THINKING & REASONING, 2005, 11 (3), 279–302

Psychology Press
Taylor & Francis Group

Differentiating insight from non-insight problems

K. J. Gilhooly
University of Hertfordshire, Hatfield, UK

P. Murphy
Brunel University, Middlesex, UK

This study aimed to investigate whether a range of tasks that have been generally classed as requiring insight form an empirically separable group of tasks distinct from tasks generally classed as non-insight. In this study, 24 insight tasks, 10 non-insight tasks, and tests of individual differences in cognitive abilities and working memory were administered to 60 participants. Cluster analysis of the problem-solving tasks indicated that the presumed insight problems did tend to cluster with other presumed insight problems, and similarly the presumed non-insight problems tended to cluster with other presumed non-insight tasks. Performance on presumed insight problems was particularly linked to measures of ideational flexibility with a different pattern of results for the non-insight tasks. Spatial insight problems were linked to spatial flexibility and verbal insight tasks were linked to vocabulary scores. The results are discussed in relation to recent developments of dual process theories of thinking.

This study aimed to investigate whether and in what ways a range of tasks that have been classed in the literature as requiring insight form an empirically separable group of tasks distinct from tasks generally classed as non-insight.

First, some key terms will be defined. A *problem* exists when someone has a goal for which they are unable to generate a suitable sequence of actions either from memory or by applying a routine method. To solve a problem requires *representing* the problem situation and goal, followed by *search* for an appropriate sequence of actions within the framework of the initial problem representation. Sometimes success is achieved within the initial representation. In other cases, the initial representation leads to an *impasse*

Correspondence should be addressed to K. J. Gilhooly, School of Psychology, University of Hertfordshire, Hatfield AL10 9AB, UK. Email: k.j.gilhooly@herts.ac.uk

This work was supported by a grant from the Leverhulme Trust (Grant no. F/00 275/A) held by KJG at Brunel University.

http://www.tandf.co.uk/journals/pp/13546783.html DOI: 10.1080/13546780442000187

(Ohlsson, 1992) in which progress halts; to break the impasse, the problem representation must be changed or *restructured* to allow fresh directions of search. A restructuring that leads to a rapid and complete understanding of how the solution can be reached is often referred to as an *insight*. Phenomenologically, insight is accompanied by an ‘‘Aha’’ experience. In an *a priori* taxonomy of problems, Weisberg (1995) stressed restructuring as a criterion for classing problems as insight or not. He pointed out that a range of problems may be solved either by insight or by trial-and-error search and such problems should be classed as ‘‘hybrid’’. These may be contrasted with ‘‘pure insight’’ tasks, which require restructuring, and ‘‘non-insight’’ tasks in which no restructuring occurs.

Insight has been studied in laboratory conditions by presenting participants with problems that initially induce misleading problem representations within which solution is impossible. In such cases, restructuring of the initial representation is crucial (Durso, Rea, & Dayton, 1994; Ohlsson, 1992; Weisberg, 1995). A classic laboratory example is provided by the *Matchsticks problem*: ‘‘Given six matches, make four equilateral triangles, with one complete match making up the side of each triangle.’’ Participants nearly always adopt an over-restricted representation of the goal and confine their attempts to two-dimensional patterns; however the problem cannot be solved unless the matches are used in three dimensions to form a triangular-based pyramid. Thus with the typically derived initial problem representation the goal cannot be reached; with the restructured goal representation, in which matches may be used in three dimensions, solution is possible. A further example is given by the *Marriage problem*: ‘‘A man in a small town married 20 different women of the same town. All are still living and he never divorced. Polygamy is unlawful but he has broken no law. How can this be?’’ (Dominowski, 1994). Participants find this difficult until they re-interpret ‘‘marrying’’ to mean ‘‘causes someone to be married to another’’ and can conclude that the man is authorised to conduct marriage ceremonies.

PREVIOUS APPROACHES CONTRASTING INSIGHT VS NON-INSIGHT TASKS

Research on insight problem solving has a long history (e.g., Duncker, 1945; Jung-Beeman et al., 2004; Knoblich, Ohlsson, Haider, & Rhenius, 1999; Ormerod, MacGregor, & Chronicle, 2002; Wertheimer, 1945; see also, Davidson, 2003, for a recent review) and a large number of different problems have been used. Weisberg (1995) listed 24 presumed insight problems that have been used widely and many more could readily be added. Given the variety of problems that have been labelled as requiring insight, it is important to establish empirically whether these different problems do indeed form a group (which is separate from non-insight problems) and whether subgroups can be established.

Some authors have cast doubt on the existence of insight, in the sense of complete, rapid understanding following restructuring, in problem solving. For example, Weisberg (1986) proposed that problem solving, even in supposedly insight tasks, involved sequences of small steps only rather than occasional major steps (such as restructuring). In Perkins's (1981) study a majority of participants reported piecemeal stepwise solution of a single presumed insight problem. More recently, MacGregor, Ormerod, and Chronicle (2001), although not denying the subjective "Aha" experience aspect of insight, have argued that the nine-dot problem, which has often been taken to be a classic example of an insight task, is tackled by the same kinds of limited look-ahead strategies, such as hill climbing, as are found in presumed non-insight problems, such as the Hobbits and Orcs task (Thomas, 1974) and that the concept of restructuring is not clear or helpful. Subsequently, Chronicle, MacGregor, and Ormerod (2004) have argued that post-solution recoding of solutions reached by search may account for the phenomenology of insight, which suggests that the insight experience and any related restructuring *follows* the solution rather than preceding or causing solution.

A number of studies have focused on differences between insight and non-insight problems. Metcalfe and Weibe (1987) contrasted five insight and five non-insight problems and found that feeling of warmth judgements (i.e., how close participants felt to solution) behaved differently for presumed insight as against non-insight tasks. With non-insight problems, feelings of warmth predicted solution imminence, but with insight problems, feelings of warmth were unrelated to solution. Schooler, Ohlsson, and Brookes (1993), in a study contrasting three insight and four non-insight problems, also found a separation between insight and non-insight problems in that concurrent thinking-aloud verbalisation interfered with the former but not with the latter. This result was interpreted as suggesting that insight problems normally involved non-verbal processes, which were overshadowed by concurrent verbalisation.

Jung-Beeman et al. (2004) carried out functional magnetic resonance imaging (FMRI) and electroencephalogram (EEG) studies of people solving Remote Associates Test (RAT) items (Mednick, 1962). In these problems, participants are presented with three problem words (e.g., *pine, crab, sauce*) and attempt to produce a single solution word (e.g., *apple*) that can form a familiar combination with each of the three problem words (e.g., *pineapple, crab apple, apple sauce*) . Bowden and Jung-Beeman (2003) had previously found that participants sometimes reported solving such problems "with insight", and sometimes reported that solutions were obtained "without insight" and that "insight" solutions were associated with primes sent to the right hemisphere as against left-hemisphere primes. ("Insight" was explained to the participants as a sudden "Aha" experience coupled with a certainty that the solution was right.) In the FMRI study it was found that "insight" solutions were associated with increased activity in the right anterior superior temporal gyrus as compared to non-insight solutions. The same brain area showed increased EEG activity beginning shortly before insight solutions were reported compared to when non-insight solutions were reported. This work suggests that there are differences in underlying neural activity between solving problems with and without insight. It may be noted that in the Jung-Beeman et al. studies, "insight" has been defined in terms of the suddenness of solution rather than in terms of restructuring. In solving an RAT item it would seem that the participant moves from an initial structure where the problem words are unrelated to one where they are tightly related via the solution word, so that restructuring always occurs in solution, but sometimes through a basically sequential procedure ("non-insight") and sometimes through highly simultaneous or parallel processes (leading to experience of "insight").

DUAL PROCESS APPROACHES AND INSIGHT

Recently, *dual process* approaches to thinking, reasoning, and decision making have been fruitfully developed by Stanovich and West (2000), Kahneman (2003), Sloman (1996), Evans (2003, 2005) and others. In these accounts two distinct cognitive systems are proposed. System 1 is seen as automatic, implicit, fast, and as generating intuitive, immediate responses. This system is assumed to be relatively old in evolutionary terms and is very similar between humans and other animals. It comprises a "*set* of systems in the brain (partially encapsulated modules in some views) that operate autonomously" (Stanovich & West, 2003, p. 182) such as reflexes, instincts, innate input modules, and processes formed through associative learning. Only the final product of such processes is available to consciousness. System 2, on the other hand, is seen as evolutionarily recent and is peculiar to humans. It permits abstract reasoning and hypothetical thinking, operates relatively slowly and sequentially, is constrained by working memory capacity, and is highly correlated with general fluid intelligence and with performance on sequentially solvable non-insight problems in general. System 2 processes are impaired by dual task activity but System 1 processes are not. Interference by dual tasks is a signature of involvement of working memory as conceptualised by Baddeley (2000) in his very influential model, and the concepts of System 2 and working memory are closely interrelated. System 2 can be seen as incorporating working memory as an integral component system involving both storage and executive control. Evidence for working memory involvement, particularly central executive involvement, may be taken as an indication of System 2 involvement. On this basis, dual task and individual difference studies have implicated working memory and hence System 2, in a range of non-insight tasks such as Tower of London (Gilhooly, Phillips, Wynn, Logie, & Della Sala, 1999b; Gilhooly, Wynn, Phillips, Logie, & Della Sala, 2002) and syllogisms (Gilhooly, Logie, & Wynn, 1999a). The two systems are seen as interacting and an important role for System 2 is to inhibit and override System 1 when appropriate, e.g., when the costs of errors are high. In the other direction, pre-attentive System 1 processes markedly influence the attentional focus of System 2 processes (Evans, 2005).

From a theoretical point of view, System 2 processes would be expected to be strongly implicated in insight problem solving by approaches that regard restructuring as requiring explicit, executively demanding processes. For example, Kaplan and Simon (1990) proposed that insight resulted from explicit search processes at the level of problem representations. On the other hand, the original Gestalt analyses of insight (see Humphreys, 1951; Kohler, 1947) proposed automatic processes that resolved "stresses" inherent in misleading representations and led to useful representations free from internal stresses. For example, Maier (1931, p. 193) stated: "The perception of the solution of a problem is like the perceiving of a hidden figure in a puzzle picture. In both cases (a) the perception is sudden, (b) there is no conscious intermediate stage; and (c) the relationships of the elements in the final perceptions are different from those which preceded, i.e., changes of meaning are involved." More recently, Ohlsson (1992) and Schooler et al. (1993) have argued that insight results from automatic, implicit, non-executive processes. Similarly, Jung-Beeman et al. (2004) argued that insight "... involves seeing a problem in a new light, often without awareness of how that new light was switched on" (p.14). Such views, which stress the role of automatic, unconscious, implicit processes in insight problem solving, would suggest that System 2 processes would not be heavily implicated in insight problem solving. Evans (2005) has discussed insight problem solving in relation to System 1 and System 2 processes and suggested complementary roles for the two systems. System 1 affects the solver's attentional focus through pre-attentive unconscious processes, while System 2 is needed for understanding the solution and in order to transfer the solution to related or analogous problems.

Overall, the general implication of the literature reviewed above seems to favour the view that insight problem solving shows a greater involvement of System 1 processes than typical incrementally solvable non-insight problems, and conversely that non-insight problems show a greater involvement or dependence on System 2 processes than typical insight problems. However, although the role of System 2/working memory in non-insight tasks has been extensively documented, the involvement of System 2/working memory in insight tasks has not hitherto been directly examined. The present study examines the possible contributions of System 2/working memory processes to insight problem solving compared with non-insight problem solving.

As has been noted above, previous studies exploring possible ways of differentiating insight and non-insight tasks have generally employed small numbers of problems. In the present study a relatively large set (n = 24) of presumed insight tasks was gathered, which varied markedly in "surface" characteristics. To provide a contrasting set, a number of problems (n = 10) generally agreed to be non-insight tasks were assembled. Participants were tested on all insight and non-insight problems and an initial cluster analysis was carried out to investigate whether insight tasks grouped with other insight tasks.

In the present study an individual differences approach is applied and a range of indicators of System 2 processes are assessed so that possible differences between insight and non-insight tasks in the degree to which they load different System 2 processes may be assessed. Details of the tests used here are given in the Method section but to summarise, we assess general, fluid intelligence ("g") by means of Raven's Matrices, crystallised verbal ability by means of Vocabulary and Category Fluency tests, standard measures of verbal and spatial working memory and short-term memory, and measures of flexibility (Letter and Figural fluency and Alternative Uses) which are generally taken to be executively loading. The resulting data were expected to provide further evidence regarding possible differentiation of insight from non-insight problems, with particular reference to the issue of the extent to which insight processes are automatic (System 1) or executive (System 2) as compared with non-insight solution processes.

METHOD

Problem tasks

A summary listing of the problems used is given in Table 1. The *presumed insight problems* were as follows: *Nine dot* (Connect nine dots in a square array by drawing four straight lines). *Inverted pyramid* (On a steel table is a £50 note. On the note is a large steel pyramid, which is balanced upside down. Remove the note without upsetting the pyramid; Ohlsson, 1992). *X ray* (A patient has an inoperable tumour in the middle of his body. A ray machine destroys tissue as well as tumours. How can the ray machine be used to destroy the tumour without damaging healthy tissue? Duncker, 1945). *Triangle* (Given a diagram of circles arranged in a triangle shape with one in the top row then two, three, and four in lower rows, how can you move three circles to make the triangle point the other way? Schooler et al., 1993). *Pound coins* (Why are 1992 pound coins worth more than 1991 pound

TABLE 1
List of tasks

Insight tasks		
Nine dot	Minimutilated checkerboard	Six matches
Triangle of circles	Pigpen	Necklace
Four trees	Farm	
Inverted pyramid	Matching socks	Murples
Duncker's candle	Duncker's radiation	Two string
Trains and bird	Water jars	
Coins	Earth in hole	Horse trading
Ocean liner	Reading in dark	Football
Marrying	Lake	
Non-insight tasks		
Hobbits and orcs	Tower of London	Tower of Hanoi
Syllogisms	Dinner party	Cards
Suspects	Heavy/light coins	Anagrams
Cryptarithmetic		

coins? After Dominowski, 1994). *Football scores* (Joe Fan has no psychic powers but he can tell you the score of any football game before it starts. How? Dominowski, 1994). *Marriage* (A man in a small town married 20 different women of the same town. All are still living and he never divorced. Polygamy is unlawful but he has broken no law. How can this be? Dominowski, 1994). *Matchsticks* (Given a diagram showing six matchsticks lying on a table, make four equal-sided triangles. Ohlsson, 1992). *Minimutilated checkerboard* (Given a 6 × 6 checkerboard with 2 black squares removed from opposite corners, can the remaining 34 squares all be covered by 17 dominoes? After Kaplan & Simon, 1990). *Pigpen problem* (Nine pigs are kept in a square pen. Build two more square enclosures that would put each pig in a pen by itself; Schooler et al., 1993). *Four trees problem* (How could you plant four trees so that each one is an equal distance from each of the others? Schooler et al., 1993). *Farm problem* (How could you divide an L-shaped piece of land into four equally shaped pieces of equal sizes? Metcalfe, 1986). *Matching socks* (There are black and brown socks in a drawer mixed in a ratio of 4 to 5. How many socks would you have to take out without looking to be sure of getting a pair of the same colour? Sternberg, 1987). *Murples problem* (There is a container of Murples. The Murples double in number every day. The container will be full in 60 days. In how many days will it be half full? After Sternberg & Davidson, 1982). *Candle problem* (You have a candle, some matches, and a box of tacks. How can you support the candle on the wall? Duncker, 1945). *The two string problem (*Given two strings hanging from the ceiling, which are too far apart to be both grasped at once, find a way of using objects in the room to tie the strings together. Maier, 1931). *Trains and bird problem* (Two trains 50 miles apart start towards each other at 25 mph. As the trains start, a bird flies from the front of one train towards the second. On reaching the second train the bird turns round and flies back to the first train, and so on until the trains meet. If the bird flies at 100 mph, how many miles will the bird have flown before the trains meet? Posner, 1973). *Earth in hole problem* (How much earth is there in a hole 3 ft by 3 ft by 3 ft? Dominowski, 1994). *Horse trading problem* (A man buys a horse for £60, sells it for £70, buys it back for £80 and sells it finally for £90. How much has he made? Dominowski, 1994). *Ocean liner problem* (At 12 noon a porthole in an ocean liner was 9 ft above the water line. The tide raises the water at a rate of 2 ft per hour. How long will it take the water to reach the porthole? Dominowski, 1994). *Reading in dark* (A man is reading a book when the lights go off but even although the room is pitch dark the man goes on reading. How? Dominowski, 1994). *Necklace problem* (A woman has four pieces of chain. Each piece is made up of three links. She wants to join the pieces into a single closed ring of chain. To open a link costs 2 cents and to close a link costs 3 cents. She has only 15 cents. How does she do it? Metcalfe, 1986). *Lake problem* (Someone walked for 20 minutes on the surface of a lake without sinking but without any form of flotation aid. How? Dominowski, 1994). *Water jar problems* (Given jugs of different sizes, solve a series of problems to get specified amounts of water. The first four tasks induce a set, which must be broken in the fifth problem. Our interest with this task was in performance on the fifth, set-breaking problem. After Luchins, 1942). On Weisberg's (1995) taxonomy, the above problems would appear to be pure insight tasks with the exceptions of the Nine dot, Triangle, and Necklace problems, which would be classed as hybrid.

Presumed non-insight problems were as follows: *Tower of London* (Manipulate five disks on pegs to match the target configuration in minimum moves; Gilhooly et al., 1999). *Syllogistic reasoning* (Given two categorical premises, draw necessary conclusion; e.g., Gilhooly et al., 1999). *Suspects problem* (Given statements by four suspects for a crime, infer who committed the crime; Schooler et al., 1995); *Hobbits and orcs* (Given a boat that can only take two creatures, how can you get three hobbits and three orcs across a river in such a way that the hobbits are never outnumbered by the orcs on either side? Thomas, 1974). *Tower of Hanoi* (Given four discs stacked in decreasing size on a peg and two empty pegs, move the discs to a target peg, one at a time, in such a way that a larger disc is never placed on a smaller disc; Egan & Greeno, 1974). *Dinner party problem* (Given five guests with specified food aversions and a list of foods, construct a menu all could eat; Schooler et al., 1993). *Cards problem* (Given three cards on a table face down, from limited information identify which suit each card is; Schooler et al., 1993). *Heavy and light coins* (Given four coins of which two are slightly heavy and two slightly light, but which look and feel identical, how could you find out which are which in two weighings on a balance scale? Schooler et al., 1993). *Anagrams* (Unscramble eight 5-letter word anagrams; Gilhooly & Johnson, 1978). *Cryptarithmetic* (Work out what numbers different letters stand for, given that D = 5, and DONALD + GERALD = ROBERT; Newell & Simon, 1972).

Individual difference measures

The *individual difference measures* encompassed fluid and crystallised intelligence, short-term and working memory capacities, and tests related to ideational fluency and flexibility. These measures were intended to reflect System 2/working memory capacities and processes. The measures were as follows.

Raven's Progressive Matrices (Raven, 1960). Participants worked through the booklet as far as they could within a 20-minute time limit. In this test, the items consist of visual patterns that are related by some rule and the participant has to identify the rule. The rules concerned vary markedly in complexity. This test is regarded as an exemplary measure of fluid general intelligence or "g" (Jensen, 1980). Multidimensional scaling of a range of tests found that Raven's test places at the centre of the solution, indicating that it is closer to all the other tests than those that are placed nearer the periphery (Snow, Kyllonen, & Marshalek, 1984). It may be that Raven's draws on processes that underpin all intelligence tests and many forms of problem. Candidate processes involved in Raven's may include maintenance of goal structures, metacognitive awareness of progress, and systematic exploration of options, all of which are executively loading System 2 processes.

Digit span. The stimuli were taken from WAIS-R digit span test (Wechsler, 1981). Digit span was taken as mainly a measure of short-term verbal memory with a relatively low central executive involvement.

Sentence span (Baddeley, Logie, Nimmo-Smith, & Brereton, 1985). Participants were read a series of statements and had to state whether each was true or false, and then recall the final word of each sentence in the correct order. Participants heard two sets of two sentences to start with, then two sets of three, and so on up to eight. The stimulus sentences were read out from sheets on which the participants' responses were also recorded. Span was measured as the maximum sequence length at which participants correctly recalled all the words in at least one of the trials. Such span tests (cf. Daneman & Carpenter, 1980) are seen as largely reflecting central executive resources (since they require appropriate attentional control and switching).

Visual pattern span. Participants were required to recall grid patterns (Della Sala, Gray, Baddeley, Allamano, & Wilson, 1999) with filled squares ranging from 2 to 15. Each pattern was presented for 2 seconds and participants were given as much time as required to complete each trial. The resulting span was taken as a measure of visuo-spatial temporary storage capacity.

Corsi blocks. This was a computerised version of Corsi's (1972) nine-block spatial span task. The sequences ranged from three to nine blocks, with two trials at each sequence length. The participant was seated in front of the computer monitor with the first practice trial on screen. They were instructed that certain of the blocks would light up in sequence and then the word "recall" would appear on the screen. Participants then had to tap the blocks with the light pen in the same order as they lit up on the screen. There were three practice trials. Block selected and time taken to select each block was recorded. The task was automatically terminated when a participant had made errors in each of the two trials of a particular span. Again, span was measured as the maximum sequence length at which participants reproduced at least one sequence correctly. The span score was taken as a measure of visuo-spatial working memory capacity for spatial information.

Corsi distance estimation (Gilhooly et al., 2002). This is a variation of the Corsi block task and was devised to be a visuo-spatial analogue of working memory sentence span tasks. Nine blocks were visible on the PC screen. One block lit up (the reference block) followed by a second block, the words "estimate distance" would appear on the screen and participants were asked to estimate how far apart the centres of the two blocks were (in inches or centimetres) and say the distance aloud. After this, the reference block lit up again, followed by a different block, and again participants had to estimate the difference between them. After this, an instruction to recall appeared on the screen and participants were to recall the two *second* blocks. Sequences began with two blocks and ranged to nine blocks. The resulting span was taken as reflecting mainly executive function, as it requires both storage and processing.

Letter fluency (Phillips, 1997). Participants were given 1 minute in which to produce as many different examples of words starting with a given letter (F) as they could. This task is generally seen as requiring development of novel strategies and strategy switching, and hence as tapping executive functions.

Figural fluency (Phillips, 1997). Participants were presented with an A4 sheet containing boxes in each of which was the same pattern of five dots. The task was to produce as many different figures as possible by joining at least two dots in each pattern of five, using straight lines, within a 1-minute time limit. This task requires developing novel strategies and inhibiting previous solutions, and hence is seen as tapping executive (System 2) functions.

Mill-Hill vocabulary test B (Raven, 1960). This test is a measure of crystallised intelligence and as an indicator of vocabulary size may well relate to performance on problems requiring detection and resolution of verbal ambiguities.

Alternative uses test (Guilford, Christensen, Merrifield, & Wilson, 1978). Participants were asked to produce as many different uses as they can think of for six common objects, which are different from the normal use. One minute is allowed per object and score is total number of distinct uses over the six objects. This test is an indicator of ideational flexibility and, as with Figural and Letter Fluency tasks, requires novel strategies and the inhibiting of previous solutions and can be regarded as a test of executive (System 2) functions.

Category fluency test (Martin, Wiggs, Lalonde, & Mack, 1994). Participants were asked to produce as many different examples of two categories (fruits and animals) and were allowed 3 minutes per category. Score was total number of instances over the two categories. This test indicates fluency of retrieval from long-term memory.

Participants

A total of 60 students at Brunel University were tested. Participants were between 18 and 35 years of age (mean age = 22.00 years); 41 female and 19 male. Participants were paid £30 for 6 hours of testing.

Procedure

Participants took part in three 2-hour sessions. The individual difference tasks were given in the first session and the problem-solving tasks over a further two sessions. Each problem was presented on a separate sheet of paper and participants were allowed time to read through each problem once, followed by 5 minutes within which to attempt a solution. The only exceptions to this were the syllogisms, the Tower of London, and the anagrams tasks. For the syllogisms, participants were given 10 minutes to answer 10 problems presented in a multiple-choice format. In the Tower of London tasks, participants were allowed 2 minutes on each of five problems. Participants were allowed 2 minutes to attempt the eight anagram problems. Scores for these three tasks were the number of problems correct in the time allowed. For the other problems, scores were obtained for whether correctly solved or not and for time to solution (non-solutions were given a time score of 300 seconds). The order of presentation of the problems was varied over participants and was such that problems of a similar type did not follow each other immediately.

Participants were allowed to use blank sheets of paper to make notes in working on most of the tasks. The exceptions were the Hobbits and Orcs, the Triangle, the Tower of London, and the Tower of Hanoi; for these problems concrete versions of the tasks were provided to work with. For all the other tasks participants were allowed to ask questions to which only a “Yes” or “No” answer was provided. Participants were asked to propose solutions as soon as they could, and time to solution was recorded.

RESULTS AND DISCUSSION

Table 2 shows mean times, or solution scores where appropriate, together with measures of skew and kurtosis. It is clear from the Table that certain problems were very rarely solved in the time limit—the Minimutilated checkerboard, the Nine dots, the Four trees, and the Farm problems. These were dropped from further analysis. Water jars was also dropped because participants found the initial set-inducing tasks more difficult than anticipated and few participants got as far as tackling the critical set breaking fifth problem, which was our target task. Analyses are principally based on time measures to solution (non-solution scored as 300 s) because time measures are more discriminating than simple solved/not solved scoring. A number of the time measures were markedly skewed and transformations were applied to reduce skews. All scores were oriented so that a larger score meant a better performance. For example, times were subtracted from 301, thus a quick solution would receive a high score. This procedure should make interpretation of the results more straightforward.

Cluster analysis of tasks

In order to investigate whether the tasks do form distinct groupings, *cluster analysis* was applied with the problems as cases and the participants as variables (known as R-analysis, Romesburg, 1984). The scores in the proximity matrix were standardised to overcome scaling differences between the variables and Ward’s method of clustering was employed. The resulting dendrogram is shown in Figure 1. It appeared that nine clusters emerged, which were mostly quite homogeneous as either insight or non-insight clusters. The clusters were as follows:

Cluster 1: Earth hole; Ocean liner; Murples; Matching socks; Football scores
Cluster 2: Dinner party; Cards; Heavy and light coins
Cluster 3: Trains and bird; Candle; X-ray
Cluster 4: Pound coins; Horse trading; Matchsticks
Cluster 5: Syllogisms; Tower of London; Necklace
Cluster 6: Pigpen; Inverted pyramid; Triangle; Marriage; Lake; Suspects
Cluster 7: Hobbits and Orcs; Tower of Hanoi
Cluster 8: Anagrams; Cryptarithmetic
Cluster 9: Two strings; Reading in dark

TABLE 2
Descriptive statistics on problem performance

		Descriptive Statistics		
	Mean	*SD*	*Skewness*	*Kurtosis*
ANAGRAMS	2.85	2.07	0.602	− 0.134
CANDLE	220.02	103.26	− 0.901	− 0.809
CARDS	223.28	86.13	− 0.524	− 1.351
CRYPTARITH	0.58	0.79	1.762	4.692
DINN PARTY	81.32	51.13	1.715	4.771
DUNCKER RAD	274.12	67.62	− 2.608	5.853
EARTH HOLE	100.02	118.39	0.813	− 1.029
FARM	292.10	39.49	− 5.656	33.603
FOOTBALL	227.55	112.44	− 1.123	− 0.472
FOUR TREES	293.67	30.76	− 6.046	39.331
HEAVYLIGHT	195.78	101.97	− 0.382	− 1.303
HOBBORCS	11.25	3.90	1.888	2.841
HORSE TRADE	142.10	104.83	0.490	− 1.398
INVERT PYRAMID	265.72	74.90	− 2.217	3.772
LAKE	141.08	133.15	0.265	− 1.843
MARRYING	185.88	130.50	− 0.422	− 1.695
MATCHES	267.08	80.41	− 2.349	4.145
MATCHSOC	166.60	113.05	− 0.092	− 1.553
MINIMUTI	295.67	25.27	− 6.356	42.176
MURPLES	136.55	114.89	0.396	− 1.415
NECKLACE	282.03	43.86	− 2.676	6.553
NINEDOTS	296.33	26.35	− 7.211	51.997
OCEANLINER	178.10	126.68	− 0.300	− 1.719
PIG PEN	249.10	85.41	− 1.448	0.633
POUND COINS	107.47	126.22	0.788	− 1.257
READ DARK	138.93	128.02	0.327	− 1.791
SUSPECTS	216.77	91.11	− 0.639	− 1.074
SYLLOGISMS	2.95	1.61	0.387	− 0.002
TOWERHANOI	25.65	10.13	0.799	− 0.400
TOWERLONDN	2.67	1.34	− 0.106	− 0.699
TRAIN & FLY	202.25	100.37	− 0.315	− 1.583
TRIANGLE	142.83	104.45	0.428	− 1.219
TWO STRINGS	253.25	90.11	− 1.649	1.058
WATERJARS	279.17	46.87	− 2.732	8.591

All measures solution times (max = 300 s), except Anagrams, cryptarithmetic, syllogisms, Tower of London, (n solved), Hobbits and Orcs, and Tower of Hanoi (moves to solve). $N = 60$. Problems listed in alphabetical order.

The average intercorrelations among clustered items and across all items were examined. The overall average correlation among all 29 problems was 0.139. The average within-cluster correlations were: for cluster 1, .413; for cluster 2, .329; for cluster 3, .279; for cluster 4, .406; for cluster 5, .243;

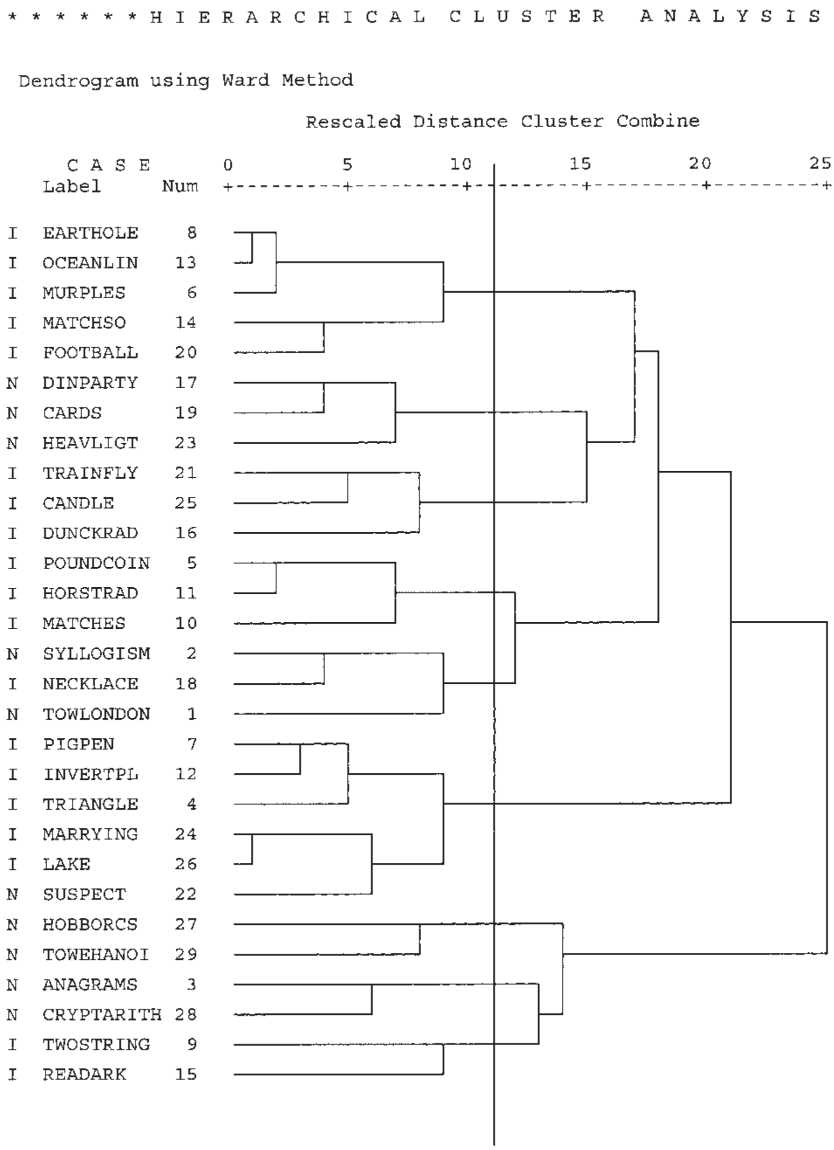

Figure 1. Cluster analysis of insight (I) and non-insight (N) tasks. Proposed "cut" of dendrogram indicated by line.

for cluster 6, .354; for cluster 7, .184; for cluster 8, .325; and for cluster 9, .214. Thus, intercorrelations within clusters were higher than the overall average.

Clusters 1, 3, 4, and 9 were exclusively insight problems, while clusters 2, 7, and 8 were exclusively non-insight. Cluster 6 was almost wholly composed of insight problems, with five insight and one presumed non-insight task. Cluster 5 was mixed with two non-insight and one presumed insight problem (Necklace). Interestingly, Weisberg (1995, p. 192) classes the Necklace problem as a hybrid that may be solved either through restructuring (insight) or non-insight means, and this is consistent with its position in a mixed cluster. Of the 17 problems in this analysis that would be classed as Pure Insight on Weisberg's (1995) scheme, all fall into predominantly insight clusters. Of the 10 non-insight problems in the above analysis, all but 1 fall into predominantly non-insight clusters. Questions remain regarding the psychological meaning of the particular clusters. For example, what underlies the largest cluster, Cluster 1 (which comprises Earth hole; Ocean liner; Murples; Matching socks; Football scores)? Some links that suggest themselves are that Earth hole, Ocean liner, and possibly Football scores seem to violate Gricean assumptions (Grice, 1989) in that the solver is asked for quantities that do not exist (How much earth is in a hole? None. How long before sea reaches porthole? It never will. How does man know score before game? There is no score before the game). Murples and Matching socks may benefit from visualising the proposed situations and perhaps running models of actions in the situations forwards and backwards. Exactly why these latter two are placed with the "Gricean" puzzles is not obvious; however, in the data these five puzzles "go together" in that those who do well or poorly at one do similarly with the others. Overall, the psychological bases for the clusters of insight problems identified here merit further analysis and testing.

Individual difference analyses

The individual difference tests were used to investigate whether insight and non-insight problem solving were predicted equally well by measures of fluid intelligence, vocabulary, flexibility, fluency, and working memory capacities, which were taken to be indices of System 2 functioning. If individual difference measures show different patterns of relationships to insight and non-insight problems, this would suggest that distinctive processes and capacities are involved in tackling the various kinds of problems.

Correlations between the individual difference measures and the problem-solving tasks were calculated. Using *t*-tests to compare the 19 presumed insight vs the 10 presumed non-insight tasks, significant differences between average correlations over the two types of tasks were found for Figural fluency (average r with insight tasks = .21 vs .11 for non-insight tasks, $t = 1.75$, $df = 27$, $p < .05$); Vocabulary (average r with insight tasks = .18 vs .06 on non-insight tasks, $t = 2.18$, $df = 27$, $p < .05$); and Alternative uses (average r with insight tasks = .12 vs .03 for non-insight tasks, $t = 1.90$, $df = 27$, $p < .05$). Thus, performance on insight tasks tended to be more strongly (positively) related to these measures than was performance on non-insight tasks. These results support the view that *cognitive flexibility* (as indexed by Figural fluency and Alternative uses tests) is particularly important for insight problem solving compared to non-insight problem solving. The relatively greater influence of Vocabulary on insight tasks than on non-insight tasks may well reflect the number of insight tasks that depend on detection and resolution of verbal ambiguities.

Composite insight task scores for each participant were initially formed by averaging solution speed scores over the 19 presumed insight tasks. Similar *non-insight task composite scores* were initially formed by converting the non-insight task scores into standard scores and averaging these. Correlations between each item and the initial composite into which it entered were obtained. Those items that had no significant correlation with their initial composite were then removed from the final composite scores. (On this basis, the Two-string and Duncker's X-ray problems were dropped from the insight composite, and the Hobbits and Orcs problem was removed from the non-insight composite). The resulting composites were normally distributed and over the 60 participants correlated with each other $r = .50^{**}$, indicating moderate overlap but with considerable unshared variance (75%). Table 3 shows the correlations of individual difference measures with composite insight and non-insight scores. The main differences between insight and non-insight tasks that emerge from this Table concern correlations with Figural fluency, Alternative uses, and Vocabulary. The results from this analysis mirror those from the comparison above of correlations when tasks were treated as cases. Insight tasks appear to draw more heavily on cognitive flexibility (Alternative uses) and fluency (Figural fluency) than do non-insight tasks. Vocabulary correlates with the insight composite score but not with the non-insight composite score. As suggested in the item-level analysis above, the importance of vocabulary for insight tasks may well reflect the importance of verbal ambiguity detection and resolution in many insight problems.

The pattern of simple correlations of working memory measures with composite insight scores is similar to that with composite non-insight scores, and may appear to support the view that working memory factors play a role in insight problem solving as well as in non-insight tasks. However, in view of the complex pattern of intercorrelations among the individual difference variables, apparent links with working memory and other

TABLE 3
Correlations

	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*
1. Insight comp.	–												
2. Non-ins comp.	.50**	–											
3. Ravens	.41**	.50**	–										
4. Sentence span	.23*	.23*	.47**	–									
5. Digit span	−.02	−.09	.14	.23*	–								
6. Visual span	.24*	.24*	.15	.48**	.21	–							
7. Corsi blocks	.08	.17	.30**	.15	.12	.09	–						
8. Corsi distance	.34*	.25*	.28*	.27*	.03	.20	.28*	–					
9. Letter fluency	.13	.03	.20	.14	.08	−.11	.22*	.00	–				
10. Figural fluency	.41**	.25*	.45**	.22*	.02	.02	.26*	.27*	.37**	–			
11. Vocabulary	.41**	.14	.32**	.18	.27*	.21	.10	.02	.34**	.02	–		
12. Alternative uses	.32*	.30*	.35**	.34**	.16	.19	.20	.17	.11	.28*	.48**	–	
13. Category fluency	.26*	.09	.35**	.23*	.14	.19	.27*	.10	.39**	.22	.53**	.50**	

Correlations among individual difference measures and composite scores on insight and non-insight tasks. $N = 60$. * $p < .05$; ** $p < .01$.

variables may be due to confoundings among variables. To take account of such intercorrelations and possible confoundings we undertook separate simultaneous multiple regressions of the insight and non-insight composites as dependent variables with the individual difference scores as independent variables.

Multiple regression analyses

The results of the multiple regression analyses for insight and non-insight composite scores are shown in Table 4.

From Table 4, it appears that when all other measures were partialled out, insight problem solving was predicted significantly by Figural fluency and by Vocabulary, while non-insight problem solving was predicted only by Raven's Matrices scores. This suggests that the two types of task are differentiated by the particular ability patterns on which they draw.

It may be that the regression results for insight tasks reflect a distinction between verbal insight tasks and spatial insight tasks such that verbal tasks only are affected by Vocabulary (which may facilitate detection of ambiguities) and spatial tasks are affected only by Figural fluency (which reflects ability to generate a range of spatial interpretations). To assess this possibility, our initial insight composite score was divided into a *spatial insight composite* (formed from Triangle, Pigpen, Matches, Necklace, and

TABLE 4
Multiple regressions: Insight and non-insight scores

	Dependent variable			
	Insight composite		*Non-insight composite*	
	Beta wt.	*t*	*Beta wt.*	*t*
Independent variable				
Ravens	0.091	0.61	0.45	2.99**
Sentence span	0.02	0.10	– 0.02	– 0.13
Digit span	– 0.18	– 1.53	– 0.22	– 1.82
Visual span	0.12	0.89	0.17	1.23
Corsi blocks	– 0.09	– 0.77	0.00	0.03
Corsi distance	0.18	1.52	0.11	0.87
Letter fluency	– 0.20	– 1.39	– 0.20	– 1.38
Figural fluency	0.42	2.94**	0.11	0.73
Vocabulary	0.48	3.11**	0.06	0.35
Alternative uses	– 0.17	– 1.19	– 0.26	– 1.77
Category fluency	0.08	0.60	0.30	1.97

Simultaneous multiple regressions of individual difference measures on insight and non-insight composite scores. ** = $p < .01$, $df = 49$.

TABLE 5
Multiple regressions: Verbal and spatial insight scores

	Dependent variable			
	Verbal insight composite		*Spatial insight composite*	
	Beta wt.	*t*	*Beta wt.*	*t*
Independent variable				
Ravens	0.08	0.61	0.11	0.92
Sentence span	0.04	0.12	0.06	0.48
Digit span	−0.19	1.39	0.04	0.18
Visual span	0.09	0.60	0.03	0.07
Corsi blocks	0.10	0.31	0.03	0.13
Corsi distance	0.21	1.62	0.07	0.73
Letter fluency	0.24	1.98	0.05	0.27
Figural fluency	0.35	2.51**	0.50	3.15**
Vocabulary	0.53	3.52**	0.12	0.71
Alternative uses	0.16	1.10	0.14	0.94
Category fluency	0.08	0.61	0.20	1.32

Simultaneous multiple regressions of individual difference measures on verbal and spatial insight composite scores. ** = $p < .01$, $df = 49$.

Candle) and a *verbal insight composite* was formed from the remaining insight tasks. These two composites were normally distributed and correlated with each other ($r = .54$**) indicating some overlap but also c. 70% unshared variance. Multiple regressions were then obtained, and the results for verbal and spatial insight tasks are shown in Table 5.

From Table 5, it appears that when all other measures were partialled out, verbal insight problem solving was predicted significantly mainly by Vocabulary but also to a lesser extent by Figural fluency, while spatial insight problem solving was predicted only by Figural fluency and neither was predicted by Raven's Matrices scores. This suggests that the two subtypes of insight task are differentiated from each other and also from non-insight tasks in terms of the particular ability patterns on which they draw.

GENERAL DISCUSSION

This study sought to add to available knowledge regarding the extent to which insight and non-insight problems could be empirically differentiated, and in the extent to which insight problem solving drew on System 2 processes. Our division of tasks into presumed insight and presumed non-insight tasks was largely based on Weisberg's (1995) *a priori* taxonomy, which sorts tasks into those necessitating restructuring (pure insight), those that do not involve restructuring (non-insight), and hybrids that may or may not involve restructuring. The cluster analysis reported here broadly supported Weisberg's taxonomy, as our presumed insight tasks strongly tended to cluster with other insight tasks and non-insight tasks clustered mainly with other non-insight tasks. This supports the notion that there are some common processes underlying performance on insight tasks distinct from those underlying non-insight tasks.

Recent dual-process approaches to thinking (Evans, 2003; Kahneman, 2003; Stanovich & West, 2000), which contrast automatic and implicit System 1 processes with slower, sequential, executive- and working-memory-loading System 2 processes, are relevant to the study of insight problem solving. Classic Gestalt approaches proposed that insight resulted from automatic (i.e., System 1) processes of representational restructuring, which arose from the inherently unstable nature of misleading problem representations (see Humphreys, 1951; Kohler, 1947). More recently, Ohlsson (1992) and Schooler et al. (1993) have also proposed that insight results from automatic, implicit, non-executive processes of the type attributed to System 1. On the other hand, Kaplan and Simon (1990) argued in favour of explicit search processes (System 2) seeking changes to the problem representation as the basis for insight problem solving. Our results favour an involvement of System 2 processes in insight problem solving. Furthermore, we found that the pattern of loading of non-insight tasks and insight tasks on individual difference measures of System 2/working memory functions differed in an interpretable way. Raven's Matrices, which involve systematic search and a heavy working memory load, predicted performance in non-insight tasks, but insight task performance was better predicted by measures of strategic switching and inhibition (e.g., Figural fluency). The inference is that insight tasks particularly draw on System 2/executive processes of switching and inhibition presumably in order to effect strategic switches of approach and inhibit strongly evoked but misleading strategies. This interpretation fits well with generally accepted views of how System 1 and 2 interact (Evans, 2003). System 1 pre-attentive processes will determine the initial problem representation and initial approaches. If the initial representation and approaches do not yield progress, as typically occurs with insight problems, an important role for System 2 is then to inhibit and override System 1 to develop alternative approaches and representations.

The results reported here provide further behavioural evidence in support of distinctions between insight and non-insight tasks, and so add to the differences reported by Metcalfe and Weibe (1987). Our results also have possible implications for the differences reported by Schooler et al. (1993). The links found here between strategy-switching functions and insight problem solving are consistent with Metcalfe and Weibe's findings that "feelings of warmth" did not increase while tackling insight problems until the solution was reached, but did increase steadily throughout the solution period for non-insight problems. Until the correct strategy switch has been made there would be no association between feeling of warmth and progress for insight tasks. In non-insight tasks a workable strategy is quickly found and applied steadily to reach solution in an incremental fashion.

The differential impairing effect of concurrent verbalising on insight tasks vs non-insight tasks reported by Schooler et al. (1993) may benefit from re-examination in the light of the distinctions found here between verbal and spatial insight tasks regarding the relative loading of spatial flexibility measures, which were important for spatial tasks, and vocabulary measures, which were important for verbal tasks, many of which involve ambiguities. Most of Schooler et al.'s insight tasks would seem to be quite spatial and thus concurrent verbalising may well interfere with appropriate spatial coding in these tasks. In contrast, the non-insight tasks in Schooler et al.'s study would seem to be predominantly verbal and so verbalising would not be expected to cause interference. This suggested interpretation could readily be tested in future studies by examining effects of verbalising on both spatial and verbal insight and non-insight tasks.

In conclusion, we suggest that the question of whether insight problem solving requires System 2 processes seems to be answered here in the affirmative, but the System 2 processes involved differ between insight and non-insight problems. Specifically, the results support a relatively large role for the System 2 executive processes of strategy switching and inhibition in insight problem solving, and a greater role for System 2 based systematic search processes in non-insight tasks. The issue has sometimes been posed in rather stark, exclusive "either – or" terms as "do insight problems involve System 2/working memory or not?". The present discussion suggests that a restructuring of the original question into one of the differential degree to which different System 2/working memory functions are involved in insight vs non-insight problems may be more fruitful.

Manuscript received 23 November 2003
Revised manuscript received 11 November 2004

REFERENCES

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*, 417 – 423.

Baddeley, A. D., Logie, R. H., Nimmo-Smith, I., & Brereton, N. (1985). Components of fluid reading. *Journal of Memory and Language*, *24*, 119 – 131.

Bowden, E. M., & Jung-Beeman, M. (2003). Aha! Insight experience correlates with solution activation in the right hemisphere. *Psychonomic Bulletin*, *10*, 730 – 737.

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

Corsi, A. T. (1972). *Human memory and the medial temporal region of the brain*. Dissertation, McGill University, Montreal.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466.

Davidson, J. E. (2003). Insights about insightful problem solving. In J. E. Davidson & R. J. Sternberg (Eds.), *The psychology of problem solving*. Cambridge: Cambridge University Press.

Della Sala, S., Gray, C., Baddeley, A., Allamano, N., & Wilson, L. (1999). Pattern span: A tool for unwelding visuo-spatial memory. *Neuropsychologia*, *37*, 1189–1199.

Dominowski, R. (1994). *Insight and instructions*. Paper presented at 11th Annual Conference, Cognitive Psychology section, British Psychological Society. New Hall, Cambridge, September 1–3.

Duncker, K. (1945). On problem solving. *Psychological Monographs*, *58*, 1–113.

Durso, F. T., Rea, C. B., & Dayton, T. (1994). Graph-theoretic confirmation of re-structuring during insight. *Psychological Science*, *2*, 94–98.

Egan, D. W., & Greeno, J. G. (1974). Theories of rule induction: Knowledge acquired in concept learning, serial pattern learning and problem solving. In L. W. Gregg (Ed.), *Knowledge and cognition*. New York: John Wiley.

Evans, J. St. B. T. (2003). In two minds: Dual process accounts of reasoning. *Trends in Cognitive Sciences*, *7*, 454–459.

Evans, J. St. B. T. (2005). Insight and self insight in reasoning and decision making. In V. Girotto & P. N. Johnson-Laird (Eds.), *The shape of reason: Essays in honour of Paolo Legrenzi*. Hove, UK: Psychology Press.

Gilhooly, K. J., & Johnson, C. E. (1978). Effects of solution word attributes on anagram difficulty: A regression analysis. *Quarterly Journal of Experimental Psychology*, *30*, 57–70.

Gilhooly, K. J., Logie, R. H., & Wynn, V. (1999a). Syllogistic reasoning tasks, working memory and skill. *European Journal of Cognitive Psychology*, *11*, 473–498.

Gilhooly , K. J., Phillips, L. H., Wynn, V., Logie, R. H., & Della Sala, S. (1999b). Planning processes and age in the 5 disc Tower of London task. *Thinking and Reasoning*, *5*, 339–361.

Gilhooly, K. J., Wynn, V., Phillips, L. H., Logie, R. H., & Della Sala, S. (2002). Visuo-spatial and verbal working memory in the five-disc Tower of London task: An individual-differences approach. *Thinking and Reasoning*, *8*, 165–178.

Guilford, J. P., Christensen, P. R., Merrifield, P. R., & Wilson, R. C. (1978). *Alternate uses: Manual of instructions and interpretations*. Orange, CA: Sheridan Psychological Services.

Grice, P. (1989). *Studies in the ways of words*. Cambridge, MA: Harvard University Press.

Humphreys, G. (1951). *Thinking: An introduction to its experimental psychology*. New York: John Wiley & Sons.

Jensen, A. R. (1980). *Bias in mental testing*. London: Methuen.

Jung-Beeman, M., Bowden, E. M., Haberman, J., Frymiare, J. L., Arambel-Liu, S., Greenblatt, R. et al. (2004). Neural activity when people solve verbal problems with insight. *Public Library of Science Biology*, *4*, 1–23. http://www.plosbiology.org

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, *58*, 697–720.

Kaplan, G. A., & Simon, H. A. (1990). In search of insight. *Cognitive Psychology*, *22*, 374–419.

Knoblich, G., Ohlsson, S., Haider, H., & Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *25*, 1534–1555.

Kohler, W. (1947). *Gestalt psychology*. New York: Liveright.

Luchins, A. W. (1942). Mechanisation in problem solving: The effect of Einstellung. *Psychological Monographs*, *54*(248).

MacGregor, J. N., Ormerod, T. C., & 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*, 176–201.

Maier, N. R. F. (1931). Reasoning in humans: II. The solution of a problem and its appearance in consciousness. *Journal of Comparative Psychology*, *12*, 181–194.

Maier, N. R. F. (1933). An aspect of human reasoning. *British Journal of Psychology*, *24*, 144–155.

Martin, A., Wiggs, C. L., Lalonde, F., & Mack, C. (1994). Word retrieval to letter and semantic cues: A double dissociation in normal subjects using interference tasks. *Neuropsychologia*, *32*, 1487–1494.

Mednick, S. A. (1962). The associative basis of the creative process. *Psychological Review*, *69*, 431–436.

Metcalfe, J. (1986). Feeling of knowing in memory and problem solving. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *12*, 288–294.

Metcalfe, J., & Weibe, D. (1987). Intuition in insight and noninsight problem solving. *Memory and Cognition*, *15*, 238–246.

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

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

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

Perkins, D. N. (1981). *The mind's best work*. Cambridge, MA: Harvard University Press.

Phillips, L. H. (1997). Do "Frontal Tests" measure executive function? Issues of assessment and evidence from fluency tests. In P. Rabbitt (Ed.), *Methodology of frontal and executive function*. Hove, UK: Psychology Press.

Posner, M. I. (1973). *Cognition: An introduction*. Glenview, IL: Scott, Foresman & Co.

Raven, J. (1960). *A guide to the Standard Progressive Matrices*. London: H. K. Lewis.

Romesburg, C. H. (1984). *Cluster analysis for researchers*. Belmont, CA: Lifetime Learning Publications.

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

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22.

Snow, R. E., Kyllonen, P. C., & Marshalek, B. (1984). The topography of ability and learning correlations. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence, Vol 2*. Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate. *Behavioral and Brain Sciences*, *23*, 645–726.

Stanovich, K. E., & West, R. F. (2003). Evolutionary versus instrumental goals: How evolutionary psychology misconceives human rationality. In D. Over (Ed.), *Evolution and the psychology of thinking: The debate*. Hove, UK: Psychology Press.

Sternberg, R. J. (1987). Coping with novelty and human intelligence. In P. Morris (Ed.), *Modelling cognition*. Chichester, UK: John Wiley.

Sternberg, R. J., & Davidson, J. E. (1982). Componential analysis and componential theory. *Behavioral and Brain Sciences*, *53*, 352–353.

Thomas, J. C. Jr. (1974). An analysis of behavior in the hobbits-orcs problem. *Cognitive Psychology*, *6*, 257–269.

Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale – Revised*. New York: The Psychological Corporation.

Weisberg, R. W. (1986). *Creativity: genius and other myths*. San Francisco, CA: W. H. Freeman.

Weisberg , R. W. (1995). Prolegomena to theories of insight in problem solving: A taxonomy of problems. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight*. Cambridge, MA: MIT Press.

Wertheimer, M. (1945). *Productive thinking*. New York: Harper & Row.