PROBLEM SOLVING BEYOND THE LOGIC OF THINGS

Textual and Contextual Effects on Understanding and Solving Word Problems

Kurt Reusser

University of Berne, Switzerland

Address:

Abteilung Paedagogische Psychologie

Waldheimstrasse 6

CH-3012 BERN

Paper to be presented at the annual meetings of the American Educational Research Association, San Francisco, April 1986. The writing of this paper was supported by a fellowship from the Swiss National Science Foundation to the author, and by the Institute of Cognitive Science in Boulder, Colorado.

Abstract

A little recognized topic in the psychological and educational literature of problem solving is the linguistic and extra-linguistic or social-cognitive structure of problem presentation contexts. Word-, story- or text-problems, presented in classroom contexts, represent specific textual and situational patterns of a certain grammaticality. To verbally present a problem to a student in an educational setting always means to somehow organize a fact for the attention of a problem solver. There is the specific structure of the problem text itself by which situations, processes, actions, number relations are implicitly or explicitly expressed, questioned, commented, and there is also the stimulative nature of the pragmatic and social psychological context which shapes the student's textbook-problem solving behavior over a long time.

The present paper outlines and discusses the results of several studies showing that

- subject matter related to or factual attitudes toward a problem frequently don't play an important part in the problem solving efforts;
- students often solve problems - correctly - without understanding them;
- directionality and the goals of too many problem solving processes are so strongly anticipated by means of various textual and contextual cues that one can hardly speak of the solution any longer as of a genuine achievement of the problem solver;
- false contextual expectations can lead to abstruse errors of understanding and to weird solution attempts.

The experiments also indicate that students can become sensitive and skillful in perceiving and capitalizing on subtle textual and contextual signs pointing to the solution or anticipating its pattern. Moreover, most textbook problems, which are seldomly ill-defined, insolvable or carry irrelevant or other complicating information, and which are almost never presented without an informative question, let students get accustomed to certain courses of processing where a simple fact like whether an equation works out evenly or doesn't, can stop the process or push it further.

<u>Problem Solving 'Beyond the Logic of Things': Textual and Contextual Effects on Understanding and Solving Word Problems</u>

Children learn to deal with text-problems as soon as they enter the highly structured setting of formal schooling in the classroom. It is very likely that over many years the kinds and types of problem texts that are used in school, as well as the pragmatic-situative environment of their solving shape the students concept and style of problem solving. Nevertheless, the structure of commonly used problem texts as well as of problem presentation contexts are barely recognized as topics in the psychological and educational literature of text-, word- or story-problem solving.

Problem solving is widely seen by psychologists and teachers as a process of analyzing a task situation by following its <u>internal factual logic</u>, or as Wertheimer (1945, p. 33) put it, "the inner requirements of the situation". The tradition of "insightful" problem solving first arose with Gestalt psychologists such as Duncker (1935) and Wertheimer (1945), and it has come to life again in cognitive science with the stress on the role of knowledge and understanding in problem solving. While for many decades psychologists discounted the phenomena of insight, planning and understanding, cognitive psychologists nowadays have adopted a quite different attitude towards thinking and problem solving. This attitude comes quite close to the views that Selz (1922), Duncker and Wertheimer had, or had at least anticipated decades ago. Building adequate problem representations, goal-directed planning, inferencing and elaborating by using one's world knowledge, testing hypotheses, applying heuristics and comprehension monitoring are seen as basic operational building blocks of problem solving, as well as of the teaching of thinking skills (cf., e.g., Chipman, Segal & Glaser, 1985; Nickerson, Perkins & Smith, 1985). This paper will not argue against this view of problem solving at all. Instead, the perspective that we are going to illustrate, is intended to complement a view of problem solving that may have been over-idealized.

To present a problem to a student in a classroom or an examination setting means to work up, to pose a factual or fictional situation for the attention of a problem solver. First, there is the wording or specific structure of the problem text itself in which situations, processes, actions, number relations are implicitly or explicitly expressed, questioned, commented upon, (not) excluded and finally, anticipated. Problem texts are grammatical in the many subtle ways they signal paths and goals pointing to the solution pattern and putting the student on the right track. Second, there is the presentational setting, the tacit structure of the pragmatic or situational context that can provide quite significant hints to the solution of a specific task, and that also shapes the student's textbook-problem solving behavior. An important piece of this "context-knowledge" of even very young students concerning textbook math problems is that the problems must always make sense, that they are always solvable, that they work out neatly, that they usually don't contain irrelevant numerical information (i.e., everything that is numerical is relevant), that everything that is relevant is mentioned in the task, and that the explicit problem question - which always accompanies the task - is a reliable guideline in imposing a mathematical perspective on the task or in anticipating (SELZ, 1922) the "operation gestalt" of the solution.

The present paper outlines and discusses the results of several experiments and thinking aloud studies that show how linguistic and extra-linguistic or situational factors facilitate or impede the comprehensibility and solvability of problems. It will be shown how

- strategies following 'the logic of things', or factual attitudes towards a problem frequently don't play an important part in the problem solving efforts;
- students often solve problems - correctly - without understanding

them;

- directionality and the goals of too many problem solving processes are so strongly anticipated by means of various textual and contextual cues that one can hardly speak of the solution any longer as a genuine achievement of the problem solver;
- ill-defined or false contextual expectations can lead to abstruse errors of understanding and to weird solution attempts.

Irrelevance and Pre-knowledge in the Missile Task

The phenomena I am going to discuss first stems from an observation in an experiment which was originally aimed at understanding the effects of irrelevant numerical and episodic information in a text problem. The main question was how irrelevant information would affect problem difficulty depending on what kind of instructional aids were provided. We will not report the whole experiment here, but concentrate on an interaction between the formulation of an applied physics problem and two levels of pre-knowledge.

In the first version of the experiment, four versions of the missile task (Table 1), manipulating the extent and the quality of irrelevant

---

Insert Table 1 about here

---

information, were given to groups of 13-14 year old high school and to 18-19 year old college students. As we hypothesized from previous observations irrelevant scriptal information was easier to recognize as irrelevant and to eliminate than was irrelevant quantitative or numerical information (Figure 1).

Insert Figure 1 about here

A rather astonishing fact, however, is also revealed in Figure 1: Considering the frequency of correct solutions, the college students with some background in math and physics weren't any better than the eighth and ninth graders in version T4, the task with the most irrelevant numerical information. This observation led us to look closely at the student's written solution protocols. What we found there led us to conduct a second experiment, this time only comparing T1 and T4, now predicting that the college students (n=39) and the high school students (n=57) should differ in T1 but not in T4. This is exactly what happened. The result of this comparison is shown in Figure 2.

Insert Figure 2 about here

First of all, the almost opposite tendency with respect to T4 is due to an effect specific to one group of college students who participated in the experiment shortly before graduation: While 11 students solved T1 and 1 student didn't, only 1 student solved T4 and 10 didn't. It became clear what had happened when we looked at the solution protocols and the retrospective reports of some of the college students. What most of the college students did, activated by the giving of the distance and the launch-time delay and their more elaborated physics knowledge, was to construct a far more complex problem space, not only more complex than was required by the task, but also far more complex than their mathematical abilities could handle. The college students tried hard to do something impossible, namely to get a grasp of the mathematica description of the trajectory of the missile. They typically interpreted the proposition [APPROACH(A,B,DIRECTLY)] as implying a curved trajectory, while the high school students took it for granted that it was a straight line. This lead to different notions of distance. A minority of college students concluded the task was unsolvable or ill-defined. Comments and questions concerning launching angles and shape of the trajectory, as well as the drawings and the many complicated calculations found in the protocols, clearly indicate the complexity of the problem spaces that the college students - so to speak the expert subjects in the experiment - were forming. The college students knew (just as well as the high school students knew the opposite) that they should be able to come up with the calculation of a trajectory. However, the task is simply not specific enough to permit these calculations. Since, as we will show later, it doesn't seem to be at all easy to recognize a given task as being unsolvable, the majority of the college students tried hard to understand mathematically what happens in the task, with only a few finally seeing how the task could be solved under very simple assumptions.

The high school students were in a more comfortable situation than their more expert colleagues. Because of their modest physics and mathematical knowledge, they could not form a complex view of the physical task space or carry out demanding mathematical calculations. Their minimal knowledge didn't lead them to perceive the task as as complex as the consideration of the information, intended to be stripped off as irrelevant, would require. Those students who thought of of a curved trajectory, of the curvature of the earth or of the value of some launching angle, soon dropped these elements from his representation, because he knew that he had never calculated them before, and that he would never be able to do so. The high school student - as well as the college student on a more expert level - was not capable of seeing more structure in the given task as his assimilative knowledge base was able to interpret: Distance ... time ... velocity ... there may only be a single formula available, v = d : t , and some knowledge how to transform it algebraically ... -> d = vt ... -> t = s:v ... and nothing more is needed than to try a little, to look for an instantiation of the formula that works out and that fits the context of previous experiences and demands. Put another way: The lack of physics and mathematical knowledge led the students eliminate things which were of no use and for which no calculational methods were known. The smart high school student, who may even know how the problem situation could be understood in a more complex way - with a good feeling for the demands for text problems - refrained from maximizing his semantic in-depth processing. He had probably often experienced that <u>in most text problems</u> in the classroom context there is a sort of <u>"prestabilized harmony", a certain match between subject-matter complexity and the mathematical knowledge required for solving a problem</u>. From the didactic point of view of those who design mathematical text problems, there is often a hidden conflict between, on the one hand, the real world complexity of task situations and mathematical instruments required to cope with them, and, on the other hand, the modesty of the student's actual domain-specific and mathematical knowledge. There is no easy way out of this dilemma, and it is also hard to see what more general impact this difficulty has on student's problem solving behavior outside the classroom, where no designer of a situation problem, a priori, guarantees solvability or an easy match between one's own (mathematical) resources and the semantic requirements of a task.

<u>Text and Context Related Difficulty Misjudgment in the Bicycle Task</u>

The problem text is just one aspect of the input in problem solving, as well as being just one of the guiding forces. Whoever observes students in classroom and homework situations can find again and again <u>how few common textbook problems force students to do an in-depth semantic analysis</u>, how many students are striving to <u>orient</u> their problem solving in response to a multitude of indicators in the <u>context</u> of the factual problem structure - and how they succeed. The experiment that we are going to report isolates a factor "<u>estimated or assumed task difficulty</u>". This was done on the basis of observations indicating that students, often before or in place of a thorough analysis of the real content of a task, use context information to assess the direction and difficulty of a solution. The single task we employed in this experiment allowed us to manipulate the factor "estimated task difficulty" in two ways: as a text factor and as a context factor. Four different settings of the cyclist task were presented to ninth graders and to college students.

Task description:

Table 2a shows the simple (S) and the more complicated (C) text version of the cyclist task. Both problem texts were accompanied by the sketch in Figure

Insert Table 2a and Figure 3 about here

3 and the basic distance-time formula. The formula was provided so that the high school students would not fail merely because they could not reproduce the formula from memory. Distance-time tasks like the one we employed belong to the basic math curriculum of the junior high school. The task is seductive because it often puts people off the scent by letting them simply average the three partial speeds. This temptation, which we found by presenting the task to a few subjects, seems to be smaller or bigger according to how the velocities in the task are expressed. The two textual versions take these observations into account. While S, a suspiciously simple looking version, might warn the average problem solver to be cautious with regard to a (v1+v2+v3)/3 solution ("What's the difficulty here?"), the reassuringly complex version C induces the subject to expect rather the opposite: The problem solver finds at least one 'difficulty' in the calculation of the partial speeds v2 and v3, and since, moreover, the following calculation (v1+v2+v3)/3 works out evenly, he may think the problem is solved. Further, two co-textual additions were constructed with the function of explicitly inducing an expectation of either high (IDH) or low (IDL) task difficulty

(Table 2b). The difficulty induction

Insert Table 2b about here

---

was made by associating the task to school-specific types of examinations. Different additions had to be developed for both the high school and the college subjects. However, the relation between IDH and IDL is assumed to be the same for both groups.

Four task settings resulted from the combination of the text versions with the co-text additions:

S(IDL) : The induced low difficulty and the simple text version support each other. Quite a large number of (v1+v2+v3)/3 -solutions are to be expected.

C(IDL) : In contrast to S(IDL), v2 and v3 cannot be directly read off from the problem text but have to be worked out before the main calculation.

This 'element of difficulty' - in interaction with the IDL-addition - will probably make many subjects rather careless. There are even more errors to be expected than with the first setting - presumably the most of all settings.

S(IDH) : This version can be viewed as the complement to C(IDL). In the face of the high difficulty context, only a few subjects ought to accept a (v1+v2+v3)/3 solution. Therefore, in this condition, the <u>most correct solutions are expected</u>.

C(IDH) : Since for many high school students, the preliminary calculations of v2 and v3 are expected to be of significant difficulty, it can be assumed that, for many of them, this difficulty is taken into account sufficiently by the IDH context. Others - and presumably most of the college students - will look for an additional difficulty and do some deeper semantic processing. Altogether we still expect a fairly large number of erroneous solutions, but significantly fewer than under the conditions C(IDL) and S(IDH).

Hypotheses:

Three main predictions underlie the experiment.

H 1 : For both text versions, a higher frequency of correct solutions is expected for the condition of an induced high-difficulty context than for the low-difficulty context.

[f+] S(IDH) + C(IDH) > [f+] S(IDL) + C(IDL) => [IDH > IDL]

H 2 : In case of the induction of an equal difficulty, a higher frequency of correct solutions is expected under the S- than under the C-Condition.

[f+] S(IDH + IDL) > C(IDH + IDL) => [S > C]

H 3 : The highest frequency of correct solutions is expected for the condition S(IDH), the lowest for C(IDL).

[f+] S(IDH) > C(IDL) ;

[f+] S(IDH)max ; [f+] C(IDL)min

Results:

68 Bernese junior high school students (median age 15) and 51 college students (median age 19) participated in the experiment which was run in groups, by two experimenters during normal class hours. The written solution protocols - Figure 4 shows some examples - were first assigned to two restrictively defined categories:

* <u>Correct solutions</u>: Subjects were assigned to this category if they first calculated the times for each of the three distance-segments, summed up the times and then put the sum as the denominator into the basic formula (cf. example A in Figure 4).

$$v(\text{average}) = \frac{3s}{\underbrace{\frac{s}{v1}}_{t1} + \underbrace{\frac{s}{v2}}_{t2} + \underbrace{\frac{s}{v3}}_{t3}} = \frac{30}{\frac{10}{9} + \frac{10}{48} + \frac{10}{18}} = 16 \text{ km/hour}$$

* <u>Incorrect solutions</u> corresponding to the pattern "average the partial speeds" (aps-pattern; cf. example B in Figure 4).

$$v(\text{average}) = \frac{v1 + v2 + v3}{3} = \frac{9 + 48 + 18}{3} = 25 \text{ km/hour}$$

This analysis turned out to be too rigorous, especially for the high school subjects, where only 10% solved the problem completely correctly. Moreover, 30% of the protocols couldn't be classified because of all sorts of errors. Therefore, three raters analyzed the protocols again along the following lines: Whoever's solution attempt clearly showed the calculation of the partial times ti - even if this was done by the wrong formula (v/s instead of s/v ; cf. examples C in Figure 4) - was assigned to the correct

---

Insert Figure 4 about here

---

solution category. Due to the very specific structure of the task (one salient erroneous solution), virtually all subjects could be classified either as solvers, according to the above (weaker) criterion, or as non-solvers of the average-partial-speed kind. The frequency data for the correct (+) and the incorrect aps-solution (-) attempts of both subject groups are shown in the figures 5a, 5b, and 5c. We will now discuss each of the three hypotheses.

Insert Figures 5a, 5b, 5c about here

---

* Effects of difficulty induction [H 1 : f(IDH) > f(IDL)] : The reader will have noticed that the high school students produced rather few correct solutions (floor effect) whereas the college students showed only a few aps-solutions (ceiling effect), which weakens the statistical analysis. But this was the price we had to pay for using the same task in on two very different school levels. The frequency data confirm our first hypothesis that there are more correct solution attempts under the IDH- than under the IDL-condition (Figure 6). The difference becomes significant for the

---

Insert Figure 6 about here

---

college students if we add some subjects. There were no differences in solution times between IDH and IDL.

* Effects of problem formulation [H 2 : f(S) > f(C)]: The direction of the S/C-comparison for both groups is consistent with the previous prediction, but only the combination of the groups shows a statistically reliable difference (Figure 7). When we later tested 50 more college students comparing only S(IDL)

Insert Figure 7 about here

with C(IDL), we found a reliable difference in the expected direction ($p<.025$). So, even an expert-like group, for which simple algebraic transformations were by no means a source of difficulty, was sensitive to the induction of a low-difficulty context. What was highly affected by the problem formulation were the solution times of the high school students, but not of the college students - which makes sense: Even if the algebraic transformations in C were not an obstacle for solving the problem, it took the high school students a lot of time to do them, while the more expert students - due to their much better subroutine- skills - showed no difference. Further, both groups also needed more time to come up with the correct solution than with the incorrect one.

* Maximum- and minimum-frequency [H 3 : S(IDH)max ; C(IDL)min] : The presentation of the very simple task associated with a high-difficulty context showed the most correct solutions, the association of the enriched task with

Insert Figure 8 about here

an induced low-difficulty context the fewest. H 3 can be confirmed for both groups (Figure 8).

Discussion:

Three out of four high school students and one out of five college students were wrong in their solution of the cyclist task, succumbing, as it were, to its seductive textual and co-textual features. The results support our view that a contextual orientation ought to be seen as a constitutive factor of comprehension in word problem solving. Students of very different levels of expertise interact with a problem text not only in a fact- or subject-matter- related way, but also orient themselves - as in the present experiment - towards a co-textual context. The most striking difference in the data is that between the conditions S(IDH) and C(IDL). Is this difference due to a direct effect of the differences in the formulation of the task, or is it to be explained as an indirect effect due to an interaction of the textual characteristics with the difficulty-induction characteristics? We can clarify this issue by analyzing the C-version solution protocols with respect to the question how well the conversions of v2 and v3 were done under the conditions IDH and IDL. What Table 3 shows is that a simple and direct impact of the conversions of v2 and v3 on

Insert Table 3 about here

the solution process - under both IDH and IDL - can be excluded: Neither are there ID-related differences in correct conversions within both subject groups, nor is there a significant difference between high school and college students. The observed difference between the groups would be even smaller if one didn't take into account the mere calculation mistakes of the high school students in the course of working out v3. Because there is a difference between C(IDH) and C(IDL), in any case, we have to conclude that there is an indirect effect of a (per se) very simple conversion operation required by the C-version of the task. We can think of this effect as follows: The induced difficulty cause the students to expect complications. A high school student under IDL hardly expects any serious difficulty. Nevertheless, the conversion of v2 and v3 provides at least a minor difficulty, but there is no reason to seek further complications. There is also the (disarming) fact that the incorrect solution works out evenly, and therefore most high school students -

and also 31% of the college students - become careless and stop without any further epistemic control. - The situation is much different when a IDH context is given: The student expects a rather complicated situation. He is probably more observant, and EINGESTELLT (Luchins, 1942) on a task with a certain amount of complications. At least some of these expectations are fulfilled - probably at little bit more for the high school than for the college students: The conversions have to be done without any mistakes, which takes at least the high school students some time, but doesn't provide a significant difficulty for them. So, there remains an unfulfilled expectation; there still must be another difficulty: the student carefully recalls (vergegenwaertigt) the problem situation again, and - in some cases - succeeds in finding the crucial point. - The student may be even more unsatisfied under the condition S(IDH). The task elicits high expectations of complications which may never be fulfilled. The expectations are not met by any conversions or the required preparatory operations. The contradiction between the activated level of expected task difficulty and the initial representation of the problem or the first problem model is maximum, and therefore challenges the student to more deeply process the problem, with the result that under S(IDH) the most correct solutions are found. So, the additional calculations in the C-versions impede the solution process not by imposing any substantial difficulty on the task, but by inducing, as a moderating variable, in-depth processing.

Therefore, the problem solver proves himself to be not only one who is driven by "the desire ... to go on from an unclear, inadequate relation to a clear, transparent, direct confrontation - straight from the heart of the thinker to the heart of his object, of his problem" (Wertheimer, 1945, p. 236), but also one who, after having built an initial problem model by using all available textual and contextual clues, allocates the amount of resources in processing time and energy he expects to use. One could put it this way: A problem solver, after or while reading a specific task, allocates resources to be used in a solution attempt, or opens, so to speak, a sort of "cognitive processing energy account", against which all steps of processing get recorded. If there is nothing to "debit" the account - given a high "energy credit" - as under S(IDH), the problem solver will seek to work himself deeper into the material as if, as under C(IDL), his expectations were more or less fulfilled. In this view, to gain a deeper level of comprehension also means to activate or to strengthen the control functions. There were quite a few subjects under IDH who first worked out (v1+v2+v3)/3 , but later discarded this solution.

The Classroom Context of Problem Solving

How a problem is understood and solved, and how difficult it is, depends in the first place on its wording as a task. However, as the following examples will show, even the situational context within which a problem solving process takes place may have a significant influence on the understanding and solving of a problem.

How the Solving of a Problem Can Become a Joke

One can study the context- or situation-dependence of problem solving processes by looking at what impact the negation of the situation has. Suppose a college student gets the following task at an oral examination in physics:

> "Show how one can measure the height of a skyscraper with the help of a barometer."

Suppose further that the student answers correctly:

> "One can determine the height of the building by reading off from the the barometer the air pressure difference between road and roof. Air pressure decreases by 1 Torr (= 1 mm Hg) approximately every 30 feet."

It could easily happen that the candidate doesn't know the answer. This would probably result in a bad grade.

Sensitized by my interest in context phenomena in problem solving, I came across a short text, reporting how a candidate answered this question in a completely different way.2) He produced not only one but a whole series of answers - not to the pleasure of the examiner - according to the report.

(a) "You take the barometer with you to the top of the roof, tie it to a long rope and lower it to the road. Then you pull it back up and measures the length of the rope. This length corresponds to the height of the skyscraper."

(b) "... or you take the barometer outside on a sunny day, put it on the ground and measure its height and the length of the shadow. Then you determine the shadow of the skyscraper and calculate the height of the building with a simple proportional equation."

(c) "You take the barometer with you going up the stairs of the building. In the course of this you mark the wall in 'barometer-units'. The only thing you have to do afterwards in order to get the height of the building, is to count the 'barometer-units'. This is, of course, a very clear but rather crude method."

(d) "You take the barometer to the top of the building. Then you lean out over the edge of the roof. You drop the barometer and measure the falling time with a stop watch. Then you determine the height of the building by the law of falling bodies: d = 1/2 gt2 ."

(e) "If you were interested in a more subtle method, tie the barometer to a rope and let it swing as a pendulum. You determine the value of g (gravitational force in the formula T = 2 |1/g) on street and on roof level. Then you can work out the height of the building from the difference between g1 and g2."

(f) "Finally, if you don't want me to commit to a physics solution, then there still are many more possibilities. For example, you could take the barometer and knock on the janitor's door. If he answers the door, the you speak as follows: 'Dear janitor, I have here an exciting barometer. If you tell me the height of the building, then it's yours."

What remains to add is that, of course, the candidate also knew the 'correct' solution. - What happened? Obviously, a problem solving process suddenly became a sort of a funny joke. I hope that everyone will feel sympathy for the refreshing originality of the candidate. The little story sheds light on the examination context as a familiar problem solving setting, and it illuminates very nicely what can happen if its more general and task-specific constraints are disregarded by the candidate. One can look at what happened from the point of view of '<u>functional fixedness</u>', a phenomenon that was discussed first by Duncker (1935) and many others after him who studied the effects of "Einstellung" (Luchins, 1942; cf. Greeno & Simon, 1984). From the functional-fixedness point of view, quite a few of the candidate's solutions would have to be seen as very <u>difficult to retrieve</u>, because they abstract, with regard to the functional character of the object barometer, rather <u>remote and unusual features</u>. A barometer is an instrument for measuring air pressure. This is its <u>primary function value</u>. It would probably be very hard to come up with application-contexts where the barometer is used as a pendulum, a shadow-producing object, or a bribe for caretakers. Whoever is able to see a barometer under such a variety of only faintly moulding features, demonstrates creative behavior, though it presupposes a context in which the behavior is also perceived as original or creative. The typical examination context is not such a context, and so there was much argument about the value of the candidate's proposed solutions in our little story. And one must almost certainly assume that many examiners, were they exposed to similar situations, would feel insulted, provoked, or made a fool of. But why, really? I am going to try first a more general answer and then explore two guesses, which follow from it.

Oral examination situations are <u>behavior settings</u> (Barker, 1968) with a defined structure. The external course of events is mostly fixed, similar to a script (SCHANK, 1977). The actors in the situation play roles with well-defined expectations. They are, as it were, partners who have entered a - temporary limited - <u>speech act contract</u>. "For interaction to succeed, (both) participants must agree in their social situation definition" (Leodolter & Leodolter, 1976; cited from Forgas, 1985, p. 19). This "social contract" (cf. Mead, 1934) includes, for the candidate and the examiner, specific speech act obligations and rights. Greatly simplified:

- The examiner is entitled and obliged
  - to question the candidate on a previously defined topic and to present problems to be worked out by the candidate,
  - to judge the candidate's responses using criteria related to the subject-matter.

- The candidate is entitled and obliged
  - to prepare himself / herself to be examined about topics previously defined or agreed upon,
  - to obediently take the posed questions and to answer them after a short period of thinking.

It is not my point to elaborate this oral examination context, even if this were possible. It is only important here to see that such a context exists, and that we can assume that the candidate, the examiner and the reader of the story know it very well. It is this examination context, whose inherent obligations are systematically and intelligently <u>ignored</u> by the candidate above. Not that the candidate wouldn't live up to the expectations of the external examination script. It can on the contrary be assumed that the candidate

- was very polite,
- gave the examiner, while articulating his responses, a well-meaning and zealous impression,
- didn't miss any questions.

Basically, there is only a single behavior expectation that gets systematically and consciously negated by the candidate: <u>The expectation to understand and to answer the question in a situationally defined way</u>, which means that a solution is only acceptable if

- it can be regarded as, in a certain sense, intellectually demanding,
- it is founded on knowledge of the physics topics previously agreed upon,
- the barometer is regarded in its central function value as an instrument for measuring air pressure.

Because of the lack of explicitness of these expectations, the candidate does not neglect the letter but the spirit of the examination context. He irritates the examiner the same way that he amuses the reader of the story. With his solutions, none of which is based on the barometer's central function, but on remote functional and dispositional features (extension, weight, exchange value) of the object, the candidate reduces the examination to absurdity.

Stimulated by this little story, I got interested in two questions:

First, would other readers also find the story funny the same way I did? Could one possibly find a consistent rank order of solution-funniness as a consistent pattern of funniness judgments?

Second, could we predict the funniness of the solutions by the degree of their deviation from the pattern of expectations implied by the oral examination context and anticipated by the examiner's explicit question?

We predicted the following task-specific rank order of funniness (listed in reverse order). (1) means that this solution is considered to be the funniest:

(7) The correct solution: It corresponds entirely with both the textual and the contextual anticipations. Nobody will perceive this solution as funny. The function value of the solution is identical with the function value of the barometer.

(6/5) The pendulum (e) and the shadow (b) solution: Both are demanding from a math or physics standpoint, even if they do not employ the barometer in its essential function at all but in some of its merely accidental properties. Both solutions require thinking and some serious knowledge. They are neither dubious nor just convenient.

(4) The barometer as a unit of measurement (c): Admittedly, the solution is not very elegant, rather hard-working, but nevertheless simple and efficient. Besides that, the candidate negates or fails to meet the expectation of presenting a physically demanding solution.

(3) The rope-solution (a): The solution is not based on domain-specific knowledge or on skill or on staying power. It is convenient, yet it can be worked out in a lying position, and it undoubtedly leads to a correct and precise result.

(2) The janitor-solution (f): It simply defies any description of an intellectually honest solution in an examination context. The candidate not seeking the solution himself, but rather tries to buy it for the price of the "solution instrument": The epitome of a bone-idle, dishonest solution.

(1) The free-fall solution (d): This solution requires some knowledge of physics, which speaks against considering it the funniest. Unlike the the case of the 'janitor-solution', where one might think that an unintelligent or desperate problem solver could see no other way out but to sell the solution instrument for the solution, no argument of mere convenience or stupidity can be brought to bear in this solution. The candidate, who knows the law of falling bodies, who proposes to

work the solution not by letting fall, e.g., a stone, but by the irreversible destruction of the solution instrument, acts negligently in a way that transcends even the specific examination context - by the deliberate destruction of something of value.

We presented the seven solutions to 8 university students and asked them to put them in a funniness rank order. Table 4 shows two things: a high

---

Insert Table 4 about here

---

correspondence between empirical and predicted rank order, and a high consistency within the empirical rank order (coefficient of concurrency after Kendall, $W = .70$; $p<.001$).

Not infrequently, exam questions require, like many text problems, a certain sensitivity or cleverness in 'reading off' the intentions, anticipations and expectations from the text and the context of a problem. Whoever has this sensitivity for context, together with intelligence and knowledge, and who deliberately doesn't take it into account, can <u>subvert</u> a problem situation until it develops into the <u>purely comical</u>.

The Authority of Contexts and its Impact on Comprehensibility

The Swiss writer Peter Bichsel in his third Frankfurt Poetics lecture (1982, 49) remarked about his reading of Goethe's "Joseph":

> Maybe I would have broken off my reading if the author had been unknown to me. For example, I could have stopped reading because I could have assumed that the book was sort of sanctimonious, or, if you want, simply a book for people and not a book for literature. You may well interpret that as snobism. But my literary judgement is dependent of the context: I read that book in the context of Goethe and in the context of German literature.

Whoever is confronted as a student in a classroom, expects something comprehensible and solvable. And he knows that he is expected to produce an answer - even when there may be none. From the standpoint of divergent fantasy, one might call the sense-seeking behavior of students creative vis-a-vis nonsensical material or unsolvable problems, and sometimes there really are some such solutions. But this isn't the rule. Most of the time one encounters rather questionable or even ugly ways in which students try to understand and solve a problem a tout prix, because of the characteristic/moulding/shaping factor of context and its dubious impact. I would like to call that questionable impact the lack of intrinsic cognitive processing. Five examples:

Example 1

> I asked the class: 'Are you sure that this result is really correct?' Most of the pupils were plainly dumfounded by the question, surprised that it should be asked. Their attitude was clearly: 'How can you expect us to question the solution you have given us?' The question was strange to them, it touched the very essentials of what school, teaching, learning meant to them. No answer. The class was silent (Wertheimer, 1945,26).

Before Wertheimer asked the students this, he showed them how one can work out the area of a parallelogram with a very troublesome, unpleasant and senseless method - but leading to a correct result. Wertheimer comments further:

> Let the reader consider whether he has not often learned things in school that way. Isn't it the way in which perhaps you have learned differential and integral calculus? Even theorems of plane and solid geometry? Of course you had good reason to feel that the teacher was teaching sensible, serious things you had to learn. But did you have the possibility of another kind of learning, of really grasping? Could you do anything but put up with and submit to the teacher's demonstration, step by step, when you were unable to see why he did just this, then that? Could you help just following obediently as the steps dropped out of the blue? (26)

While following an explanation or a demonstration, students often don't become really challenged enough to understand what is presented to them, and to evaluate it by means of their own criteria of consistency and comprehension quality. Such criteria might not yet be available, but what does teaching do to help these develop?

The following examples were collected over the last few years. They may shed some more light on what's going on here.

Example 2: How old is the captain? 3)

97 First and second graders were given the following task:

"There are 26 sheep and 10 goats on a ship. How old is the captain?"

76 students "solved" the problem using the numbers in the task. - Or a similar task and a verbal protocol:

"There are 125 sheep and 5 dogs in a flock. How old is the shepherd?"

Protocol: ... 125 + 5 = 130 ...this is too big, and 125 - 5 = 120 is still too big ... while ... 125 : 5 = 25 ... that works ... I think the shepherd is 25 years old.

Example 3: Boats in the port

"Yesterday 33 boats sailed into the port and 54 boats left it. Yesterday at noon there were 40 boats in the port. How many boats were yesterday evening still in the port?"

We gave this task to 101 fourth- and fifth-graders with the result that

- 100 children produced a numerical solution; only one fifth grader rejected the task by writing that it was ill-defined and unsolvable,

- only 28 children doubted their result when they were asked afterwards to judge their certainty of having solved the problem correctly; 5 of the 28 children said that their solution was wrong;

- only 5 children out of the 101 called the formulation of the problem into question by saying the task was somehow difficult or strange, and this, after being asked to comment on it.

Example 4: The starting freight train

Looking at a very long and heavy freight train which is pulling out, you can sometimes make the following observation:

The engineer first backs up a little bit so that the couplings sag. Only after this he slowly pulls out. Why is this action useful?

Within the 48 (out of 55) 14 to 16 years old high school students, who didn't solve the problem correctly 4), there was one student who simply admitted : 'I don't know.' Among the remaining 47 erroneous solutions, there weren't only ugly solutions but also quite a number of adventurous attempts of explaining the facts. A short selection:

. "It could be a trick question because the engineer himself doesn't pull out. The train is what moves, the engineer only drives it."

. "It could be the sign of departure"

. "In order to prevent the engine, if it is cold, from heating up jerky."

. "In order to warm up the couplings."

. "The engineer does it from habit"

. "It looks better, if it is done so."

Example 5: The dispensoric theory of education

What distinguishes thinking people from others are their critical abilities. Cultures emerge and decline. This is a law of all biological life. You can actually find an overall structural dialectic between innovation and stagnation. The Greek philosophers, above all of them Euklyptos, have long since pointed out this fact. It is even true for the climate and the change of seasons. Human society resembles a garden, in which the most beautiful plants occur besides ugly weeds. In order to acquire a refrigerator, a worker in England has to work ten hours, in Argentine about ten times that much. On the other hand there is hardly a village in Africa, where you couldn't find a transistor radio. Education in Africa is different from education in the United States or in Europe. The validity of a mathematical formula is not restricted by the borders of continents. The subject of the natural sciences is nature. If natural science is everything, then everything is an object of natural science. Therefore, landscapes, forests and transistor radios build a unit together. What counts in boxing is to knock somebody out. The stronger wins against the weaker. Beauty as a category of nature doesn't play any role in boxing. The phenomena of this world have to be described and ordered before you can put them into a theory. Nothing else is the basis of dispensoric theory, which claims to capture the phenomena of the world in a certain totality. Trying to apply the theory to education, means to found a comprehensive theory of education

> which ultimately gets its final confirmation from practice, where practice simply has to be understood as individual and societal behavior. The dispensoric theory of education therefore isn't merely an epistemological principle but above all it provides an orientation for changing and improving the individual and societal conditions of life, which eventually will be capable of abolishing cultural and societal differences (From: W. Reyem, Dispensoric Theory and critical society, Oldenburg 1980, p. 33).

The text is syntactically correct, made up in a pseudo-scientific or scholarly way, and it even roughly follows a grammatical text pattern: general philosophical introduction, relatively concrete and diverse pieces of evidence, claim, scope and practical relevance of the theory, there is even a complete reference. But the text is, as intended by its author (Meyer, 1981), complete nonsense, put together from general and empty phrases, and <u>inconsistent on every (macro) level of deeper understanding</u>. Nevertheless, college students, teachers and university students were so much taken with this text that they spent hours trying to interpret it.

Meyer gave the text to his undergraduates in education shortly before graduation, saying that the text represented the newest educational theory - with the result:

> "In a two-hour class there was discussion about: The goals of dispensoric theory, its anthropological, philosophical and metatheoretic foundation, its method. None of the future graduate students uncovered the text as rubbish. The homework was done bravely..."

I presented the same text to 11 former teachers and at that time graduate students in education at the university (in two groups of 7 and 4 participants) asking them for a structured statement. The students didn't know that this was to be an experiment. They were in class with me, and they knew me well. The students got written instructions and some questions about the text: Here are the instructions:

(a) You have about 10 minutes to think about the enclosed text.
(b) Study the text carefully. Do you understand what its basic meaning is?
(c) What does the text mean to you?

(d) After you have gone through the above questions, please answer the following questions.

The multiple choice questions (all assuming, of course, that the text made sense) were concerned with the (in)compatibility of four principles in educational theory with the text, and with the judgement of the adequacy of several titles given to the text. Furthermore, the students were asked to write a concise one-sentence summary of the text. After finishing, every student got an additional sheet:

> You have worked for some time on a text which you probably hadn't read before. Were you able to express your impression concerning content and comprehensibility of the text on your answer sheet?
> If you want to add something, please do it here.
> . I have nothing to add : ...
> . I asked myself / I'd like to add the following thing: ....

Results:

- None of the subjects broke out of the context of trying to do a good job: Nobody walked out or protested by not working on the task or by writing nasty comments. All students obediently handed in their almost completely filled in sheets;
- 5 out of 11 used the additional sheet in order to express their doubts or their displeasure about the style and the content of the text ("additively composed", "shallow text", "very bad style");
- 8 students wrote a one-sentence summary;
- all students judged the difficulty of the text as being high.

Conversations with the students after the experiment showed clear signs of the social pressure created by the situation:

- As a university student I wasn't able to do anything other than to assume that the text was ok and the difficulties of comprehension were solely mine.
- I said to myself: Ok, I have to understand that; and then I read the text over and over until I thought I understood it.

- First, I just didn't know at all any more. But the experimenter, at the university, you know, had an effect on me like an authority. That made me perceive the text as sensible.

- Was I really so stupid? I was frustrated when the others - after exchanging some helpless glances - all started to write. I simply had to make the text mean something to me.

- This experience reminds me of the Milgram experiment.

- I'm shocked about my trust in authority.

The students who raised clear doubts about the content of the text, suppressed them or adopted the context until the additional sheet quasi-opened a valve to express them.

Discussion:

These examples clearly show that there are factors in the whole classroom setting which can heavily impair the quality of comprehension in problem solving. The studies and observations highlight the difficulties that students of all ages have in both rejecting an ambiguous or apparently senseless or unsolvable task and in simply admitting that one is unable to come up with a sensible solution. Classroom contexts seem to be authoritarian in the way that they maintain a leitmotif of sense expectation similar to what Hoermann (1976) called "sense constance" (Sinnkonstanz). Whoever as a student gets a textbook problem or any kind of text-related task assumes it to be basically sensible, unambiguous and solvable. And he feels strongly that he is required to produce a solution, to 'assimilate' the sense, even where there is none. There are at least two interpretations to what seems to be an 'always-answering-schema': First, it may reflect the pressure to answer sensibly, created by an authoritarian social context. Classroom problem solving, particularly the extreme case of solving a problem on the blackboard while talking aloud, has always had an aspect of self-presentation and competition, which may even include a moral component. Second, always responding to a question may simply reflect the cognitive failure to understand what a question really means. As Langeveld (1984) from a developmental psychological view point said, <u>early understanding of questions</u> by children is not yet strongly related to its <u>content</u>, but rather has a <u>rule-like communicative feature</u> of always eliciting an answer. While the first interpretation probably fits the examples (3), (4) and (5) quite well, the second interpretation may be adequate in explaining the number crunching and adventurous guessing behavior of (2), perhaps also partly of (3) and (4).

<u>Problem solving situations are role-defined social-cognitive and epistemic behavior settings, embedded in and legitimized by the broader institutional authority of schools</u>. Problem presentation contexts anticipate in many ways the structure of the legitimate solution space. To question the setting or parts of it as a fundamental restructuring of a problem situation model, seems to be extremely difficult. Not only because of the courage needed to leave the field (Lewin & Dembo, 1931), but also because it is normally quite hard to see why a situation is opaque, ambiguous or unsolvable. Also students get almost no experience in solving ill-defined or unsolvable textbook problems. <u>Almost every systematic dealing with ambiguity and unsolvability is factually excluded from textbooks, from curricula, and from the school setting</u> where it even seems alien.

<u>Sometimes it is Easier to Solve a Problem than to Understand It</u>

Problem texts contain a variety of signs pointing to or anticipating the solution (Reusser, 1985b). There are railings along which one can feel one's way on a solution path about which one may be not quite certain, but which is far from being completely dark. The way text problems are formulated and how they work out can provide subtle hints to the problem solver which may let him accept a solution even if he doesn't understand it. To come up with a correct solution and be quite sure about it, may not always mean, that one understands it, even if the solution was inferred by several steps. The _discrepancy between the acceptance of a solution and its understanding_ by the problem solver may even go so far that the problem solver is neither willing nor able to see the discrepancy at all. This may have something to do with some questionable preconceptions of teachers - and psychologists - about how students solve text problems. The last two examples will illustrate this facet of our story.

How the Phrasing of the Problem May Foil its Deeper Understanding

The designing and the results of a first study can be briefly summarized. 56 fifth to eighth graders were given the following problem in class with one of the problem questions.

> 30 students in a class were asked if they read or play an instrument in their spare time: 16 students read, 13 students play an instrument, and 5 students have neither of these hobbies.
>
> (a) How many students enjoy both hobbies in their spare time?
> (b) How many students who play an instrument, do not read?
> (c) How many students who read, do not play an instrument?

We found two things: First, there happened to be no false solutions in the group which got question a; all students correctly responded with "4". Second, in the groups with either question b or c, the rate of correct solutions dropped from 100% to 25% for (b) and 29% for (c), respectively, with the error "4" strongly dominating for the false solutions (66.5% in b, 62.5% in c). In addition, we collected a number of thinking aloud protocols. These protocols show very clearly that the students who first produced the correct answer "4" to version a, could rarely solve version b or c. There were several cases where the child, even if he/she solved version a before, again came up with the same answer "4" to b or c. However, several children who solved both versions of the problem and responded with the most frequent solution "4" showed considerable signs of doubt or hesitation in their protocols about the correctness of their answer and about the quality of their understanding or analysis of the problem.

Obviously, most students in the experiment who determine "4" as the correct answer in version a, don't make use of an adequate problem model, e.g., a correct set- or Venn-diagram where the answers to all versions of the problem can be read off. The fact that the answer "4" is the most frequent response independent of the explicit problem question indicates that the solution "4" doesn't necessarily indicate a understanding of the problem. What then is the source of this error? - Consider the phrasing or the grammatical form of the problem:

WS=30 : 16=R , 13=PI , 5=NONE . How many [...] ?

First, a (whole)set is introduced. It gets connected to its succeeding information by colon. Then, separated by semicolons, three quantity propositions occur, followed, finally, by a question. What could make more sense and what could be more reasonable than to assume that the three quantity propositions after the colon are the breakdown of the first mentioned quantity, which itself is interpreted as the wholeset that gets broken down? So, from this analysis and from the thinking aloud protocols, it becomes quite clear what the subjects do: They add the three subset quantities (16+13+5) and relate the sum to the wholeset. For a subject who does this mainly for "syntactic" reasons (stimulated by how the numbers are outlined and connected in the problem text), and not because he/she fully understands what the addition and the comparison (subtraction) operation mean, it is a very small step to take the result of the comparison operation (34-30=4) for the correct answer of the problem or - by default - for the most reasonable guess. When the problem structure is not really understood, this default strategy works, whatever the problem question may be. Moreover, the present task turns out to be especially suitable - because of the magical arrangement of the numbers - to be processed not only blindly but also in a consistently wrong way.

The problem solver in the classroom context is accustomed to calculate numbers even if he has only a vague, and probably insufficient, understanding of the problem. That this strategy is successful in a very broad range of classroom problem solving, should make us think not only about the invariance of problem presentation contexts, but also about the characteristics of problem texts we commonly employ in text books.

<u>To work out (un)evenly - a reliable hint for being on the right (wrong) track</u>.

Our last example deals with a rather unusual, but we think interesting, and hardly explored phenomena. I have observed more than once, how embarrassed students get if they feel themselves caught using thinking processes which are considered to be unelegant; how they seem to be subject to a certain censorship which they don't realize or, perhaps, want to admit. The <u>out-loud thinker in the classroom, the blackboard-problem-solver seems to be sometimes more factually oriented, more reflexive and deductive than the 'private' thinker</u>, possibly because of the didactic context which highly values deductive, insightful problem solving steps and ignores its darker side - the diverse processes of restructuring and generating new hypotheses. A typical and unfortunately quite reliable sign that one is on the solution path is the observation that the intermediate and/or final calculations for a problem work out evenly. Since this type of guiding forces or clues are not considered to be the sort of inferences students should rely on while solving math problems, it shouldn't surprise anyone that such clues are not reported by the students (in examination contexts, on the blackboard). Maybe students are <u>not even aware</u> of using this kind of internal feedback, or they may <u>suppress</u> it after the fact, and therefore sometimes report verbalizations to the teacher or experimenter that are rather idealized, cleaned-up versions of what they actually did (see Schoenfeld, 1983 for related observations).

The following protocol is due to a rather accidental constellation, where we were not looking at this kind of phenomena at all. We gave the following task to a former elementary school teacher:

1175 swiss francs ought to be shared amongst three siblings, in fact, inversely proportional to their age. A is 12 years old, B 18 years and C 21 years.

Here is the slightly shortened protocol:

" ... inversely proportional, i.e., the youngest gets most, the oldest least .. (10'') ... well, instead of 12 : 18 : 21 inversely, i.e. 21 : 18 : 12 ... one can cancel that down to 7 : 6 : 4 ... (30'') ... altogether there are 1175 francs to distribute ... as a unit one takes probably best 1/17th , because 7 + 6 + 4 = 17 ... ok I am going to divide the amount by 17 and then do the conversion (on a piece of paper)

```
1175 : 17 = 69,1
 155
   20
    3
```

—> [1] (after 2'): There is a flaw in my reasoning! The proportion has to be different, of course ... 1/4 : 1/6 : 1/7 ... it has to be reciprocal .. that gives me another unit ... 4 x 6 x 7 = 1/168th ... i.e., 1/84th works also ... ok, now I get things straight ... (works on paper)

$$\frac{1}{4} : \frac{1}{6} : \frac{1}{7} = \frac{21}{84} : \frac{14}{84} : \frac{12}{84}$$ ... that turns out to be for the unit

.s2

```
1175 : 84 = 13,9  => 14     ... now convert that ... 21 x 14 = 294
                                                      14 x 14 = 196
                                                      12 x 14 = 168
                                                                ---
                                                                658
```

—> [2]:(after 2 '): (calculates) 1175 : 47 = 25 .... —> Ah! I see, of course, you musn't divide by 84 ... this is not the correct unit ... you can ignore that and work only with the numerator ...21 + 14 + 12 = 47 . The inverse proportion is 21 : 14 : 12 , this corresponds to the task: The youngest should get the most and the oldest the least ... ok, now ... (calculates on paper)

```
A -> 21 x 25 = 525.-
B -> 14 x 25 = 350.-
C -> 12 x 25 = 300.-
               ------
               1175.-    ok, now it's correct."
```

There are three crucial places in the protocol, where the subject significantly changes her problem model and comments on her changes. At

—> [1] it is most likely, that the division doesn't work out evenly, which leads to a restructuring or at least makes the subject re-evaluate whether she had been wrong up to that point;

—> [2] the solving processes was driven one step further because not all the money could be distributed;

—> [3] here everything works out evenly, which induced the subject to accept the solution.

It is interesting that in a later conversation R. firmly believed that her thinking process was only guided by insightful inferential steps. From how she recalled her solution process, it was obvious that she didn't notice the places [1], [2], and [3] to be of any significance for the driving of her thinking or the final acceptance of the solution. On the contrary: My attempt to show to the subject how she most likely had also followed very pragmatic control decisions while solving the problem, first elicited quite strong defence mechanisms - probably based on a very strong ethos of insightful, pure subject-matter related thinking that the former teacher possessed. It required a careful reconstruction using the tape that finally let her agree with my interpretation of her solution.

Discussion and Educational Significance

The moral of our story is that classroom word problem solving is more - or also less - than the urgent analysis of a factual structure, in the sense that it is essentially and constitutively a species of social-cognitive activity. As a process of making sense of a problem text it is inherently tuned to its presentation context, to the classroom as a "behavior setting" (Barker, 1968). Word problem solving, as well as other types of language uses is inextricably tied to its surrounding social psychological environment and to the processes or strategies that regulate this context, or are derived from it (Clark, 1985; Forgas, 1985; Smith, 1983; Van Dijk, 1983; Van Dijk & Kintsch, 1983). A full understanding of our findings requires not only study of the failures or strengths in individual concepts, skills or procedures - as has been focused on in most research in problem solving - but also requires the understanding of the "social contract negotiated in the classroom between teachers and students" (Kilpatrick, 1985, p. 12). The problem solvers in our studies did far more than build up a problem representation or a problem model by deriving it from the "text base" (Kintsch, 1974) and their domain-specific knowledge. As language users do in general, the problem solver relied on <u>contextual as well as textual strategies</u> (Van Dijk & Kintsch, 1983) in order to capitalize on the problem-posing context as a diverse informational source. <u>Problem solvers build more or less adequate situation models</u> (Van Dijk & Kintsch, 1983) which are both textually and contextually based constructions including the representation of

- the <u>factual or 'real' task structure</u> which is often cued to a high degree by features of salience and focus, and by all sorts of directional hints in the didactically worked-up problem text, anticipating the course and goal-pattern of the solution (Reusser, 1985b);

- the <u>general characteristics of textbook problems</u>: well-defined with one solution which the teacher already knows; the solution is obtainable with one's own resources; calculations working out evenly indicate being on the right track; confinement to relevance and non-ambiguity: everything that is relevant to the solution is stated in the text, and everything that's stated is relevant; the explicit problem question is always present and highly informative; all problems are solvable;

- the <u>classroom as a "format"</u> (Bruner, 1985), as a <u>"social-cognitive and metacognitive matrix"</u> (Schoenfeld, 1983, p. 330), or as a behavior setting with its "social grammar" (Ervin-Tripp, 1972): It consists of norms and expectations regarding attitudes and strategies subjects ought to adopt (or avoid) while working on a problem, such as trying hard, being successful, always producing an answer, applying recently acquired knowledge and skills, focusing on the explicit problem question, etc.

What is the educational and scientific significance of these studies?

1. I think the studies highlight aspects and strategies of "understanding" in text-related problem solving which go beyond the mastery of concepts and discrete skills, and which, for the most part, have been neglected so far in the literature; - aspects too that probably also hardly get focussed and reflected on by students, teachers and textbook designers.

2. What we need above all are new types of textbook problems which more naturally enforce that kind of understanding that gestalt psychologists like Wertheimer and Duncker were concerned with and have described very beautifully. Text problems, as they are usually employed in classroom problem solving, maintain a set of 'grammatical features' or invariant properties which make them susceptible to the generation of task-specific biases and to all kinds of artifactual solution strategies. Students can become sensitive to and even very skillful in perceiving and capitalizing on very subtle but powerful cues pointing to the solution and anticipating its pattern. For example, by letting most elementary arithmetic and algebra word problems work out evenly, one imposes an aesthetic feature on a supposed real world context that doesn't really exist outside the textbook world. The unfortunate effect is that students start relying on this aesthetic feature by using it - successfully - as an on-line monitoring and off-line strategy for checking whether they are on the right track, or of solution adequacy, respectively.

Other important properties of didactically worked up problems include consistently used key words, the presence of informative questions, and the restriction that only relevant information is used (Reusser, 1985b). Moreover, many textbook math problems are not intellectually challenging because they are formulated as semantically poor, disguised equations instead of as thinking stories (Willoughby, Bereiter, Hilton & Rubinstein, 1981) or situation problems (Reusser, 1985a), which don't allow students to bypass a thorough semantic analysis in order to solve them. Math situation problems, for example, are seen as verbal descriptions of mathematical actions and episodes which contain an important goal, or which are structurally unsatisfactory: fragmentary, contradictory, or containing a gap. Situation problems provide comprehension starting points rather than being "locked up" and well-cued tasks. As such they can be used by teachers to initiate and foster processes of text comprehension and of mathematization (Reusser, 1986) by their students. - The major result observed in most of our studies is the extent to which textbook problem solving contexts can impair the quality of comprehension. Most problems do not ensure that the student has to "feel the difficulty in a situation" (Dewey, 1910) in order to generate a sensible question which could be seen as an intrinsic or semantic function of the problem situation (Reusser, 1986). Students are also not normally urged to control their solutions in a way that they relate the answer back to the raised question and to some metacognitive criteria of comprehension quality. This leads to the next conclusion.

3. I think that the deeper reason for the situational and contextual influences on understanding and solving of our problems lies in a fundamental weakness of the student's epistemic control behavior. Most of our subjects showed very weak schemata or epistemological standards of comprehension quality, of truth, and of coherence. These factors are not studied well

enough yet, but have an important educational impact. We should keep Dewey's postulate in mind that students not only should learn how to solve a problem, but also should learn how to control - working on one's own - the adequacy of a solution in some demanding, intersubjective way. Our subjects showed very strong tendencies in their understanding to rely on textual and contextual properties non-intrinsic or alien to the task structure, rather than to monitor the course of their ongoing comprehension and to evaluate the final state of comprehension by their own domain-related or topic-intrinsic epistemologies (Reusser, 1984). There is currently a growing body of research recognizing the importance of comprehension monitoring, epistemological standards and belief systems (Baker, 1985; Kitchener, 1983; Markman, 1977; Ryan, 1984; Schoenfeld, 1983; Wood, 1983). Many major questions are still open: How do epistemological standards of comprehension quality emerge in cognitive development? How can they be described properly, and how can they be taught or strengthened in order to establish them as dominant guiding forces in comprehension, as intrinsic components of the process of comprehension in "self-directed men" (Riesman, 1950).

4. There is even a more fundamental question connected to the previous reflections. It can be illustrated by a widely neglected characteristic in Wertheimer's (1945) monograph about "productive thinking". The issue is, ultimately, an ethical one. It is the question of the personality of the problem solver, of his/her overall style of problem solving and attitude toward objects and problem-situations which reflects the "sincerity of his attitude toward truth" (235).

> I want to remark that the feature of straightness, honesty, sincerity, does not seem peripheral in such a process. Generally speaking, it is an artificial and narrow view which conceives of thinking as only an intellectual operation, and separates it entirely from questions of human attitude, feeling and emotion ...This is especially clear in one example, in the transition from a blind egocentric view with its

> emotional ingredients to the latter steps. But even seemingly mere intellectual processes involve a human attitude - that kind of willingness to face issues, to deal with them frankly, honestly, and sincerely. Although I have referred to this fact only briefly in other chapters, it seems essential in many cases of productive thinking, including even our problem in pure elementary geometry(179).

Wertheimer made a very sharp distinction between processes which he called "structurally blind", "arbitrary", "ugly" and "foolish", and processes he vividly described as "honest", "sincere", "positive and reasonable". This distinction was based on ultimately the didactic philosophy of Wertheimer that clearly emphasizes the important role that the social-cognitive setting plays in problem solving, especially, how it shapes the metacognitive or epistemological mentality of the personality of the problem solver.

> Thus problems of personality and personality structure, structural features of the interaction between the individual and his field are basically involved. In connection with the latter we have also to realize the structure of the social situation, the social atmosphere one is in, the "philosophy of life" developed in the behavior of the child or person in his surroundings; the attitude toward objects and problem-situations eminently depends upon these factors. So also the social atmosphere in the schoolroom is sometimes of considerable importance for the development of genuine thinking. In the solution of this kind of problem it is more helpful at times to create the right mood than it is to force on the subject certain operations or drill(64).

In my own judgement, Wertheimer's view has some problematic aspects when he looks at problem solving from a largely ahistoric and idealistic viewpoint. In his dualistic picture of productivity in thinking there is basically little theoretical room for the unrestricted play of fantasy in finding a solution, and for the gradual improvement and development of epistemological - or ultimately ethical - standards. The way Wertheimer describes many phenomena may even reinforce the tendency among teachers and students to suppress socially unacceptable problem solving strategies in the classroom, to sweep them under the rug, so to speak. I can easily agree with a description of an ethos of honesty and sincereness in thinking which mirrors the Cartesian ideal of clarity, transparency and logical consistency - of rationality. It seems to me, however, that Wertheimer seems to normatively reward this kind of thinking, and denegrate other forms, and thus contributes to the tabooization of associative, tentatively scanning (Claparede, 1934: tatonnement), disjointed and trial-and-error-like thinking, which is by no means only driven by "the desire ... to go on ... straight from the heart of the thinker to the heart of his object, of his problem" (237), but which is nevertheless a constitutive and essential part of even expert problem solving (cf. Selz, 1922; Koestler, 1966).

In light of Wertheimer's analysis, and with respect to the current discussion about the development of epistemological standards, it seems to me useful to look at two issues separately: the issue of on-line monitoring of comprehension or progress toward solution, and the issue of comprehension per se, the acceptance of a solution after having checked it carefully. In other words, there are two contexts. There is the context of hypothesis generation or solution finding, i.e., how the fruitful hypothesis actually gets cued. And there is the context of its testing and evaluation, i.e., how the solution stands up to close examination. While in the first context even expert problem solvers - and certainly every problem solver in real life - will capitalize on every available and remote clue, novice problem solvers have to learn that there is this second context of careful testing of one's solution (hypotheses) against intersubjective structural standards. To rely on contextual and situational factors in the on-line guidance of problem solving and comprehension is not inherently bad. Ultimately, context is not 'beyond' the intrinsic logic of things, it is an essential and constitutive part of it. Where most of our subjects really fail is the evaluation of their solutions. After they have found them, they don't evaluate or test them seriously. The main problem to be addressed is two-fold: making teachers, students and designers of textbooks aware of the diversity of processes and strategies that can play a role in finding, reporting and justifying a 'rational' solution to even a simple text problem; and studying how students can be taught to test their solutions against <u>increasingly demanding epistemological standards of clarity, consistency, of proof and explanation</u>.

If and how a solution to a text problem is successfully found, depends on many factors we don't all manipulate consciously and insightfully. Classroom problem solving has a tendency not to take into consideration, or even to suppress non-intrinsic factors that maintain no inner relations to, or are not derived from, the content of the problem. But these factors exist in the different manifestations of trial and error behavior, of guidance by surface features of problem texts and of reliance on social-cognitive cues from the context. Whoever denies this, overemphasizes the well-orderedness, the pure fact-relatedness, even the rationality of thinking. This observation doesn't mean, however, that teachers should not continue to <u>uphold</u> the <u>standards of insight and comprehension</u> in problem solving.

Footnotes

1. This paper is a revised and abridged version of chapter 5 of my dissertation "Problemloesen in wissenstheoretischer Sicht: Problematisches Wissen, Problem- formulierung und Problemverstaendnis. Bern 1983. The data of our studies were collected between 1979 and 1983.

2. A. Calandra (1964) Die Barometer Geschichte. Schweizer Schule, 18, 1980.

3. The example stems from an unknown French source.

4. The engineer does this so that there is a maximum amount of slack between each car, so that in pulling out, the engine must overcome the friction of only one car at a time.

References

- Baker, L. (1985) How do we know when we don't understand? Standards for evaluating text comprehension. In: Forrest-Pressley, D.L., MacKinnon, G.E. & Waller, G.T. (eds.) Metacognition, cognition, and human performance. Vol. 1. New York: Academic Press. 155-205.

- Bichsel, P. (1982) Der Leser. Das Erzaehlen. Frankfurter Poetik-Vorlesungen. Darmstadt: Sammlung Luchterhand.

- Barker, R.G. (1968) Ecological psychology. Stanford: Stanford University Press.

- Bruner, J. (1985) The role of interaction formats in language acquisition. In: Forgas, J.P. (Ed.) Language and situations. New York: Springer.

- Chipman, S.F., Segal, J.W. & Glaser, R. (Eds.) (1985) Thinking and learning skills. 2 volumes. Hillsdale, N.J.: Lawrence Erlbaum.

- Claparede, E. (1934) La genese de l'hypothese. Archives de Psychologie, Vol. XXIV (Extrait). Geneve: Librairie Kundig.

- Clark, H.H. (1985) Language use and language users. In: Lindzey, G. & Aronson, E. (Eds. Handbook of Social Psychology. Reading, MA : Addison-Wesley.

- DEWEY, J. (1910) How we think.

- van Dijk, T.A. (1983) Episodic models and discourse processing. University of Amsterdam. Department of general literary studies.

- van Dijk, T. & Kintsch, W. (1983) Strategies in discourse comprehension. New York: Academic Press.

- Duncker, K. (1935) Zur Psychologie des produktiven Denkens. Berlin: Springer.

- Ervin-Tripp, S. (1972) On sociolinguistic rules: Alternation and co-occurrence. In: Gumperz, J.J. & Hymes, D. (Eds.) The ethnography of communication. New York: Rinehart, Holt & Winston.

- Forgas, J.P. (1985) Language and social situations: An introductory review. In: Forgas, J.P. (Ed.) Language and situations. New York: Springer.

- Greeno, J.G.& Simon, H.A. (1984) Problem solving and reasoning. Learning Research and Development Center. Technical Report.

- Hoermann, H. (1976) Meinen und Verstehen (To mean - to understand). Frankfurt.

- Kintsch, W. (1974) The representation of meaning in memory. Hillsdale, N.J.: Lawrence Erlbaum.

- Kitchener, K.S. (1983) Cognition, metacognition, and epistemic cognition. A three-level model of cognitive processing. Human Development, 26, 222-232.

- Koestler, A. (1966) Der goettliche Funke. Bern: Scherz.

- Langeveld, M.J. (1984) Grundsaetzliches - bezogen auf Piagets Kinderpsychologie. In: Lippitz, W. & Meyer-Drawe, K. (Hrsg.) Kind und Welt. Phaenomenologische Studien zur Paedagogik. Forum Academicum. Athenaeum-Hain-Hanstein. 149-158.

- Lewin, K. & Dembo, T. (1931) Untersuchungen zur Handlungs- und Affektpsychologie. X. Der Aerger als dynamisches Problem. Psychologische Forschung, 15, 1-144.

- Luchins, A.S. (1942) Mechanization in problem solving. Psychological Monographs, 54,6 (whole 248).

- Markman, E.M. (1977) Realizing that you don't understand. A preliminary investigation. Child Development, 48, 986-992.

- Mead, G.H. (1934) Mind, self and society. Chicago: University of Chicago Press.

- Meyer, W. (1981) Die dispensorische Erziehungstheorie. Die Zeit vom 30.November 1981.

- Nickerson, R.S., Perkins, D.N. & Smith, E.E. (1985) The teaching of thinking. Hillsdale, N.J.: Lawrence Erlbaum.

- Reusser, K. (1984) Problemloesen in wissenstheoretischer Sicht. Problematisches Wissen, Problemformulierung und Problemverstaendnis. Dissertation. Univ Berne

- Reusser, K. (1985a) From situation to equation. On formulation, understanding and solving 'situation problems'. Technical Report No. 143. Institute of Cognitive Science. University of Colorado at Boulder.

- Reusser, K. (1985b) What's in a text problem? A structural and functional analyses of a texttype? Manuscript submitted.

- Reusser, K. (1986) Situation-Problem-Solver (SPS): A cognitive simulation model of text comprehension, question generation and mathematization in solving situation or story problems. University of Colorado. Institute of Cognitive Science.

- Riesman, D. (1950) The lonely crowd. New Haven, Conn.: Yale University Press.

- Ryan, M.P. (1984) Monitoring text comprehension: Individual differences in epistemological standards. Journal of Educational Psychology, 76,2, 248-258.

- Schank, R.C. & Abelson, R.P. (1977) Scripts, plans, goals and understanding. Hillsdale, N.J.: Lawrence Erlbaum.

- Schoenfeld, A.H. (1983) Beyond the purely cognitive: Belief systems, social cognitions, and metacognitions as driving forces in intellectual performance. Cognitive Science, 7, 329-363.

- Selz, O. (1922) Zur Psychologie des produktiven Denkens und des Irrtums. Bonn.

- Smith, P.M. (1983) Social psychology and language: A taxonomy and overview. Journal of Language and Social Psychology, 2, 163-183.

- Wertheimer, M. (1945) Productive thinking. Enlarged edition, Chicago: University of Chicago Press. 1982.

- Willoughby, S.S., Bereiter, C., Hilton, P. & Rubinstein, J.H. (1981) Real Math: Thinking story book. Level 2. La Salle,Ill.: Open Court.

- Wood, P.K. (1983) Inquiring systems and problem structure: Implications for cognitive development. Human Development, 26, 249-265.

T1: Two missiles, which were launched several thousand kilometers away, are approaching each other. Missile A is flying with an average speed of 9000 km per hour and missile B with an average speed of 21000 km per hour. Determine their mutual distance exactly 5 minutes before the collision.

T2: Two missiles, which were launched 12000 kilometers away, are approaching each other. Missile A is flying with an average speed of 9000 km per hour and missile B with an average speed of 21000 km per hour. Determine their mutual distance exactly 5 minutes before the collision.

T3: Two missiles which were launched 12000 kilometres away, are approaching each other. Missile A is considerably slower than missile B. Missile A is flying with an average speed of 9000 km per hour and missile B with an average speed of 21000 km per hour. Both missiles will collide in 5 minutes. The power of the crash will be so extreme that within some fraction of a second both missiles will be completely destroyed. The missile parts will burn up in the atmosphere. Determine the mutual distance between the missiles exactly 5 minutes before the destructive collision.

T4: Two missile are launched 12000 km from each other. Missile A is started 20 seconds before missile B. Both missiles are approaching each other directely, missile A with an average speed of 9000 km per hour, missile B with an average 21000 km per hour. Exactly 5 minutes before the collision a third missile C is started. Its average speed is 15000 km per hour and it is also flying towards the point of collision of missile A and B. Determine the distance between missile A and B at the time of the launching of missile C.

Table 1: The four versions of the missile task.

	story information	numerical information
T1:	minimal	only relevant
T2:	minimal (identical to T1)	n + 12000 km
T3:	elaborated	n + 12000 km (identical to T2)
T4:	further extensions	n + 12000 km + 20 seconds + 15000 km/hour

The task is solvable by a simple distance-time calculation:

$$\text{DISTANCE(d)} = \frac{V_{\text{MISS(A)}} + V_{\text{MISS(B)}}}{60} \cdot 5 = 2500 \text{ km}$$

Table 2a: The cyclist task

S: A cyclist is going over a hill. His uphill speed is 9 km/h (v1), downhill it is 48 km/h (v2), and on the succeeding flat part he is riding 18 km/h (v3). All three parts are of equal length, namely 10 km each (d). What is the average speed of the cyclist for the whole trip?

[Reminder: speed (v) = distance (d) : time (t)]

C: A cyclist is going over a hill. His uphill speed is 9 km/h (v1). His downhill speed (v2) is by 6 km/h less than six times the speed of the uphill ride. The speed for the succeeding flat part (v3) is by 10.5 km/h less than the arithmetical mean of the first two speeds (v1, v2). All three parts are of equal length, namely 10 km each (d). What is the average speed of the cyclist for the whole trip?

[Reminder: speed (v) = distance (d) : time (t)]

Table 2b: The additional co-text

IDL: (induced difficulty low)

(both age groups)

The following mathematical task was presented in 1979 at a high school entrance examination
in the Canton of Zurich. The task should hardly be difficult for you since at that time almost 70 % of the candidates solved it correctly.

IDH: (induced difficulty high)

(high school students)

The following mathematical task is not that simple. It was presented last year
at the college entrance examination in the Canton of Solothurn, where only one third of the candidates solved it correctly.

(college students)

The following mathematical task is not that simple. It was presented last year in the canton of Solothurn at the final math examination of College, where only one third of the candidates solved it correctly.

		$V_2 + V_3$ correctly converted	solution frequency
IDH	HS	79%	32%
	Coll	92%	83%
IDL	HS	88%	6%
	Coll	100%	69%

Table 3: Correct conversions of V_2 and V_3, and correct solutions, under condition C(IDH) versus C(IDL) for the high school (HS) and the college students (Coll).

	Subjects										
	1	2	3	4	5	6	7	8	Σ(ranks)	rank ord.(emp)	rank ord.(theo)
a: Rope	3	2	3	3	2	2	3,5	2	20,5	3	3
b: Shadow	5	5	6	5	4	3,5	5,5	5,5	39,5	5	5./6.
c: B. as meas. unit	6	4	4	4	5	5	3,5	3,5	35	4	4
d: Free fall	2	3	2	1	1	1	2	1	13	1	1
e: Pendulum	4	6	6	6	6	3,5	5,5	5,5	42,5	6	5./6.
f: Janitor	1	1	1	2	3	6	1	3,5	18,5	2	2
g: Barometer	7	7	7	7	7	7	7	7	49	7	7

Table 4: Funniness ratings of barometer solutions

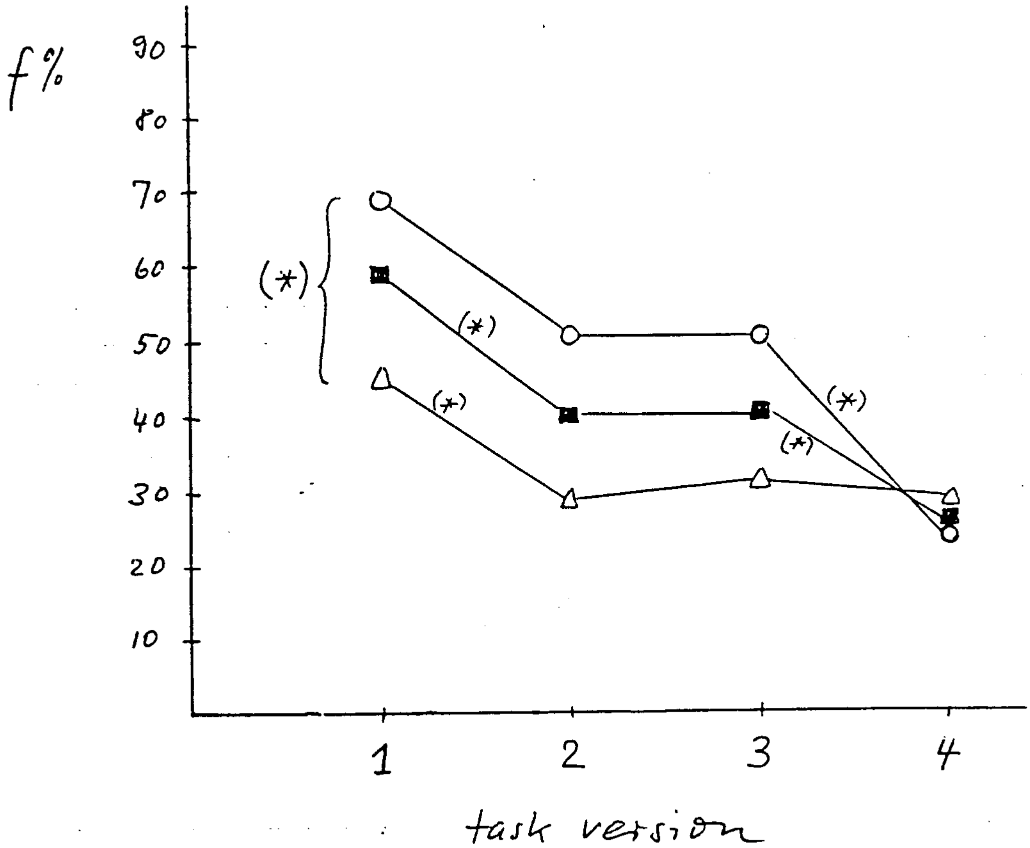

Figure 1: mean percent of correct solutions for the different missile task versions and the different student groups.

○ College students
△ High school students
■ College and High school students
(*) statistically significant differences

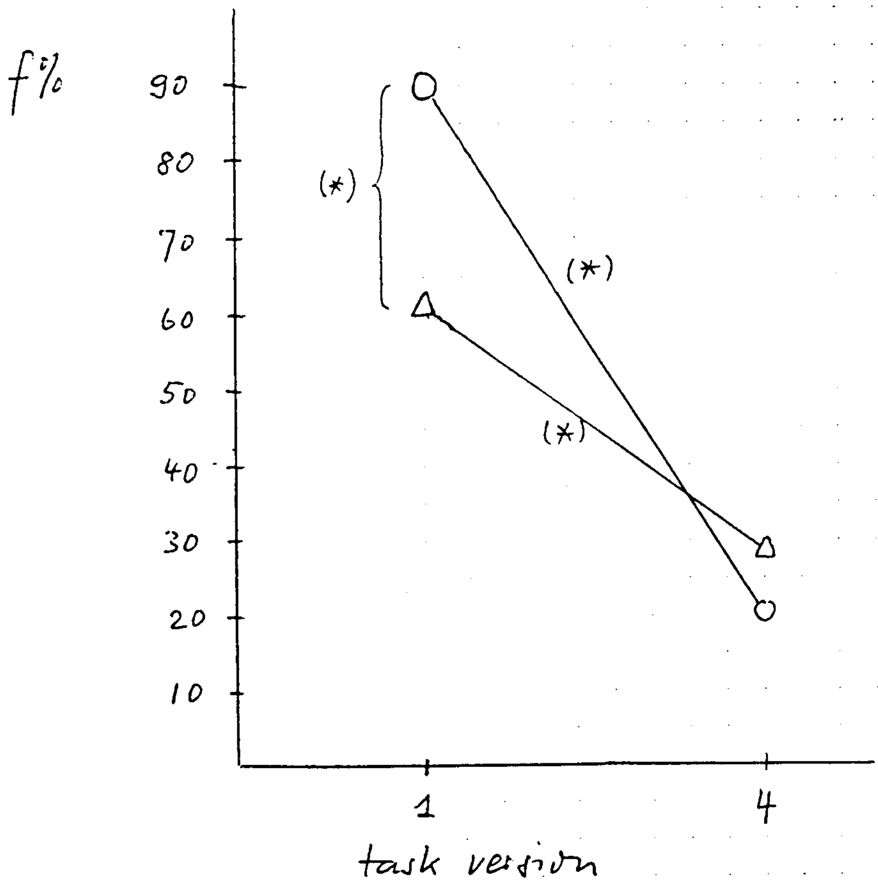

Figure 2 : mean percent correct for college students and high school students on version 1 and 4 of the missile task

O College students
Δ High school students

(*) statistically significant differences

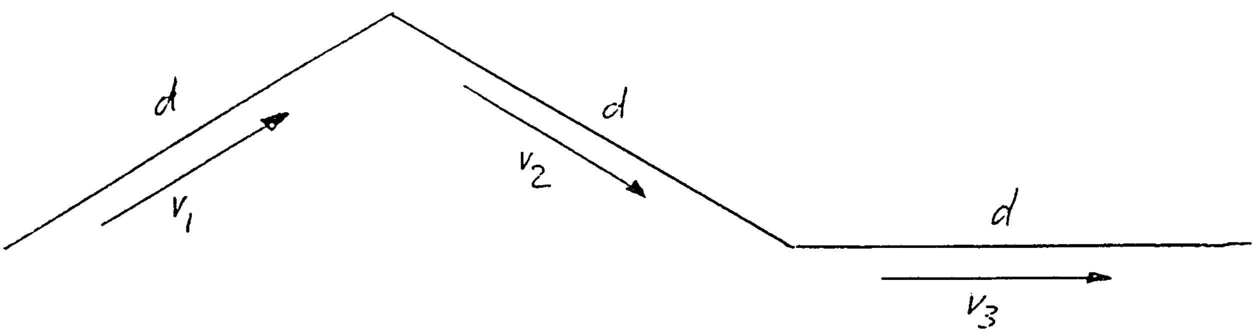

Figure 3: Sketch accompanying the cyclist task

A:

S (IDH)

V_1 9 km/h auf 10 km $\rightarrow$ $1\frac{1}{9} = \frac{80}{72}$ h

V_2 48 km/h auf 10 km $\rightarrow$ $\frac{5}{24} = \frac{15}{72}$ h

V_3 18 km/h auf 10 km $\rightarrow$ $\frac{5}{9} = \frac{40}{72}$ h

30 km $\qquad \frac{135}{72}$ h

$v = \frac{s}{t}$ $|\cdot t$

$v \cdot t = s$ $|:v$

$t = \frac{s}{v}$

$\frac{10}{9} = 1\frac{1}{9}$ h

$\frac{\cancel{105}}{\cancel{48}\,21}$ $\qquad \frac{\cancel{10}5}{\cancel{18}\,9}$

$\frac{\overset{2}{\cancel{30}} \cdot \overset{8}{\cancel{72}}}{\cancel{135}\ \ \overset{}{\cancel{45}}\ \ \cancel{15}\ \ 5} = \underline{\underline{16\ \text{km/h}}}$

B:

C (IDL)

$V_1 = 9$ km/h

$V_2 = 6 \cdot 9$ km/h $- 6$ km/h $= 48$ km/h

$V_3 = \left(\frac{48\text{ km/h} + 9\text{ km/h}}{2}\right) - 10{,}5$ km/h $= 18$ km/h

$V = \frac{9\text{ km/h} + 48\text{ km/h} + 18\text{ km/h}}{3} = \frac{75\text{ km/h}}{3} = \underline{25\text{ km/h}}$

$\underline{\underline{V = 25\text{ km/h}}}$

$V_1 = V_1$

$V_2 = 6 \cdot V_1 - 6$ km/h

$V_3 = \left(\frac{V_1 + V_2}{2}\right) - 10{,}5$ km

C:

S (IDH)

10 km : 9 km/h = 1 Std. 6 Min

10 km : 48 km/h = 0,1 = 6 Min

100

10 km : 18 km/h = 0,5 = 30 Min

100

10

1 Std. 6 Min.

+ 6 Min.

+ 30 Min.

1 Std. 42 Min. = 102'

30 km : 102 Min = 0,20 = $\underline{\underline{20\text{ km/h}}}$

300

96

D:

C (IDL)

V_1 : 9 km/h — 10 km = $\frac{8}{10}$

V_2 : 48 km/h → 10 km = $\frac{48}{10}$

V_3 : 18 km/h → 10 km = $\frac{18}{10}$

$\frac{75}{10}$

$v = \frac{s}{t}$

$\underline{t = \frac{v}{s}}$

30 km : $\frac{75}{10}$ — 30 : 75 = $\underline{\underline{4\text{ km/h}}}$

$\frac{1}{10}$ — 6 min.

$\frac{75}{10}$ — 450 Min — 7,5 h $\qquad \frac{75 \cdot 6}{450}$

Figure 4 : solution protocols for the cyclist task

difficulty induction +

oblem
rmulation +

		IDH f	IDH (f%)	IDL f	IDL (f%)	IDH+IDL f	IDH+IDL f%
S	+	8	(45%)	3	(21%)	11	(34%)
	–	10	(55%)	11	(79%)	21	(66%)
C	+	6	(32%)	1	(6%)	7	(19%)
	–	13	(68%)	16	(94%)	29	(81%)
S+C	+	14	(38%)	4	(13%)		
	–	23	(62%)	27	(87%)		

Figure 5a : Solution frequencies of High school students in the bicycle task.

difficulty induction +

roblem
ormulation +

		IDH f	IDH (f%)	IDL f	IDL (f%)	IDH+IDL f	IDH+IDL (f%)
S	+	13	(100%)	10	(77%)	23	(88%)
	–	0	–	3	(23%)	3	(12%)
C	+	10	(83%)	9	(69%)	19	(76%)
	–	2	(17%)	4	(31%)	6	(24%)
S+C	+	23	(92%)	19	(73%)		
	–	2	(8%)	7	(27%)		

Figure 5b : Solution frequencies of College students in the bicycle task.

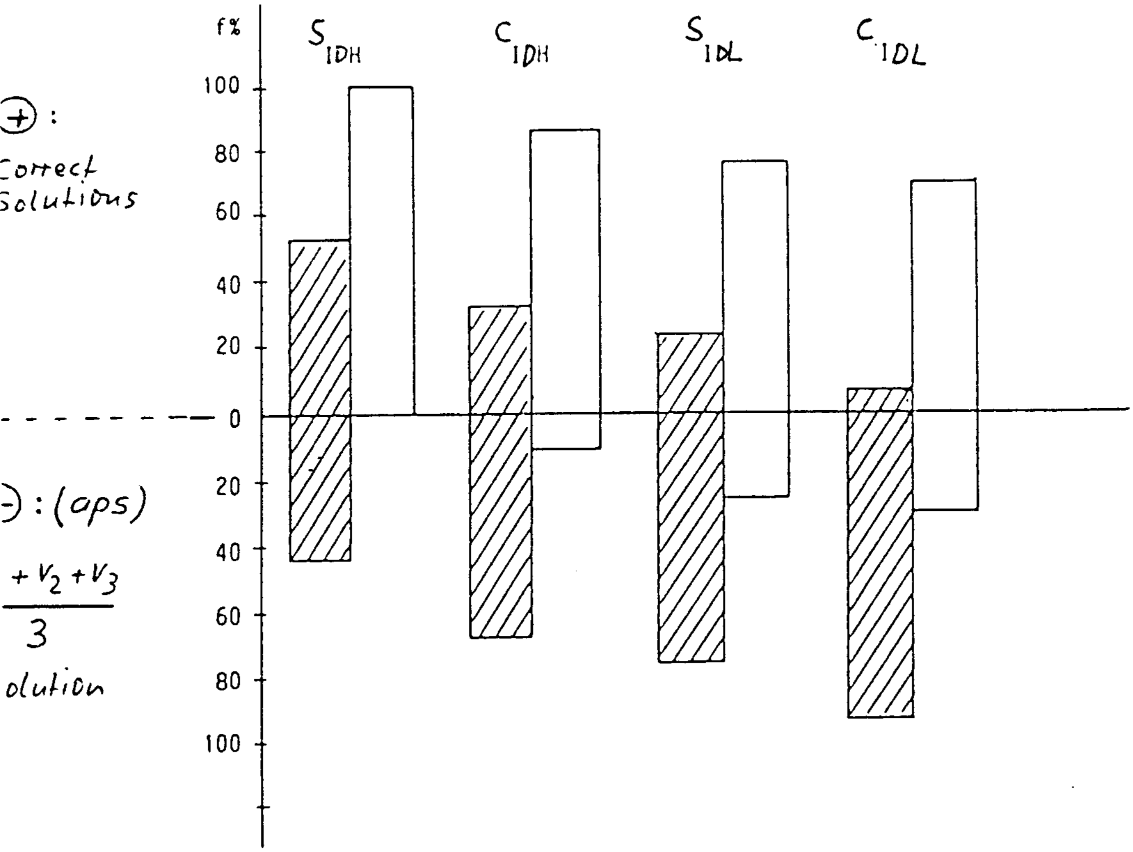

Figure 5c: Solution frequencies in the bicycle task under two types of difficulty induction (IDH, IDL) × two task formulations (S, C).

☐ ~~correct solutions~~ College students

▨ ~~aps — error~~s Jr. High school students

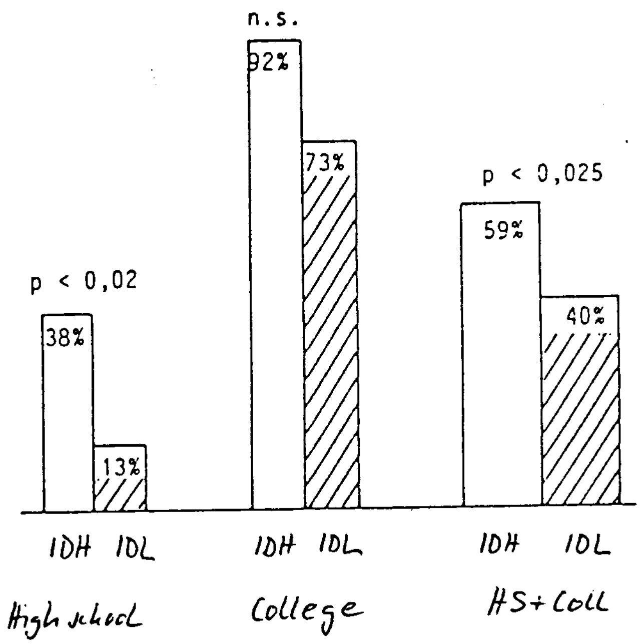

Figure 6: Effects of difficulty induction on solution frequencies in the cyclist task

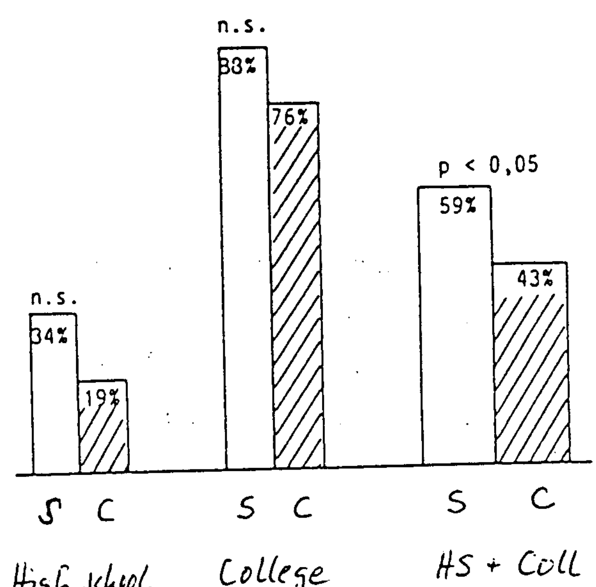

Figure 7: Effects of problem formulation on solution frequencies in the cyclist task.

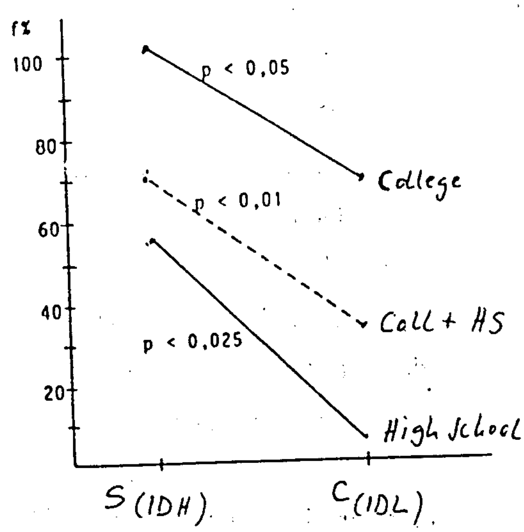

Figure 8: Mean percent correct answers as a function of S (IDH) and C (IDL).