Towards the Relation between Language and Thinking – the Influence of Language on Problem-Solving and Memory Capacities in Working on a Non-Verbal Complex Task

Christina Bartl (christina.bartl@ppp.uni-bamberg.de)
Lehrstuhl Psychologie II; Otto-Friedrich-Universität
D-96045 Bamberg, Germany

Abstract

This study focuses on the „classical" topic of the relation between language and thinking. Empirical studies investigating the interaction of verbalization and problem solving show inconsistent results. Studies differ with respect to the instruction of verbalization and the characteristics of the task. The aim of the study is to compare the performance in a non-verbal problem with and without language. For this purpose we investigate the performance of six groups of subjects working under different conditions: some of them were disturbed in their language behavior, others were encouraged to verbalize. It could be shown that though they had to work on a non-verbal problem, subjects disturbed in their linguistic behavior showed a worse performance than controls. Furthermore it can be shown that „thinking aloud" in itself does not guarantee an improvement of performance. Moreover there are specific aspects of thinking aloud supporting problem solving. Case studies reveal interesting results with respect to the specific structure of „helpful" verbalization. The differences found cannot be explained by different memory loads or by the degree of distraction.

Introduction

When children work on a difficult task you can observe them talking to themselves. Verbalization seems to help to carry out activities which are close to the maximum a child can achieve according to his or her developmental state (Diaz & Berk, 1992). Several methods in psychotherapy make use of the positive effect of self-verbalization on action regulation. Meichenbaum & Goodman (1971) developed a program for impulsive children. They trained them to talk to themselves while working on cognitive or motoric tasks. The verbalization, first aloud, later whispering and finally tacitly can improve problem solving performance and help impulsive children to control themselves. Similar observations can be made when subjects work on complex problems: being in a crucial situation, they spontaneously begin to verbalize aspects of the problem or their own behavior. Although we are not yet really sure which specific aspects of verbalization facilitate the organization of cognitive processes, the need to "talk to oneself" while solving problems seems to be almost irresistable.

Experiments concerning the treatment of "thinking aloud" and "self-verbalization" however do not give clear evidence. On the one hand there are a lot of studies which support the hypothesis that thinking aloud can improve problem solving behavior in non-verbal tasks, such as the Tower of Hanoi, Raven's Standard Progressive Matrices or other non-verbal intelligence tasks (see Berry & Broadbent, 1984; Franzen & Merz, 1976; Gagné & Smith, 1963; Hussy, 1987). On the other hand, studies give indication for a contrary assumption. Phelan (1965) forced subjects to think aloud while working on concept learning tasks and found negative effects on their performance. Several studies indicate no differences in performance between a verbalizing group and controls (e.g. Deffner, 1989; Lass, Klettke, Lüer & Ruhlender, 1991; Rhenius & Heydemann, 1984).

How can we explain such inconsistencies? Looking exactly upon the experimental designs of the cited studies, the hypothesis arises that differences in the instruction of "thinking aloud" lead to contradictory results. According to the model of Ericsson & Simon (1980, 1984), verbalization can only improve performance when it contents aspects not included in working memory before. So meta-cognitive processes such as self-explanation or self-reflection could modify and improve problem solving strategies. The study of Chi et al. (1989) could demonstrate that better achievers have more spontaneous utterances reflecting and evaluating their own behavior. Other studies show that treatments enforcing meta-cognitive processes have positive effects on subjects' problem solving capacities (Beradi-Colletta et. al., 1995; Chi et al., 1994; Tisdale, 1998).

Another reason for the different results could be found in the experimental designs. Subjects instructed to think aloud were compared with controls without any specific instruction. One could argue that the controls also verbalized, either loudly or silently. Obviously it is necessary to control, whether the subjects indeed refrained from verbalizing. One method to suppress verbalization is the technique of "articulatory suppression". In this treatment subjects have to carry out speaking routine tasks, so that their motoric speech system is permanently busy. Yet studies in the context of Baddeley's phonological loop model (Baddeley, 1990) show, that articulatory suppression has negative effects on subjects' memory loads. The influence of articulatory suppression on problem solving processes has not been investigated yet.

The aim of this study is to explain the reasons for the different findings. Firstly we investigate if an instruction of self-explication and self-evaluation has stronger positive effects on problem solving than the unspecific instruction to think aloud. Secondly we test the hypothesis that suppression of the internal dialogue leads to a worse problem solving performance.

Besides the impacts of different language treatments on problem solving performance, we will have a look at possible influence of the treatments on attention and on memory for elements of the problem space.

Method

Subjects

The experiment was carried out with 60 subjects, 32 females and 28 males, who were distributed to six experimental groups by random. The age varied between 14 and 40 years with a mean of 23 years. Most of the subjects (n = 54) were students of different fields of the universities Bamberg, Erlangen and Trier. Furthermore one pupil and five postgraduated persons joined the experiment.

Task

In this study we used the computer-aided scenario "BioBeetle", which can be classified as an interpolation problem (Dörner, 1976) in the tradition of the Tower of Hanoi. In the "BioBeetle" scenario, subjects have to transform a given initial state into a target state by using given operators in a specific sequence. The scenario is semantically embedded as a laboratory for breeding beetles where the morphology of beetles can be transformed by radiation. The beetles differ in six characteristics (color, shape, form of feet, eyes, tentacles and mouth). Each of the characteristics can have two values (p.e. the color can be red or green), so that 26 animals can be distinguished. The morphology of the beetles can be transformed by twelve kinds of radiation (= operators). Several radiations only change one characteristic of the beetle, others affect several morphological attributes. Furthermore some radiations need certain conditions to work, so that the subjects first have to achieve sub-ordinate targets for using them.

All information necessary to solve the breeding problems is given in pictures. Starting the simulation, the subjects see on the left side of the screen the initial beetle and on the right side the target state (see figure 1).

In the left upper corner the actual state ("aktueller Zustand") of the beetle is visualized. In the right corner there is the target state ("Ziel"). In order to transform the beetle, subjects have to press the buttons named with Greek letters. Clicking on the buttons named with Greek letters, the subjects get information about the prerequisits and the effects of the chosen radiation. Then they can decide whether or not they want to apply it. If the necessary conditions for a radiation are not given, nothing will happen. After a successful modification, the actual state of the beetle appears on the screen.

The information about the operators is also given in pictures (see figure 2). As a result, the subjects always know the actual state of the beetle, the target state and the impact of any possible measure. Their problem is to apply the operators in the right sequence in order to attain all the characteristics of the target state. Looking at figure 2 the first line ("IN") represents the prerequisits. The radiation is not applicable unless the actual state of the beetle has the characteristics shown in the first line. When all prerequisits were given and the radiation is applied the beetle will change its morphology the way shown in the line below ("OUT"). In this case the radiation changes the color, if the beetle has an oval shape and long tentacles.

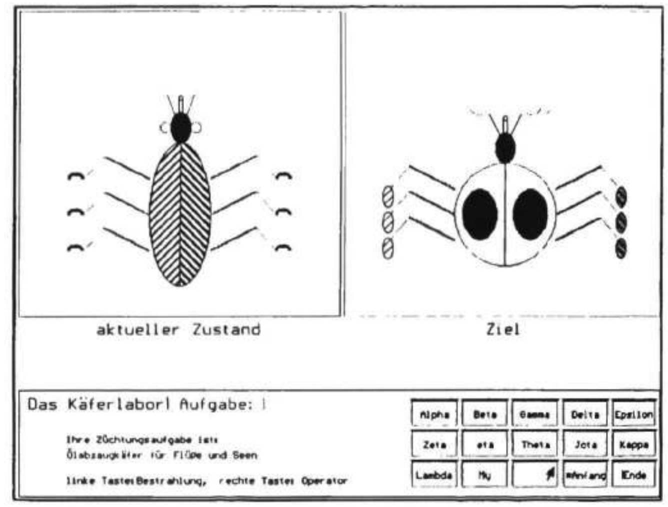

Figure 1: Surface of the computer-aided task "BioBettle".

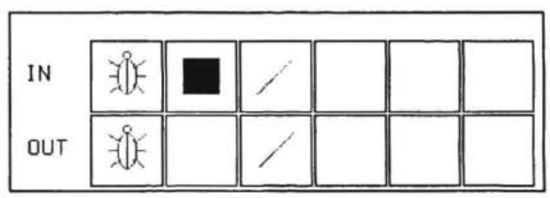

Figure 2: Pictoral information about the radiations effect.

Treatment

The subjects worked on the "BioBeetle" problem under six different conditions (10 subjects in each group). One part of the sample was encouraged in their verbal behavior by different treatments (treatment 1: "thinking aloud" and treatment 2: "self-reflection"). Others were disturbed in their inner dialogue (treatment 3: "no verbalization" and treatment 4: "articulatory suppression"). In order to determine whether an influence of articulatory suppression can be attributed to the specific verbal interference or to an unspecific impairment of attention, another group of subjects carried out a comparable motoric, but non-verbal routine task (treatment 5: motoric interference) while solving the problem. A group of controls completes the experimental design. The experimental treatments are described below:

Treatment 1 ("thinking aloud"): Before they started to work on the BioBeetle, these subjects got the instruction to verbalize their thoughts, that is to vocalize everything that would come into their mind. They were reminded of the instruction after each task.

Treatment 2 ("self-reflection"): The subjects were interrupted in intervals of four minutes. They were asked to describe what they were doing, to evaluate their strategy and to give themselves some instructions for their future behavior. In the rest of the time the subjects could verbalize aloud or silently as they liked.
Treatment 3 ("no verbalization"): In this group the subjects got the instruction to refuse verbalization, loud and tacit, as completely as possible. They should not ask themselves questions nor give themselves instructions.
Treatment 4 ("articulatory suppression"): Subjects of this group had to carry out a verbal routine task while solving the problem. They were instructed to speak two figures (32 and 43) in change again and again. We supposed that while being busy with speaking continuously, they wouldn't be able to verbalize problem-relevant topics as well as without this distraction.
Treatment 5 ("motoric interference"): Subjects were busy with a motoric routine task while solving the problem. In contrary to treatment 4, they did not carry out a task requiring the motoric system of speech, but knocked a rhythm on the table.
The sixth group ("controls") got no specific instruction.

Procedure

At the beginning of the experiment, subjects worked on three easy tasks, so that they could get used to the handling of the computer-aided scenario. Then they had to work on three experimental tasks of increasing difficulty. The first task requires the application of seven radiations in the right sequence, the second eight and the third twelve. The subjects had the possibility to restart each task and to go back to the initial state of the beetle, when they believed their sequence of operations to be misleading. Each task was aborted when the subject did not reach the target state within 30 minutes.
At the end of the experiment we tested the subjects' memory load. First the subjects had to memorize the specific effects of the radiations. Then they got a list of the pictoral codes of the twelve radiations (see figure 2) and had to assign the pictures to the labels "Alpha" to "My". In the second part of the memory test the subjects' capability to recognize the operators was evaluated.

Results

Effects of Verbalization on Problem-Solving Performance

A differentiated operationalization of problem solving performance was developed by analyzing the solution times. When a task was not solved the subject got no points for it. The maximum score of five points was achieved, when the solution time was among the shortest 20% of the sample (solution time $\leq$ 20th percentile). Four points were achieved when the solution time was between the 20th to 40th percentile and so on. In this way a maximum score of 15 points could be achieved in the three experimental tasks. The Kruskal-Wallis analysis shows significant differences between the six experimental treatments ($\chi^2 = 14.10$, df = 5, $p < 0.05$). The highest scores were achieved in the treatment "self-reflection", followed by "motoric interference" and controls (see figure 3). In figure 3 black lines in the boxes mark the median of the group. Within the boxes there are all cases from the 25th to the 75th percentile. The "whiskers" show the smallest resp. the highest value.

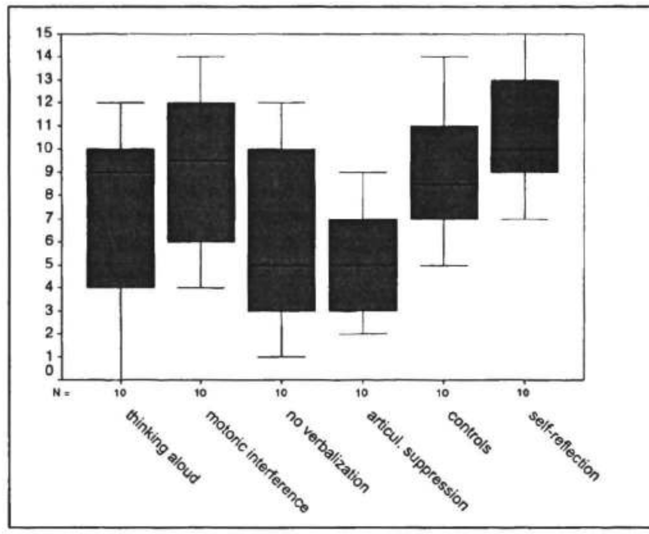

Figure 3: Box-plots of the number of points achieved in the experiment.

The "thinking aloud"-subjects showed a worse performance that the groups mentioned. This supports the hypothesis that self-explication and self-evaluation has stronger positive effects on problem solving than the unspecific instruction to think aloud. Moreover the better results of the controls indicate that the demand of thinking aloud could even disturb subjects, so that the results are even worse than without any treatment. Whereas the "self-reflection" treatment obviously improves problem solving performance, unspecific thinking aloud does not guarantee an improvement.
The two groups which were disturbed in their linguistic behavior show the worst performance. Articulatory suppression shows even stronger negative effects on problem solving that the mere instruction not to verbalize. The variances within the "no verbalization" group is significantly higher than within the "articulatory suppression" group ($F = 6.28$, $p < .05$). This result indicates that the subjects were more or less successful in suppressing their "inner dialogue" or that they followed the instruction more or less carefully. Obviously, the articulatory suppression treatment allows better control.
Besides the solution times, another indicator for problem solving performance is the number of operator applications. Some subjects find the solution directly, others go a long way round or restart the task several times, until they reach the target state. "Economy of radiation use" was one criterium explicitly mentioned in the instruction. For the following evaluation cases of subjects were excluded when they gave up before the maximum time for the task was exceeded. Looking at the number of radiations used, the six groups differ significantly ($F = 3.4311$, df = 5, $p < 0.05$). Comparing single treatments, there are significant differ-

ences between the group "articulatory suppression" and the groups "thinking aloud", "self-reflection" and controls. When solving the problem under "articulatory suppression", subjects needed more radiations to reach the goal. In contrast, subjects working under the treatments "self-reflection", "thinking aloud", and the controls made use of the radiations more carefully and economically.

Effects of Verbalization on Memory Load

Investigating the influence of verbalization on memory load, we looked at the active reproduction of elements of the problem space. A reproduction of the effects of the twelve radiations required the following steps: the subjects had to memorize how many of the six characteristics were affected by the operator. There were radiations which changed only one aspect of the beetle and others which required a specific value of all the six characteristics to work. Secondly the subjects had to memorize which characteristics were affected and which not. Thirdly they were asked to differentiate which characteristics were requirements and which of them were changed by the radiation. Finally they had to remember in which way the values changed (for example a change of color from red to green or vice versa). These aspects were summed up and for any correct answer subjects gained one point. A maximum of 160 points could be achieved.

In the recognition test, subjects had to assign pictures of the radiation effects (see figure 2) to their label (Alpha, Beta, Gamma, ..., My). For each correct answer they got twelve points, so that they could achieve a maximum of 144 points. The subjects were given the opportunity to name several alternatives. When they named two alternatives and one of them was right, they got six points, when they named three alternatives four points could be achieved and so on.

Figure 4 shows the scores achieved by the six experimental groups in the memory tests. The scores are counted in percentages of the maximum score. In general all subjects showed a bad performance. They achieved an average of 32% of the maximum score in the active reproduction and 36% of the maximum in the recognition task.

One explanation of the bad performance is the difficulty of the task. A maximum score in the active reproduction requires remembering the exact effects of twelve partly complex operators. But in solving the problem, it is not necessary to remember the effects of the operators in detail, because the pictures representing the effects can be activated on the screen at any time. Therefore it is not necessary for the subjects to memorize the operators in detail.

As to the number of points achieved in the active reproduction, analysis of variances shows significant differences between the groups ($F = 4.1076$, $df = 5/54$, $p < .05$). In the recognition task, the experimental groups do not differ. Single tests (Tukey-b) show significant differences between the controls, who achieved the highest scores, and the groups "no verbalization", "articulatory suppression" and "motoric interference".

The results show that experimental treatments handicapping the verbalization of subjects lead to a worse memory performance. However, the verbal motoric task ("articulatory suppression") does not show stronger effects than a non-verbal motoric interference. As a result, one cannot be sure whether the effects on memory load are results of an unspecific impairment of attention or of the specific verbal impairment.

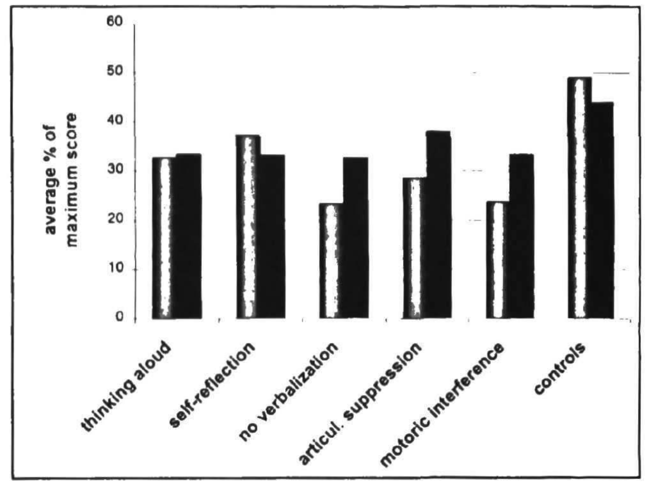

Figure 4: Average percentage of maximum score in the memory tasks. The light bars represent the active reproduction, the dark bars the recognition task.

One interesting result is that the group "motoric interference" shows comparably bad scores in the memory tests, though showing a good problem solving performance. In contrast, the group "articulatory suppression" shows the worst problem solving performance, whereas their memory loads do not differ significantly from other groups with the exception of controls.

In general, there is no correlation between the scores in the memory test and the points achieved in the problem (Kendall-Tau-b = .11, n = 60, $p < .05$). An interaction of problem solving performance and memory load can only be observed in the first experimental task (Kendall-Tau-b = .22, n = 60, $p < .05$). Subjects who can remember the operator's effects in more detail, take advantage at the beginning of the experiment, later on this advantage is not decisive any more. However, a negative correlation between the results in the memory test and the number of false clicks can be observed (Kendall-Tau-b = -.1990, n = 50, $p < .05$). Subjects who have worse memory performance also make more mistakes in handling the mouse. This supports the hypothesis that differences in memory loads are caused by an impairment of attention.

Case Studies

The results comparing the six experimental treatments did not fulfil the expectations in the whole. It could be shown that a disturbance of the "inner dialogue" has negative effects on problem solving performance and that "self-reflection" can improve performance. Mere thinking aloud however does not lead to the expected improvement. Furthermore the variances within the "thinking aloud"-group are high. The performance within the group is heterogenous: some subjects seemed to take advantage of verbalization, others seem to be handicapped by this instruction. The span between the performance of the "thinking aloud"-subjects varies between 0 and 12 of a maximum of 15 points.

There seem to be certain aspects of verbalization helping to solve problems more efficiently. The instruction to recapitulate one's own actions regularly, to criticize it and to give self-instructions for future strategies, leads to a clear improvement of performance.

In the following section two corner cases of thinking aloud subjects will be compared regarding to their verbalization. The comparing method refers to the method of "comparative casuistics" proposed by Jüttemann (1990). An exact analysis of the verbal data can on the one hand illustrate how subjects differ in following the instruction of thinking aloud. On the other hand, a comparative case study can show, which aspects of verbalization could improve problem solving. For the comparison the verbal data of the first task was transcribed and evaluated as to linguistic categories.

For the corner-case analysis two subjects were elected, who were in the same group, but showed different performance. One subject with a very good performance ("Logobene") will be compared with the worst subject of the "thinking aloud"-group ("Logomale"). Both subjects are female and students of psychology of the university of Bamberg.

Let us first have a look at the scores achieved by the two corner cases: The subject "Logomale" could not solve the task. After she worked on the problem for 30 minutes" the task was given up. In contrary "Logobene" solved the first task within 10 minutes and got the maximum score of 5 points. While "Logomale" applied 44 radiations, "Logobene" only needed 10 to reach the target state of the beetle.

Questions, answers and self-instructions: As with the number of utterances, the two subjects show clear differences. "Logobene" shows 126 utterances in ten minutes, whereas "Logomale" makes not more than 71 utterances while working 30 minutes on the solution. Their verbal data were analyzed with respect to the three "classical" linguistic categories of questions, answers/statements and requests (in this case: self-requests resp. self-instructions). The category sequences make the differences of the verbal data of the two corner cases obvious (see figures 5 and 6).

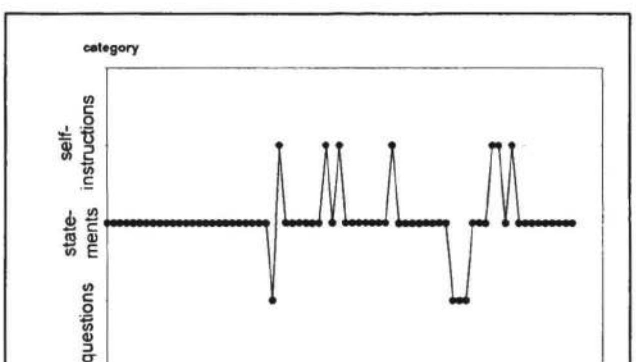

Figure 5: "Logomale": Sequence of questions, statements and self-instructions in the thinking aloud protocol.

"Logomale's" whole thinking aloud protocol shows only four questions. Moreover, there are only seven self-instructions. The main part of "Logomale's" verbalization are statements, e.g. "here the eyes can be changed", "here the color will not be changed" or "the beetle is green now and has small eyes". In the first ten minutes of the problem solving procedure, "Logomale" exclusively makes statements. Afterwards there are some questions and self-instructions, when "Logomale" deliberates whether she should start the task once more or not. until the end, however, statements dominate her thinking aloud.

The sequence of "Logobene's" verbal data shows a completely different pattern (see figure 6). She formulates hypothesis and asks herself questions from the beginning. The questions are often followed by answering statements and that leads to self-instructions. In the middle of the solution time, there is a vivid change of these categories in the sequence described. A dominance of statements can only be found at the end, when the solution is almost found and the problem turned to a solvable task.

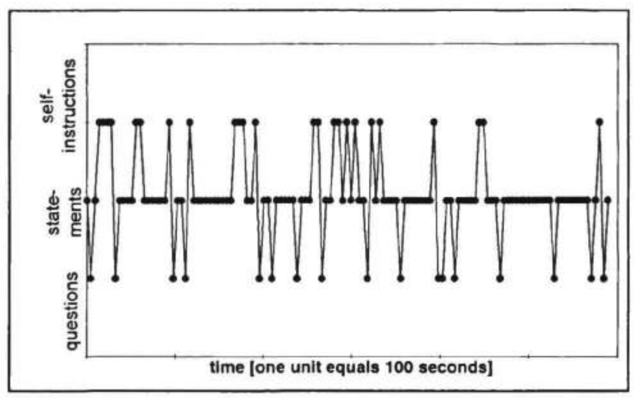

Figure 6: "Logobene's" vivid soliloquy: Sequence of questions, following statements and self-instructions.

In summary it can be shown that the thinking aloud of "Logomale" is no inner dialogue, but more a commentation of her activities. In contrary, "Logobene" talks vividly to herself. She asks herself questions and gives answers. Frequently a question-answer-sequence is followed by a self-instruction for future behavior.

Discussion

Summarizing the findings, the different treatments lead to striking effects on the performance of the "BioBeetle" tasks. The two experimental groups which were disturbed in their "inner dialogue" achieved the worst performance. Subjects who were encouraged to self-reflection showed the best performance. The other groups, "motoric interference", "thinking aloud" and controls show an average performance. Moreover it can be shown that there are strong differences in memory load. The controls show the best memory performance. The group "motoric interference", though belonging to the best problem solving groups achieve significantly lower scores in the reproduction task, as well as the groups "no verbalization" and "articulatory suppression". There is no correlation between memory load and problem solving performance, but between memory load and deficits in concentration.

Differences in problem solving performance can therefore not be explained by different memory loads, nor by different degrees of distraction. One conclusion of this study is, that a specific instruction to describe and evaluate the own behavior, which includes meta-cognitive processes leads to an improvement of problem solving. Thinking aloud in itself can not guarantee an improvement. Moreover there a some indications that an excessive commentation of action even has negative effects. The case studies could illustrate individual differences in the pattern of verbalization, both invoked by the same instruction to "think aloud". This gives some clues for explaining the high variances within the group of "loud thinkers". Some of them vividly talk to themselves, others are busy with describing the actual state and the actions carried out. The results of the case analysis led to a following study in which the patterns found ("commentation" vs. "questions, answers and self-instruction") are used as experimental treatments. Preliminary results of this experiment, in which a different scenario was used, show that this effect is robust over different settings (Bartl, in preparation).

The study has a second central finding: A disturbance of the "inner dialogue" leads to negative effects on problem solving performance. Both experimental groups that were disturbed in their linguistic behavior showed a comparably bad performance. "Articulatory suppression" has stronger effects than the instruction "not to verbalize". The high variances in the latter treatment indicate that the subjects followed this instruction more or less carefully. This experimental treatment produces the same problems as the "thinking aloud" instruction: the subjects differ in the way of following them, which leads to heterogeneous groups with high variances. In summary the results indicate a close relationship between language and thinking, whereby a soliloquy, including meta-cognitive processes, seems to improve problem solving capacities.

Acknowledgements

The experiments were conducted within the research project "Autonomie", Az. Do 200/12 supported by the Deutsche Forschungsgemeinschaft (DFG). The "BioBeetle" problem was designed by Dietrich Dörner. Special thanks to Patricia Cammarata for carrying out the experiments.

References

Baddeley, A. D. (1990). Human Memory. Theory and Practice. Hove: Lawrence Erlbaum Associates.

Bartl, C. (in preparation). Exploring an Island. The influence of commentations and meta-cognitive processes on complex problem solving. Pre-print available at Lehrstuhl Psychologie II, University of Bamberg, Markusplatz 3, D-96050 Bamberg.

Beradi-Coletta, B., Buyer, L.S., Dominowski, R.L. & Rellinger, E.R. (1995). Metacognition and Problem Solving: A Process-Oriented Approach. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 21 (1), 205-223.

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

Chi, M.T.H., Bassok, M., Lewis, M.W., Reimann, P. & Glaser, R. (1989). Self-Explanations: How Students Study and Use Examples in Learning to Solve Problems. *Cognitive Science*, 13, 145-182.

Chi, M.T.H., de Leeuw, N., Chiu, M.-H. & LaVancher, Ch. (1994). Eliciting Self-Explanations Improves Understanding. *Cognitive Science*, 18, 439-477.

Deffner, G. (1989). Interaktion zwischen Lautem Denken, Bearbeitungsstrategien und Aufgabenmerkmalen? Eine experimentelle Prüfung des Modells von Ericsson und Simon. (Towards the interaction of thinking aloud, solution strategies and task characteristics?). *Sprache & Kognition*, 8, 98-111.

Diaz, R.M. & Berk, L.E. (Eds.) (1992). Private Speech: From Social Interaction to Self-Regulation. London: Erlbaum.

Ericsson, K.A. & Simon, H.A. (1980). Verbal reports as data. *Psychological Review*, 87, 215-251.

Ericsson, K.A. & Simon, H.A. (1984). Protocol Analysis: Verbal Reports as Data. Cambridge, Ma: M.I.T. Press.

Franzen, U. & Merz, F. (1976). Der Einfluß des Verbalisierens auf die Leistung bei Intelligenzprüfungen: Neue Untersuchungen. (The influence of verbalization on the performance of intelligence tasks: new investigations). *Zeitschrift für Entwicklungspsychologie und pädagogische Psychologie* 8, 117-139.

Gagné, R.M. & Smith, E.C. (1963). A study of the effects of verbalization on problem solving. *Journal of Experimental Psychology*, 68, 12-18.

Hussy, W. (1987). Zur Steuerfunktion der Sprache beim Problemlösen. (The controlling function of speech in problem solving). *Sprache und Kognition*, 6 (1), 14-22.

Jüttemann, G. (1990). Komparative Kasuistik. (Comparative casuistics). Heidelberg: Asanger.

Lass, U., Klettke, W., Lüer, g. & Ruhlender, P. (1991). Does thinking aloud influence the structure of cognitive processes? In: R. Schmid & D. Zambarbieri (Eds.). Oculomotor Control and Cognitive Processes. Amsterdam: Elsevier.

Meichenbaum, D.H. & Goodman, J. (1971). Training impulsive children to talk to themselves: A means of developing self-control. *Journal of Abnormal Psychology*, 77, 115-126.

Phelan, J.G. (1965). A replication of a study on the effects of attempts to verbalize on the process of concept attainment. *Journal of Psychology*, 59, 282-293.

Rhenius, D. & Heydemann, M. (1984). Lautes Denken beim Bearbeiten von RAVEN-Aufgaben. (Thinking aloud while working on RAVEN matrices). *Zeitschrift für Experimentelle und Angewandte Psychologie*, 31, 308-327.

Tisdale, T. (1998). Selbstreflexion, Bewußtsein und Handlungsregulation. (Self-reflection, consciousness and action regulation). Weinheim: Psychologie Verlags Union.