Verbalization and problem solving: insight and spatial factors.

K. J. Gilhooly   E. Fioratou   N. Henretty

School of Psychology

University of Hertfordshire

Abstract

Two groups of participants attempted eight examples of each of four different problem types formed by combining insight v. non-insight and verbal v. spatial factors. The groups were given different verbalization instructions *viz*., Silent (N=40) or Direct Concurrent (N=40). There were significant differences between insight and non-insight tasks and between spatial and verbal tasks in terms of solution rates and latencies. Significant interactions between the verbal v. spatial factor and verbalization condition on solution rates and latencies reflected a greater (negative) effect of verbalizing on spatial as against verbal problems. However, no significant interactions of the insight v. non-insight factor with verbalization condition on solution rates or latencies were found. These results favoured the "business as usual" view of insight problem solving as against the "special process" view which predicted larger effects of verbalization for insight problems as against non-insight problems.

Introduction

Problem solving is a key function of the human cognitive system and considerable progress has been made in understanding how people solve well defined problems (e.g., the Tower of Hanoi tasks) in which the starting conditions, the goal and the possible actions are presented clearly and unambiguously (Egan & Greeno, 1974). Such well defined problems can be solved by heuristic search within the original representation. However, less progress has been made in understanding how people deal with tasks in which the initial way of representing the task is misleading and must be changed to permit solution. An example of this second type of problem is the 6 matchsticks task ("Given 6 matchsticks on a table make 4 equilateral triangles."). This problem typically induces an initial representation within which a solution cannot be found. Participants typically search possible configurations of dimensions is required so that the matches can be formed into a triangular based pyramid which meets the goal.

The matchstick problem is an "*Insight problem*" in which the initial representation has to be changed or "re-structured" in order that solution can be attained (Weisberg, 1995). The Towers tasks are examples of "*Non-Insight problems*" in that the initial representation is adequate to allow solution through search processes.

Explaining how re-structuring occurs remains a major challenge for cognitive theory despite a long history of experimental research from the Gestalt psychologists (e.g., Kohler, 1947) to more recent information processing approaches (e.g., Ash & Wiley, 2006; Chronicle et al., 2004; Fleck & Weisberg, 2004; Gilhooly & Murphy, 2005; Ohlsson, 1992).

3

Approaches

Two broad alternative approaches to explaining insight problem solving are currently in contention. One approach may be labelled "business as usual" and argues that re-structuring in insight problem solving occurs through small incremental and reportable steps that change the initial representation following failed attempts (Chronicle et al., 2004; Fleck & Weisberg, 2004; Perkins, 1981; Weisberg, 2006). The second approach may be labelled "special process" and argues that re-structuring requires ineffable, un-reportable processes (such as spreading activation, Ohlsson, 1992) that operate unconsciously to change the problem representation and lead to solutions which are phenomenologically sudden and surprising to the solver (Jung-Beeman, Bowden, Haberman *et al*., 2004; Kohler, 1947; Maier, 1933; Ollinger, Jones & Knoblich, 2006; Schooler, Ohlsson & Brooks, 1993).

A key method in distinguishing the "business as usual" and the "special process" approaches is to examine effects of verbalizing or "thinking aloud" (Ericsson & Simon, 1993; Gilhooly & Green, 1996) during insight problem solving (Fleck & Weisberg, 2004; Schooler, Ohlsson & Brooks, 1993). According to the "business as usual" view the processes involved in solving insight problems are as reportable as those involved in non-insight tasks and attempting to verbally report steps taken need not necessarily affect performance. However, the "special process" view holds that since the important processes are unconscious and un-reportable, attempts to verbally report steps during insight problem solving will be disruptive and interfere, through "verbal overshadowing", with the natural course of insight problem solving.

Types of verbalization

Before discussing previous findings on verbalization effects in insight tasks, it will be useful to briefly distinguish different types of verbalization or thinking aloud. Ericsson and Simon (1993) classify verbalization procedures by Time of verbalization (concurrent or retrospective) and by Type of verbalization. Three Types of verbalization are distinguished: Type 1, where information in a verbal code in focal attention is vocalised directly; Type 2, where information in focal attention but not in a verbal code is recoded into a verbal code and then vocalised, and Type 3, in which participants are asked to verbalize reasons and explanations. Ericsson and Simon (1993, p. *xvii*) argued that "Type 3 verbalization forces subjects to change their thought sequences in order to generate and verbalize overtly the information sought." Hence Type 3 verbalization is a *reactive* method, which affects processing during the task.

 Nonreactive think aloud methods are sought when the aim is to obtain verbal records (protocols) to indicate how the target task is normally carried out. On the basis of an extensive literature review, Ericsson and Simon concluded that Types 1 and 2 verbalizing (which can be labelled together, "concurrent direct verbalizing") are non-reactive in that they do not affect the type of processing people adopt, but they may cause some slowing effects. We have also found such non-reactive effects in studies of non-insight problem solving (Gilhooly et al., 1997; Gilhooly et al., 1999).

Previous studies

Schooler, Ohlsson and Brooks (1993) reported that the effect of verbalization was to depress significantly solution rates on insight problems compared to silent controls but that no effect of verbalisation was found with non-insight problems. Schooler *et al*. (1993) interpreted their results as indicating that special unreportable processes were required to solve insight tasks and that verbalizing – even when designed to be minimally reactive – interfered by biasing processing into verbalizable conscious forms.

Although the study has been very influential, a number of aspects of Schooler *et al*.'s (1993) study are problematic. It may be noted that in Schooler *et al*.'s (1993) key Experiments 3 and 4, the four non-insight problems were predominantly verbal in character while at least two of the three insight problems used (triangle of coins and the rope problem) could be regarded as having a large spatial element. Verbalization effects would be expected to be stronger for spatial problems than for verbal problems because verbalization would tend to induce a switch from more appropriate spatial coding to a less appropriate verbal coding. Hence, Schooler *et al*.'s (1993) results could be at least partly due to a confounding of the insight v. non-insight factor with the verbal v. spatial factor.

The exact instructions used for verbalising might also be a factor in Schooler et al.'s (1993) finding. Ericsson and Simon (1993, p.*xxx*) noted that Schooler *et al*.'s (1993) verbalization instructions in the key experiments on effects of verbalising were not quite standard for concurrent direct verbalization. In Schooler *et al*.'s (1993) Experiments 3 and 4 in which the think aloud instructions were intended to lead to direct concurrent reporting, participants were told to verbalize "anything you read, questions you ask yourself and so forth" and it may be that participants were thus cued to read the problem statement more often than

6

control participants and hence the initial interpretation might become stronger and more resistant to change.

Fleck and Weisberg (2004) pointed out that the training given for thinking aloud by Schooler *et al* (1993) was very brief and the only practice task given to familiarise participants with direct concurrent think aloud was a non-insight task. Thus, think aloud training may have been insufficient to overcome pre-existing tendencies to explain and justify solution attempts; and use of a non-insight practice problem may have biased participants to treat the subsequent insight problems as if they were non-insight tasks. Some support for these views emerged from Fleck and Weisberg's (2004) protocol analysis study of a single insight task, *viz.*, Duncker's (1945) candle problem. Using instructions close to Ericsson and Simon's ideal for non-reactive verbalizing and more extensive think aloud training than Schooler *et al.* (1993), Fleck and Weisberg (2004) found no effect of direct concurrent think aloud (N = 34) v. silent controls (N = 18) for the candle problem. Weisberg (2006, p.334) also reported no effects of direct concurrent think aloud in a study of three insight tasks with Ns of 55 in both control and experimental groups.

Finally, the number of tasks exemplifying insight and non-insight problems in Schooler *et al.*'s (1993) study was relatively small (three insight v. four non-insight) and this raises questions about the representativeness of the task sample and the reliability of composite scores based on the example items.

7

Present Study

 Although the possible existence of verbal overshadowing in insight tasks is important for establishing the nature of the processing underlying insight problem solving, and for assessing the utility of think aloud methods in studying insight problem solving, the evidence base is problematic, in that verbalization effects had been studied in very few insight problems (three in Schooler *et al.*'s (1993) main studies and one in Fleck & Weisberg (2004)); it was not completely clear whether Schooler *et al.*'s (1993) verbalization instructions were truly direct concurrent (Type 1 or 2) or Type 3 and there was a possible confounding of verbal/spatial factors with insight/non-insight. Hence, the present study examined possible effects of verbalizing using clearly direct concurrent verbalization methods and a considerably larger set of problems than hitherto. Moreover, the problem set used in the present study included spatial and verbal insight and non-insight problems as a check on whether effects may be stronger for spatial problems. Eight problems representing each combination of insight/non-insight and spatial/verbal factors were used, giving 32 problems in total. Verbalization instructions followed Ericsson and Simon's (1993) guidelines and appropriate levels of training in verbalization procedures were given. Overall, the "business as usual view" would predict no effect of direct concurrent verbalizing on insight or non-insight tasks; the "special process" view of insight would predict impairing effects of direct concurrent verbalising on insight problems but no effect of direct concurrent verbalising on non-insight tasks. The experiment was a mixed design with one between factor (Verbalization instructions) and two within factors (Insight v. Non-insight; Verbal v. Spatial).

Method

Materials

8 <u>verbal insight</u> problems viz., *Ship* (At noon a ship's porthole is 4 metres above the waterline. The tide rises at 1 metre per hour. How long will it take the water to reach the porthole?), *Fan* (Joe has no psychic powers but he can tell you the score in any football game before it starts. How?), *Socks* (There are of black and brown socks in a drawer mixed in 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?) , *Hole* (How much earth is there in a hole 2m by 3m by 2m?), *Lake* (Someone walked for 20 mins. on the surface of a lake without sinking but without any floatation aid. How?) , *Lilies* (The Lilies in a lake double in number everyday. The lake will be covered in 60 days. In how many days will it be half covered?) , *Horse trading* (A man buys a horse for £60, sells it for £70, buys it back for £80 and finally sells it again for £90. How much has he made?), *Dark* (A man is reading a book when the lights go off. Although the room is pitch dark the man goes on reading. How?).

8 <u>spatial insight</u> problems viz., *Triangle of coins* (Given coins in a triangle shape with 1, 2, 3, 4 coins in the rows, move 3 coins so that the triangle faces the other way.), *Pigpen* (9 pigs are kept in a square pen. Build 2 more square enclosures that would put each pig in a pen by itself.) , *6 matches* (see above), *Cake problem* (Given a circular cake can you cut it into 8 equal pieces using only 3 straight cuts?), *6 coins* (Given 2 rows of 3 coins in 3 moves make a circle of coins), *Rope* (A prisoner divides a rope in half and escapes from a high tower. How?), *8 coins* (transform an arrangement of 8 coins into one where each coin touches exactly 3 others), *Farm* (How could you divide an L-shaped piece of land into 4 equally shaped pieces of equal sizes?), *Cherry* (Given 4 matches arranged to represent a glass and a dot representing a cherry in the glass, move 2 matches so cherry is outside the glass).

8 <u>verbal non-insight</u> problems viz., *Suspects* (Given statements by 4 suspects infer which one committed the crime), *Plan errands* (Devise optimal plan to complete errands given travelling times and opening hours of various shops)*, Dinner party* (Given 5 guests with specified food aversions and a list of foods, make up a menu all could eat.), *Couples* (Given information about what colors members of couples were wearing, infer what color a specific individual is wearing), *Anagrams* (unscramble 8 five-letter single solution anagrams), *Cards* (Given 3 cards on table face down and minimal information work out which suit each card is), *Water* (Given containers of varying sizes get a specified quantity of water), *Flowers* (Given limited information about flowers given from four male to four female partners deduce who gave which flowers to whom.)

8 <u>spatial non-insight</u> problems viz., *Hobbits & Orcs* (Given a boat that can only hold 2 creatures, how can you get 3 hobbits and 3 orcs across a river in such a way that the hobbits are never outnumbered in minimum moves), *Ward-Allport Tower of London* (Manipulate 5 differently coloured discs on equal size pegs to match target configuration in minimum moves) , *4-disc Tower of Hanoi*, *Matrices* (Ravens matrices items), *Cards* (Given limited information deduce spatial layout of cards on a table), *Heavy/light coins* (Given 4 coins of which 2 are slightly light and 2 are slightly heavy, find out which are which in 2 weighings on a balance scale), *Peg solitaire* (Remove as many pegs as possible; pegs jumped over are removed), *Wolf and chicken* (Move a wolf, chicken and cabbage from one side of river to another subject to constraints.).

Participants: 80 students at University of Hertfordshire. (36 male; 44 female; Mean age =22.01 yrs, SD = 3.70 yrs). Participants were paid £7 per hour for 2 x 2hrs sessions.

Procedure: Participants were assigned randomly to a Concurrent Direct think aloud group (N = 40) or to Silent Working Control group (N = 40).

Problems were presented in random orders. Verbalizations were digitally recorded for protocol analyses which will be reported in a separate paper.

Direct concurrent think aloud instructions were as follows-

"In this experiment we are interested in what you think about when you find answers to some problems that I am going to ask you to answer. In order to do this I am going to ask you to **think aloud** as you work on each problem. What I mean by think aloud is that I want you to tell me **everything** you are thinking from the time you first see the question until you give an answer. I would like you to talk aloud **constantly** from the time I present each problem until you have given your final answer to the question. **I don't want you to try to plan out what you say or try to explain to me what you are saying.** Just act as if you are alone in the room speaking to yourself. It is most important that you keep talking. If you are silent for any long period of time I will ask you to talk. Do you understand what I want you to do?" Any questions were then answered.

"Good, now we will begin with some practice problems."

The order of practice problems was randomised. Two practice tasks were insight problems and two were non-insight tasks.

Practice problems:

1. I want you to multiply two numbers in your head and tell me what you are thinking as you get an answer. What is the result of multiplying 24 x 15?

2. How many windows are there in your parents' house?

3. Two men play 5 games of chess and each wins an even number of games without any ties. How could that be?

4. A woman didn't have any driving insurance. She didn't stop at a railway crossing and went the wrong way down a one way street for hundreds of yards. A policeman saw all this but did nothing. Why?"

A prompt was given after 15s silence "Please keep talking".

A prompt was given after 15s silence "Please keep talking and explain what you are thinking and why".

All participants were tested individually and attempted all tasks (presented in random orders). Experimenter confined feedback for proposed solutions by saying only "Yes, that is the solution" or "No, that is not the solution. Please keep trying."

A maximum time of 4 mins was allowed per problem.

Results

Solution rates by problems

Since averaging data over problems might obscure differences among problems and to identify tasks that may display floor or ceiling effects we first analysed average solution rates within 4 mins for each problem by verbalization condition, as shown in Table 1 below.

INSERT TABLE 1 ABOUT HERE

From Table 1 it can be seen that out of 32 F- test comparisons between solution rates with and without verbalising, only three (Lilies, Socks, Raven) were significant at the .05 level (2

tail). Even if the criterion is relaxed to a 1 tail significance, which is justifiable as the overshadowing hypothesis is directional, at least for insight problems, only one more comparison becomes significant (Cheap Necklace). A simple tally of number of comparisons for each problem type in the direction predicted by the overshadowing hypothesis out of eight indicates that for Verbal Insight the tally was 2/8, for Spatial Insight it was 7/8, for Verbal Non-Insight it was 1/8 and for Spatial Non-Insight it was 7/8. Thus, a verbal overshadowing effect was present for 14/16 of spatial tasks and for 3/16 verbal tasks, which yielded a $\chi^2(1) = 15.18$, $p < 0.001$, $\varphi = .69$. However, a verbal overshadowing effect was only present for 9/16 insight and 8/16 non-insight tasks indicating no significant association of verbal overshadowing with whether a task was insight or not ($\chi^2(1) = 0.12$, ns, $\varphi = .06$). Overall, these tallies suggest that the verbalisation effect is largely confined to Spatial problems and is not affected by whether the problems are of the Insight type or not.

Solution rates by problem types

We combined the problem scores into composite averages over examples of each type. Since a few problems showed floor effects (Cheap Necklace, Coins, Matches, Tower of Hanoi) or ceiling effects (Horse, Couples,) these were omitted from the composite scores per problem types. A problem was considered to show a floor effect if the control solution rate was less than .20 and a ceiling effect if the control solution rate was greater than .80. The resulting average solution rates per problem type are shown in Table 2.

INSERT TABLE 2 ABOUT HERE

A 2x2x2 factorial Anova with two within factors (Insight/Non-insight and Spatial/Verbal), and one between factor (Verbalization condition: Silent and Direct concurrent verbalization)

was carried out with solution probability scores averaged over problem types as the dependent variable. The Anova results are shown in Table 3.

INSERT TABLE 3 ABOUT HERE

From Table 3 it can be seen that there were significant differences between Insight and Non-Insight problems and between Verbal and Spatial problems in average difficulty levels. Overall, in this set of problems, the Non-insight tasks and the Spatial tasks were somewhat more difficult than the Insight and the Verbal tasks respectively. However, the results indicated that there was no main effect of Verbalization condition and, importantly, *no interaction* effect between the Insight factor (Insight v Non-insight) and Verbalization condition on solution rates. Thus, these results replicate the lack of verbalization effects on insight tasks reported by Fleck & Weisberg (2004) and Weisberg (2006, p.334) and counter the finding of such effects by Schooler *et al*. (1993). There was a significant interaction effect between the Verbal/Spatial problem factor and Verbalization condition on solution rates ($F(1,78) = 4.57$, $p < 0.05$). The pattern of the interaction shown in Fig. 1 suggests that the spatial problems were somewhat impaired by verbalizing but verbal problems were somewhat aided by verbalizing.

INSERT FIG. 1 ABOUT HERE.

Latencies by problems

As latencies may be regarded as more sensitive measures of solving success than correct/incorrect measures (Gilhooly & Murphy, 2005), analyses were also carried out on latencies (non-solutions = 240s). Latencies per problem type and verbalization condition are shown in Table 4 below.

INSERT TABLE 4 ABOUT HERE

From Table 4 it can be seen that out of 32 *F* test comparisons between solution rates with

and without verbalising, only one (Dinner) was significant at the .05 level (2 tail) and this

result was *opposite* in direction to the overshadowing hypothesis. Even if the criterion is

relaxed to 1 tail significance at the .05 level, only one more comparison becomes significant

(Cheap Necklace) and this result is in the direction predicted by the overshadowing

hypothesis. A simple tally of number of comparisons for each problem type in the direction

predicted by the overshadowing hypothesis out of eight indicates that for Verbal Insight the

tally was 2/8, for Spatial Insight it was 8/8, for Verbal Non-Insight it was 1/8 and for Spatial

Non-Insight it was 7/8. Thus, a verbal overshadowing effect was present for 15/16 of spatial

tasks and for 3/16 verbal tasks, which indicated a $\chi^2(1) = 18.28$, *p* <0.001, $\varphi = .76$. However

a verbal overshadowing effect was only present for 10/16 insight and 8/16 non-insight tasks

indicating no significant association of verbal overshadowing with whether a task was

insight or not ($\chi^2(1) = 0.51$, *ns*, $\varphi= .13$). Overall, these tallies suggest that the verbalisation

effect is largely confined to Spatial problems and is not affected by whether the problems are

of the Insight type or not. These tallies, as with the solution rate analysis, suggest again that

the verbalisation effect is largely confined to Spatial problems and is not affected by whether

the problems are of the Insight type or not.

 Latencies by problem types

The average latencies per problem type and verbalisation condition are shown in Table 5.

INSERT TABLE 5 ABOUT HERE

A 2x2x2 factorial Anova with two within factors (Insight/Non-insight and Spatial/Verbal), and one between factor (Verbalization condition: Silent and Direct concurrent verbalization) was carried out with latencies averaged over problem types as the dependent variable. The results are shown in Table 6.

INSERT TABLE 6 ABOUT HERE

From Table 6 it can be seen that there were significant differences between Insight and Non-Insight problems and between Verbal and Spatial problems in average latency scores. Overall, in this set of problems, the Non-insight tasks and the Spatial tasks scored higher on the latency measuret than the Insight and the Verbal tasks respectively. The results in Table 6 also indicate that, as with the solution data, there was no main effect of Verbalization condition and no interaction effect between the Insight factor (Insight v Non-insight) and Verbalization condition on latencies. Thus, these results again replicate the lack of verbalization effects on insight tasks reported by Fleck & Weisberg (2004) and Weisberg (2006, p.334) and counter the hypothesis that verbalisation effects would be particularly strong for insight latencies. There was an interaction effect between the Verbal/Spatial problem factor and Verbalization condition on latencies (F (1,78) = 6.20, $p < 0.01$). The interaction pattern shown in Fig. 2 suggests that spatial problems were slowed to a greater extent by verbalizing than were verbal problems.

INSERT FIG.2 ABOUT HERE

Analysis of simple effects of verbalization condition on latencies found that for spatial tasks, $F(1,78) = 3.67, p < 0.05$, but for verbal tasks, $F(1,78) = 2.78$, *ns*.

16

Discussion

On the "special process" view of insight put forward by Schooler *et al*. (1993), Ohlsson (1992), Jung-Beeman et al. (2004) and others, according to which insight problem solving involves un-reportable processes, we would have expected to find a significant interaction of insight v. non-insight problem type with verbalization condition, with particularly negative effects of Concurrent Direct verbalization on insight tasks. On the "business as usual" view put forward by Perkins (1981), Fleck and Weisberg (2004) and others, it would be expected that direct concurrent verbalization would not differentially affect insight *v*. non-insight problems.

Overall, the lack of interactions found in the present study between the Insight factor and Verbalization condition on solution rates and latencies supports the "business as usual" view and does not support the "special process" view that un-verbalizable processes are particularly involved in insight problem solving. It may be argued that the present report is relying on a null result and that such a result may be due to lack of statistical power in the study reported here. One may be note that Schooler *et al*.'s (1993) key interaction results showed a 35% difference between solution rates for insight tasks under verbalising v. silent conditions. This difference in proportions translates into an effect size (*h*) of .54 which is a medium sized effect (Cohen, 1988). The interaction of Problem Type X Verbalisation, $F(1,38) = 5.22$, reported by Schooler *et al*. (1993, p.174) represents a partial eta^2 of 0.12, which again translates into a medium sized effect. Thus, Schooler *et al*. (1993) obtained a significant medium sized Problem Type X Verbalisation interaction with group Ns of 20. The present study was more powerful than Schooler *et al*. (1993), in that its group Ns were 40 rather than 20. Furthermore, the composite measures used were more reliable, being based on 8 items each as against the 3 or 4 used by Schooler *et al*. (1993), which increases

17

the power of the present study. Overall, the absence of a Problem Type X Verbalisation

interaction in the present study cannot be easily attributed to lack of power as compared to

Schooler *et al*. (1993). It can also be noted that the analyses on individual problems found

little sign of significant verbal overshadowing even with multiple one tail significance

testing over 64 comparisons (32 problems x two performance measures yielded 5

comparisons in the predicted direction, significant at $p < 0.05$, one-tail).

 We suggest that the previous results which indicated verbal overshadowing in insight

problem solving were largely due to confounding of insight and spatial task factors. Thus,

the significant interactions of the Verbal *v*. Spatial problem factor with Verbalization on

solution rates and latencies in the present studies are consistent with the view that previous

reports of impairing effects of verbalizing in insight tasks (Schooler *et al*., 1993) may have

reflected a confounding of insight tasks with spatial tasks; such tasks are generally impaired

by verbalization because of the need to re-code spatial information into verbal form for

reporting purposes (Ericsson and Simon, 1993). An examination of the degree to which

individual problems matched the predictions of overshadowing indicated that such

predictions are generally accurate in direction for spatial tasks but not for verbal tasks

irrespective of whether the tasks involved insight or not.

In terms of theoretical issues regarding insight, the present results suggest that insight

problem solving does not depend on processes that are disrupted by verbalisation. This

conclusion offers support to the usefulness of verbal protocol methods in studying insight

problem solving. The exact processes involved in restructuring under the "business as usual"

view need further clarification. Explicit heuristic search processes aimed at changing the

problem space as suggested by Kaplan and Simon (1990) are candidates for further research.

For example, solvers could deliberately decide to examine each word in a verbal insight

problem for ambiguity in order to seek an alternative interpretation. Such explicit processes

would be expected to involve the central executive of working memory and support for the

role of executive processes has been found in a number of individual difference studies of

insight problem solving (Gilhooly & Murphy, 2005; Murray & Byrne, 2005; Fleck, 2008;

Ash & Wiley, 2006).

References

Ash, I.K. & Wiley, J. (2006). The nature of re-structuring in insight: an individual

differences approach. *Psychonomic Bulletin and Review*, **13**, 66-73.

Chronicle, E., 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 & Cognition, **30**, 14-27.

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

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.

Ericsson, K.A. & Simon, H.A. (1993). *Protocol analysis: Verbal reports as data*.

Cambridge, MA: MIT Press.

Fleck, J.I. (2008). Working memory demands in insight v. analytic problem solving.

European Journal of Cognitive Psychology, **20**, 139-176.

Fleck, J.I. & Weisberg, R. (2004). The use of verbal protocols as data: an analysis of insight

in the candle problem. *Memory & Cognition*, **32**, 990-1006.

Gilhooly, K J. and Green, C. (1996). Protocol analysis: theoretical background. In Richardson, J (Ed.), *Handbook of qualitative research methods for psychology and the social sciences*. Leicester: BPS Books.

Gilhooly, K. J., McGeorge, P., Hunter, J., Rawles, J. M., Kirby, I., Green, C., & Wynn, V. (1997). Biomedical knowledge in diagnostic thinking: the case of electrocardiogram (ECG) interpretation. *European Journal of Cognitive Psychology, 9,* 199-223.

Gilhooly, K. J. & Murphy, P. (2005). Differentiating insight from non-insight problems. *Thinking & Reasoning*, **11**, 279-302.

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

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

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

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

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

Murray, M.A. & Byrne, R.M.J. (2005). Attention and working memory in insight problem solving. *Proceedings of Cognitive Science Society*, **27**, 1571-1575.

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.

Ollinger, M. Jones, G. & Knoblich, G. (2006). Heuristics and representational change in two-move matchstick arithmetic tasks. *Advances in Cognitive Psychology*, **2**, 239-253.

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

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

Weisberg, R. (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.

Weisberg, R. (2006). *Creativity*. Hoboken, NJ: Wiley.

Table 1: Mean solution rates over 4 minute trials by problem and verbalization condition, inferential statistics (F, exact 2-tail p and partial eta^2) for verbalisation effects. * = p<.05, 2 tail; [a] = p<.05, 2 tail.

Problem	Silent	Think aloud	$F(1,78)$	2 tail p	Part. Eta^2
Verbal Insight					
Earth	.73	.78	0.26	.61	.003
Football	.38	.45	0.46	.50	.006
Horse	.85	.90	0.45	.51	.006
Lake	.70	.78	0.57	.45	.007
Lilies	.70	.48	4.29	.04*	.052
Ocean	.53	.65	1.28	.26	.016
Reading	.73	.75	0.06	.80	.001
Socks	.80	.60	3.90	.05*	.048
Spatial Insight					
Cake	.50	.35	1.84	.18	.023
Cherry	.43	.33	0.84	.36	.011
Cheap necklace	.13	.03	2.92	.09 [a]	.036
Coins	.13	.10	0.12	.73	.002
Farm	.53	.63	0.81	.37	.010
Matches	.05	.00	2.05	.16	.026
Pigpen	.33	.23	0.99	.32	.013
Triangle	.65	.58	0.46	.50	.006

Table 1 (Continued): Mean solution probabilities over 4 minute trials by problem and

verbalization condition, inferential statistics (F, exact 2-tail p and partial eta^2) for

verbalisation effects.

Problem	Silent	Think aloud	$F(1,78)$	2 tail p	Part. Eta^2
Verbal Non-Insight					
Bachelors	.33	.38	0.22	.64	.003
Couples	.88	.85	0.10	.75	.001
Day	.70	.78	0.57	.45	.007
Dinner	1.00	1.00	----	---	---
Flower	.70	.80	1.05	.31	.013
Pint	.70	.70	0.00	1.00	.000
Suspects	.40.	.55	1.80	.18	.023
Anagrams	.64	.69	0.47	.68	.002
Spatial Non-Insight					
Heavy/light coins	.58	.55	0.05	.82	.001
Peg	.59	.47	2.28	.13	.029
Cards	.50	.53	0.05	.83	.001
Tower of Hanoi	.18	.18	0.00	1.00	.000
Wolf	.80	.65	2.23	.14	.028
Latin	.65	.59	1.21	.28	.015
Raven	.54	.41	4.81	.03*	.058
Tower of London	.44	.55	2.41	.13	.030

Table 2: Mean solution probabilities over 4 minute trials by problem type and verbalization condition, averaged over problems per type. (SDs in brackets).

	PROBLEM TYPE			
	Insight		Non-insight	
	Verbal	Spatial	Verbal	Spatial
VERBALIZATION				
Silent (N=40)	.65 (.24)	.49 (.31)	.54 (.25)	.41 (.22)
Think aloud (N=40)	.64 (.25)	.42 (.24)	.61 (.22)	.35 (.16)

Table 3. ANOVA on solution scores. N=80. Key: Vb = Verbalization condition; Nonins.= Non-insight; Ins = Insight; Verb = Verbal; Spat = Spatial; P. = Partial. * =p <.05. ** = p<.01

Source	SS	df	MS	F	P. eta^2
Vb. Condition	.02	1	.02	.15	.002
Error Vb	12.37	78	.15		
Insight/Nonins	.37	1	.37	17.79**	.186
Insight X Vb	.03	1	.03	1.63	.020
Error Ins	1.60	78	.02		
Verb/spatial	2.97	1	2.97	82.28**	.513
Verb/spatial X Vb	.13	1	.13	4.57*	.055
Error Verb/spatial	2.81	78	.04		
Ins X Verb/spatial	.00	1	.00	0.02	.000
Ins X Verb/spat X Vb	.03	1	.03	1.67	.021
Error	1.24	78	.02		

Table 4: Mean latencies in secs. by problem and verbalization condition, inferential statistics

(F, exact 2-tail p and partial eta^2) for verbalisation effects. * = p< .05, 2 tail; ^{a} = p<.05, 1 tail.

N=40.

Problem	Silent	Think aloud	$F(1,78)$	2 tail p	Part. Eta^2
Verbal Insight					
Earth	98.0	94.9	0.02	.88	.000
Football	160.8	151.4	0.16	.69	.002
Horse	115.9	92.9	1.91	.17	.024
Lake	93.2	82.2	0.26	.61	.003
Lilies	137.0	161.7	1.48	.23	.019
Ocean	157.6	151.6	0.09	.75	.001
Reading	103.5	96.5	0.11	.74	.001
Socks	126.4	152.8	2.09	.15	.026
Spatial Insight					
Cake	174.4	196.8	1.56	.21	.020
Cherry	178.3	198.2	1.39	.24	.018
Cheap necklace	235.1	239.9	3.23	.08 ^{a}	.040
Coins	228.3	229.9	0.04	.84	.001
Farm	177.05	179.7	0.03	.87	.000
Matches	236.3	240.0	1.82	.18	.023
Pigpen	201.9	216.3	1.25	.27	.016
Triangle	149.7	175.5	2.08	.15	.026

Table 4 (Continued): Mean latencies in secs. by problem and verbalization condition, inferential statistics (F, exact 2-tail p and partial eta^2) for verbalisation effects.* = p<.05. N =40.

Problem	Silent	Think aloud	$F(1,78)$	2 tail p	Part. Eta^2
Verbal Non-Insight					
Bachelors	222.4	217.9	0.39	.53	.005
Couples	79.2	65.9	2.69	.11	.033
Day	90.4	77.5	1.69	.19	.021
Dinner	87.2	70.2	5.49	.02*	.066
Flower	166.9	153.9	1.14	.29	.014
Pint	147.3	140.1	.14	.71	.002
Suspects	211.5.	196.9	.14	.71	.002
Anagrams	116.2	119.3	1.34	.25	.017
Spatial Non-Insight					
Heavy/light coins	183.9	180.9	0.04	.84	.001
Peg	134.8	145.9	0.84	.36	.011
Cards	186.4	200.5	1.11	.29	.014
Tower of Hanoi	126.0	128.28	0.03	.87	.000
Wolf	125.9	150.8	2.13	.15	.027
Latin	28.6	25.12	1.65	.20	.021
Raven	36.7	38.69	.61	.44	.008
Tower of London	31.56	35.32	2.30	.13	.029

Table 5: Mean latencies in s., by problem type and verbalization condition, averaged over 8 problems per type. (SDs in brackets.).

PROBLEM TYPE.

	Insight		Non-insight	
	Verbal	Spatial	Verbal	Spatial
VERBALIZATION				
Silent (N=40)	124.08	176.25	153.44	142.06
	(52.59)	(35.65)	(30.77)	(28.63)
Think aloud (N=40)	123.01	193.25	145.17	149.63
	(49.92)	(22.61)	(23.68)	(30.96)

Table 6. ANOVA on latency scores. N=80. Key: Vb = Verbalization condition; Nonins.= Non-insight; Ins = Insight; Verb = Verbal; Spat = Spatial. P. = partial. * = p<.05. **=p<.01.

| Source	SS	df	MS	F	P. eta^2
Vb. Condition	512.39	1	512.39	.16	.002
Error Vb	255405.3	78	3274.44		
Insight/Nonins	20404.47	1	20404.47	30.73**	.283
Insight X Vb	666.51	1	666.51	.1.01	.013
Error Ins	151785.96	78	663.92		
Verb/spatial	117206.66	1	117206.66	175.29**	.692
Verb/spatial X Vb	4148.46	1	4148.46	6.20**	.074
Error Verb/spatial	52152.93	78	668.63		
Ins X Verb/spatial	139320.92	1	139320.92	250.54**	.763
Ins X Verb/spat X Vb	41.63	1	41.63	.07	.001
Error	43374.76	78	516.09		
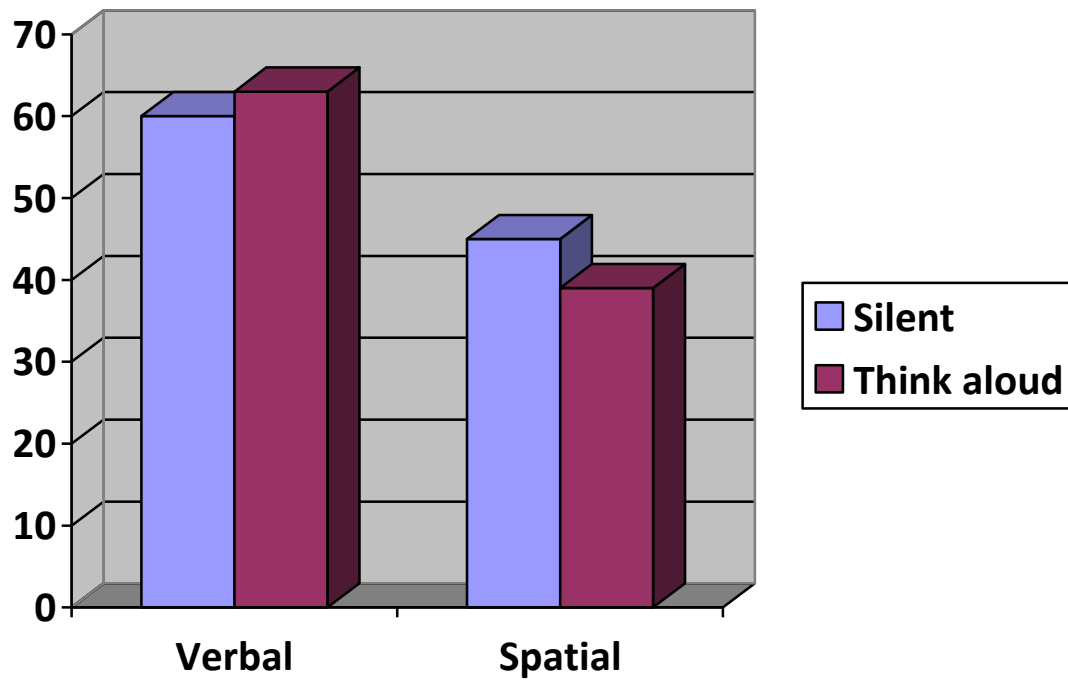
Fig.1 Mean percentage of correctly solved verbal and spatial problems under Think aloud and Silent conditions.
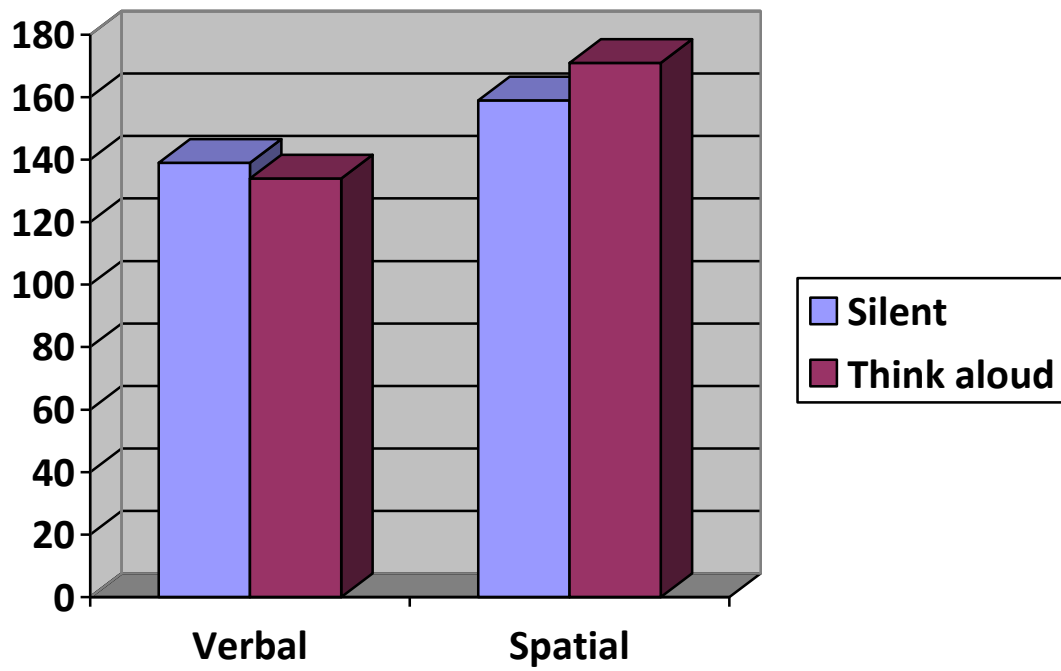
Fig. 2. Average latencies (seconds) of verbal and spatial problems under Think aloud and Silent conditions.