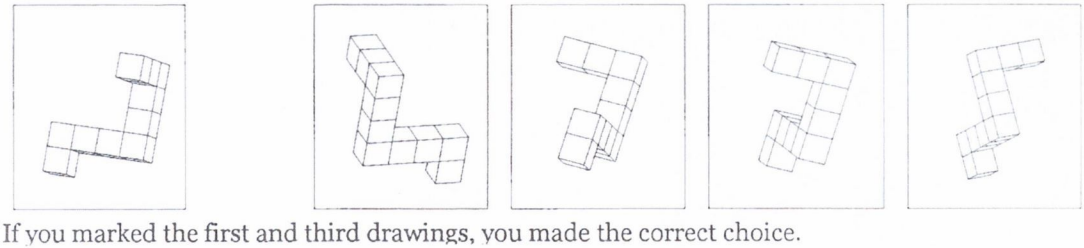


Here are two drawings of a new figure that is different from the one shown in the first 5 drawings. Satisfy yourself that these two drawings show an object that is different and cannot be "rotated" to be identical with the object shown in the first five drawings.

Now look at this object:

Two of these four drawings show the same object. Can you find those two? Put a big X across them.

1.

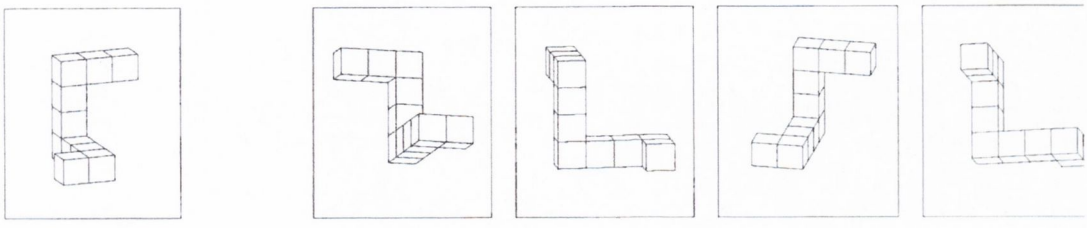


If you marked the first and third drawings, you made the correct choice.

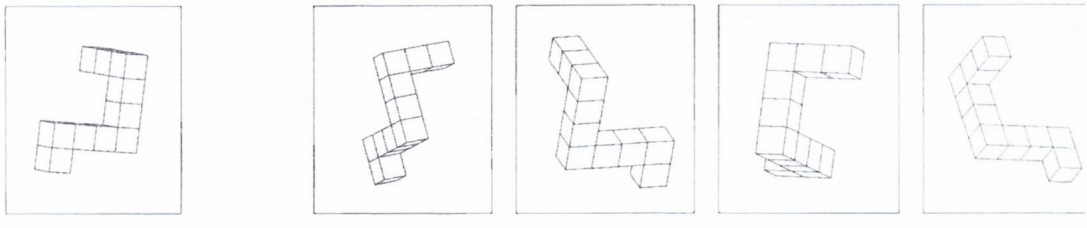
## Mental Rotations Test – Version A – Practice Sheet 2

Here are three more problems. Again, the target object is shown twice in each set of four alternatives from which you choose the correct ones.

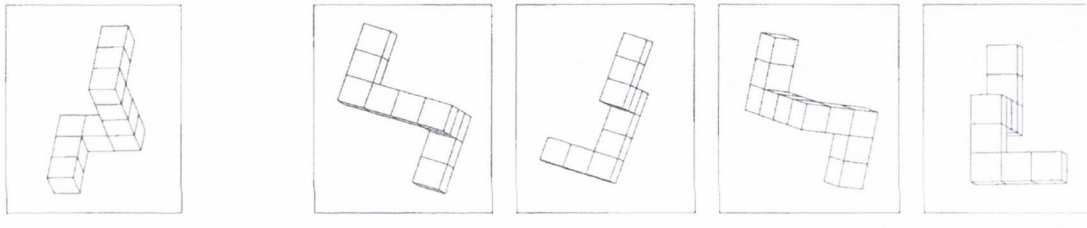
2.a



3.a



4.a



Correct Choice:

2: second and third
3: first and fourth
4: first and third

When you do the test, please remember that for each problem set there are two and only two figures that match the target figure.

You will only be given a point if you mark off both correct matching figures, marking off only on these will result in no marks

**The Cognitive Failures Questionnaire (Broadbent et al. 1982) as used in Experiment 6**

The following questions are about minor mistakes which everyone makes from time to time, but some of which happen more often than others. We want to know how often these things have happened to you in the last six months. Please circle the appropriate number.

| | Very often | Quite often | Occasio-nally | Very rarely | Never |
|---|---|---|---|---|---|
| Do you read something and find you haven't been thinking about it and must read it again? | 4 | 3 | 2 | 1 | 0 |
| Do you find you forget why you went from one part of the house to the other? | 4 | 3 | 2 | 1 | 0 |
| Do you fail to notice signposts on the road? | 4 | 3 | 2 | 1 | 0 |
| Do you find you confuse right and left when giving directions? | 4 | 3 | 2 | 1 | 0 |
| Do you bump into people? | 4 | 3 | 2 | 1 | 0 |
| Do you find you forget whether you've turned off a light or a fire or locked the door? | 4 | 3 | 2 | 1 | 0 |
| Do you fail to listen to people's names when you are meeting them? | 4 | 3 | 2 | 1 | 0 |
| Do you say something and realise afterwards that it might be taken as insulting? | 4 | 3 | 2 | 1 | 0 |
| Do you fail to hear people speaking to you when you are doing something else? | 4 | 3 | 2 | 1 | 0 |
| Do you lose your temper and regret it? | 4 | 3 | 2 | 1 | 0 |
| Do you leave important letters unanswered for days? | 4 | 3 | 2 | 1 | 0 |
| Do you find you forget which way to turn on a road you know well but rarely use? | 4 | 3 | 2 | 1 | 0 |
| Do you fail to see what you want in a supermarket (although it's there)? | 4 | 3 | 2 | 1 | 0 |
| Do you find yourself suddenly wondering whether you've used a word correctly? | 4 | 3 | 2 | 1 | 0 |

| | | | | | |
|---|---|---|---|---|---|
| Do you have trouble making up your mind? | 4 | 3 | 2 | 1 | 0 |
| Do you find you forget appointments? | 4 | 3 | 2 | 1 | 0 |
| Do you forget where you put something like a newspaper or a book? | 4 | 3 | 2 | 1 | 0 |
| Do you find you accidentally throw away the thing you want and keep what you meant to throw away as in the example of throwing away the matchbox and putting the used match in your pocket? | 4 | 3 | 2 | 1 | 0 |
| Do you daydream when you ought to be listening to something? | 4 | 3 | 2 | 1 | 0 |
| Do you find you forget people's names? | 4 | 3 | 2 | 1 | 0 |
| Do you start doing one thing at home and get distracted into doing something else (unintentionally)? | 4 | 3 | 2 | 1 | 0 |
| Do you find you can't quite remember something although it's 'on the tip of your tongue'? | 4 | 3 | 2 | 1 | 0 |
| Do you find you forget what you came to the shops to buy? | 4 | 3 | 2 | 1 | 0 |
| Do you drop things? | 4 | 3 | 2 | 1 | 0 |
| Do you find you can't think of anything to say? | 4 | 3 | 2 | 1 | 0 |

**Remote Associates Test (adapted from Bowden & Beeman, 2003)**

***Instructions***

For each set of three words in this booklet, a fourth word can be used with each to form a new common compound word or phrase. For example, in the set *age/mile/sand*, the word STONE can be used with each to form the three new compound words or phrases, STONE–age, mile-STONE and sand-STONE.

Have a go at this practice set presented below. Try and find a fourth word that will combine with these three to form three new words or phrases. Give yourself about 15 seconds to think about it.

Practice set: loser/throat/spot
Solution:

For each of the following 24 sets, find the fourth word that can be used to form three new common compound words or phrases with that set. Try not to spend too long at any one problem. You have 6 minutes to complete as many sets as you can.

If you have any questions, please ask the experimenter.

**Test Items (nb: original items were presented four per page without solutions)**

night/wrist/stop *solution: watch*
aid/rubber/wagon *solution: band*
cracker/fly/fighter *solution: fire*
safety/cushion/point *solution: pin*
measure/worm/video *solution: tape*
dream/break/light *solution: day*
fish/mine/rush *solution: gold*
river/note/account *solution: bank*
print/berry/bird *solution: blue*
date/alley/fold *solution: blind*

hound/pressure/shot *solution: blood*

right/cat/carbon *solution: copy*

cross/rain/tie *solution: bow*

chamber/mask/natural *solution: gas*

tank/hill/secret *solution: top*

pine/crab/sauce *solution: apple*

cover/arm/wear *solution: under*

change/circuit/cake *solution: short*

man/glue/star *solution: super*

illness/bus/computer *solution: terminal*

type/ghost/screen *solution: writer*

note/chain/master *solution: key*

end/line/lock *solution: dead*

reading/service/stick *solution: lip*

**The Questionnaire upon Mental Imagery (Sheehan,1967) used in Experiment 6**

The aim of this questionnaire is to determine the vividness of your imagery. Each of the items will bring certain images to mind. You are to rate the vividness of each image by reference to the rating scale given below. For example, if your image is "vague and dim" you would give it a rating of 5. Record your answer in the brackets provided after each item. Before you turn to the items on the next page, familiarise yourself with the different categories on the rating scale. A copy of this scale is printed on each page. Please do not turn over a page until you have completed each item, in the order given, on the page you are doing. Do not turn back to check on other items you have done. Try to do each item separately – independent of how you may have done other items.

The image aroused by an item may be:

| | |
|---|---|
| Perfectly clear and as vivid as the actual experience | .....Rating 1 |
| Very clear and comparable in vividness to the actual experience | .....Rating 2 |
| Moderately clear and vivid | .....Rating 3 |
| Not clear or vivid, but recognisable | .....Rating 4 |
| Vague and dim | .....Rating 5 |
| So vague and dim as to be hardly discernible | .....Rating 6 |
| No image present at all, you only "know" that you are thinking of the object | .....Rating 7 |

---

An example of an item on this questionnaire might be one which asked you to consider an image which comes to your 'mind's eye', of a red apple. If your visual image is "moderately clear and vivid" you would check the rating scale and mark 3 in the brackets as follows:

| **Item** | **Rating** |
|---|---|
| Seeing a red apple | (3) |

When you have understood these instructions, turn to the next page and begin.

| **Imagine** | **Rating** |
|---|---|
| 1. Feeling the warmth of a moderately hot bath. | ( ) |
| 2. Seeing, for a relative or friend, the different colours worn in some familiar clothes. | ( ) |
| 3. The sensation of fatigue. | ( ) |
| 4. Smelling an ill-ventilated room. | ( ) |
| 5. Seeing, for a relative or friend, the characteristic poses of head, attitudes of body, etc. | ( ) |
| 6. Tasting granulated (white) sugar. | ( ) |
| 7. Performing the act of running upstairs. | ( ) |
| 8. The sensation of a sore throat. | ( ) |
| 9. Hearing an ambulance siren. | ( ) |
| 10. The sensation of drowsiness. | ( ) |
| 11. Feeling sand. | ( ) |
| 12. Seeing, for a relative or friend, the exact contour of face, head, shoulders and body. | ( ) |
| 13. Feeling linen. | ( ) |
| 14. Tasting jelly. | ( ) |
| 15. Hearing the sound of hands clapping in applause. | ( ) |
| 16. Tasting salt. | ( ) |
| 17. Smelling the scent of a rose. | ( ) |

---

The image aroused by an item may be:

| | |
|---|---|
| Perfectly clear and as vivid as the actual experience | .....Rating 1 |
| Very clear and comparable in vividness to the actual experience | .....Rating 2 |
| Moderately clear and vivid | .....Rating 3 |
| Not clear or vivid, but recognisable | .....Rating 4 |
| Vague and dim | .....Rating 5 |
| So vague and dim as to be hardly discernible | .....Rating 6 |
| No image present at all, you only "know" that you are thinking of the object | .....Rating 7 |

| **Imagine** | **Rating** |
|---|---|
| 18. Hearing the mewing of a cat. | ( ) |
| 19. Smelling fresh paint. | ( ) |
| 20. Seeing, for a relative or friend, the precise carriage, length of step, etc., in walking. | ( ) |
| 21. Hearing the sound of a car horn. | ( ) |
| 22. Performing the act of springing across a gutter. | ( ) |
| 23. Feeling the prick of a pin. | ( ) |
| 24. Smelling cooking cabbage. | ( ) |
| 25. Performing the act of drawing a circle on paper. | ( ) |
| 26. Smelling roast beef. | ( ) |
| 27. Performing the act of reaching up to a high shelf. | ( ) |
| 28. Tasting oranges. | ( ) |
| 29. The sensation of hunger. | ( ) |
| 30. Performing the act of kicking something out of the way. | ( ) |
| 31. Hearing the sound of children singing. | ( ) |
| 32. Tasting your favourite soup. | ( ) |
| 33. Feeling fur. | ( ) |
| 34. Seeing the colour and shine of silverware. | ( ) |
| 35. The sensation of repletion, as from a full meal. | ( ) |

---

The image aroused by an item may be:

| | |
|---|---|
| Perfectly clear and as vivid as the actual experience | .....Rating 1 |
| Very clear and comparable in vividness to the actual experience | .....Rating 2 |
| Moderately clear and vivid | .....Rating 3 |
| Not clear or vivid, but recognisable | .....Rating 4 |
| Vague and dim | .....Rating 5 |
| So vague and dim as to be hardly discernible | .....Rating 6 |
| No image present at all, you only "know" that you are thinking of the object | .....Rating 7 |

**Instructions for Visual Elevator Task as per Test of Everyday Attention Manual (Robertson et al, 1997) as used in Experiment 7**

"Say

*Try to imagine that during your trip, you decide to stay in a large hotel, many stories high. While you are staying there, you find that the indicator in the elevator that tells you what floor you are on is not working properly.*

Show the subject the first visual elevator example page.

Say

*Look at this series of pictures. As you can see, each one shows an elevator. Every so often there is a large arrow, like this one. An arrow pointing down means that the elevator is going down, so you need to reverse count. An arrow pointing up means that the elevator is going up. What I want you to do is count out the floors. Say 'up' and 'down' when you come to the large arrows, as this avoids counting them. I will point to each one in turn as you say the number. Remember the big arrows are not floors, they only tell you which way the elevator is going. So, in this example, you would say – one – two – down – one – up – two. Now you try.*

Repeat as often as necessary until the person has comprehended the task. Do not proceed until you are sure that the subject has performed both practice items correctly on his or her own.

Say

*Now try and do the same with the next set of pictures. Work as quickly and accurately as you can. Count out loud as you move along the elevators."*

**Instructions for Digit Span used in Experiment 7 (taken from WAIS-R, 1981)**

*Forwards*

I will call out sequences of digits and I would like you to repeat them back to me in the order I gave them to you. For example, if I said 9-7-1, you would say:

*Backwards*

Now I will call out more sequences of digits but this time I want you to repeat them back to me in the reverse order to that I gave you. For example, if I said 9-7, you would say:

**Plus-Minus Task used in Experiment 7**

| +3 | -3 | +3,-3 |
|---|---|---|
| 93 | 94 | 38 |
| 95 | 49 | 84 |
| 82 | 55 | 65 |
| 28 | 58 | 68 |
| 43 | 98 | 78 |
| 31 | 47 | 71 |
| 40 | 72 | 41 |
| 85 | 60 | 74 |
| 77 | 24 | 21 |
| 46 | 89 | 27 |
| 23 | 81 | 19 |
| 15 | 61 | 34 |
| 42 | 18 | 17 |
| 11 | 45 | 32 |
| 83 | 70 | 91 |
| 36 | 73 | 59 |
| 67 | 16 | 54 |
| 79 | 99 | 56 |
| 64 | 29 | 50 |
| 52 | 39 | 92 |
| 44 | 26 | 88 |
| 12 | 80 | 33 |
| 20 | 37 | 57 |
| 75 | 66 | 97 |
| 86 | 62 | 96 |
| 48 | 69 | 63 |
| 87 | 14 | 90 |
| 51 | 30 | 25 |
| 10 | 76 | 53 |
| 22 | 35 | 13 |
| Time | | |
| Cost | | |

**Instructions for Map Search Task as per Test of Everyday Attention Manual (Robertson et al, 1997) as used in Experiment 7**

"Say

*The symbol here (show knife-and-fork symbol from cuebook) shows where restaurants can be found in the Philadelphia area. There are many symbols like this on the map.*

Point to one at left side of map. Also, indicate to subjects that the symbols are found all over the map, left and right, top and bottom. Check that the subject can see the symbol clearly. Turn the map over so the subject cannot scan it while you give further instructions. The cuebook is left open at the symbol in front of the subject during the test, to remind him or her of the symbol being sought.

Say

*Let's say you are with a family member or friend. They are driving while you are navigating. You want to know where restaurants are located in case you decide to stop for a meal. What I would like you to do is to look at the map for two minutes and circle as many symbols as you can. I will stop you once a minute has gone by to ask you to stop pens. OK?*

When the subject indicates that they have understood (reiterate the instructions if they have not) turn the map over to reveal the symbols, give them a red pen and begin timing. After one minute, ask the subject to change pens and hand them a blue pen. At the end of two minutes ask the subject to stop.

If the subject feels that they have completed the task before the two minute time

limit, or if they assume that they have done so by reaching the right hand edge of the map, ask them to continue searching for any symbols which they might have missed until the end of the time limit."

**Sentence Span Task used in Experiment 7**

**Instructions**

For the sentences that follow, you must decide whether they are true or false and indicate this by pressing the relevant button on the keyboard. Each sentence will be displayed for 3 seconds, if you have not responded in this time a brief prompt will appear before the next sentence is shown. Sentences will be shown in two blocks of three, four etc.

In addition to making a true-false decision, you will also need to remember the last word in each sentence. At the end of each block of sentences, you will be asked to recall all the last words in the order in which they appeared. We will start with a practice trial of two sentences.

If you have any questions, please ask the experimenter.

***Sentences used***

**Practice Sentences**

A giraffe has a long neck

Newspapers are carved in stone

**Test Sentences**

Some cars use a petrol engine

There are four sides on a square

Snow will only fall in hot weather

Money grows on a money tree

Dublin is Ireland's capital city

Beethoven wrote classical music

Beer is a type of drink

Milk is a toxic gas

Snakes have a long nose

A razor is a kitchen tool

Wordsworth was a poet

Australia is an island

All rocks are orange in colour

Books can be printed on paper

Elizabeth is the English queen

Geography is a school subject

A computer is a living creature

Televisions are powered by water

CD stands for compact disc

A grocer shop sells fruit

Margaret is the name of a girl

A sledge is a kind of boat

Norway is on the African continent

Goblins are a type of animal

The Beatles were a famous pop group

Einstein was a genius

You can swim in a swimming pool

The moon is made of cheese

There are six fingers on a hand

A vampire sleeps in a coffin

A vegetarian eats meat

You can borrow books from a library

The letter k is a vowel

You can sit on a chair

Mice live in the sea

A skyscraper is a low building

**Instructions for Sustained Attention Response Task (SART) used in Experiment 7**

Press the spacebar every time you see a number EXCEPT when the number 3 appears. When this happens make no response. We will start with a practice trial.

Participants were also told that a symbol would appear after each number and that there was still to time to make a response at this stage if they had not had time while the number was on screen, although this was what they should aim for.

## PUBLICATIONS

Murray, M. A. and Byrne, R. M. J. (2004). Improving performance on insight problems. In L. McGinty and B. Crean (Eds.) *Proceedings of the 15th Artificial Intelligence and Cognitive Science Conference*. Ireland: Artificial Intelligence Association of Ireland