to comply with such requirements, the normal course of their internal processes is altered (e.g., rote serial rehearsal may increase and attention may be shifted).

Apart from the effects of imposing specific requirements for verbalization, hearing overt verbalizations generally facilitates memory retrieval and storage. The first component of this effect is most likely the greater durability of the accoustical sensory store, which only spans a couple of seconds and is furthermore subject to masking by subsequent verbalizations. The second, longer-term, component is a facilitation in retrieval, due to the specific cues supplied accoustically by the overt verbalization. The strength of the second effect would depend on how selective the overt verbalizations were with respect to the retrieval task. Increases in the verbalization of material other than that which was to be remembered should decrease this effect.

THINKING-ALOUD PROCESSES

In the previous section we reviewed experiments in which the subject was instructed to verbalize or rehearse the actual stimuli. Only in the arithmetic tasks was the subject required to verbalize thoughts that he himself generated in the course of performing a task. We turn now to our main concern, thinking-aloud (TA) instructions, where the subject is specifically ask to vocalize those self-generated symbols while he performs his task. Again, we will be seeking evidence on the extent to which the TA activity affects the performance of the main task.

Thinking-aloud activity is not entirely alien to everyday life, and almost all subjects have probably had some experience of it before they come to the laboratory. Students at school occasionally have to explain their solutions of problems aloud to their fellow students in order to show how the solutions were generated. In other situations, people explain or describe their solution attempts to others, so that the listener can tell them where their thinking is in error. Even more frequently, people just communicate their thinking to others. In many cases, at least, such verbalization requires considerable intermediate processing prior to articulation, and is distinctly different from the verbalization of ongoing cognitive processes. In order to characterize the differences, we will describe three different levels at which a subject can verbalize his thought processes and their content.

Levels of Verbalization

1. A first level of verbalization is simply the vocalization of covert articulatory or oral encodings, as required in the tasks we have presented in the previous sections. At this level, there are no intermediate processes, and the subject needs expend no special effort to communicate his thoughts. A distinction must still be made, however, between cases where the subject is directing his communications to himself, and those where he wishes to communicate with other people. The self-directed verbalizations have been found to be more idiomatic and to use more idiosyncratic referents than communications directed to others (Werner & Kaplan, 1963). In which category a particular example of verbalization falls may depend on the subject's interpretation of the instructions, as well as upon their actual content.

2. A second level of verbalization involves description, or rather explication of the thought content. We assign to this level verbalizations that do not bring new information into the focus of the subject's attention, but only explicate or label information that is held in a compressed internal format or in an encoding that is not isomorphic with language (e.g., information about odors). We will have to add to our model the requisite recoding processes. We will refer to verbalization requiring such recoding as Level 2 verbalization.

Since explication or recoding requires processing time for the subject but does not replace other processing involved in the task performance, a subject who is verbalizing at this second level can be expected to take more time for the task than one who is not verbalizing. However, we would hypothesize that such recoding does not change the structure of the process for performing the main task.

3. A third level of verbalization requires the subject to explain his thought processes or thoughts. An explanation of thoughts, ideas, or hypotheses or their motives is not simply a recoding of information already present in STM, but requires linking this information to earlier thoughts and information attended to previously. Level 2 verbalization does not encompass such additional interpretative processes.

If, in the normal course of task performance, a subject pays attention to the relational or generating processes that are bringing new information into his STM (and insofar as he *can* pay attention to them), the verbalization would fall within the compass of Level 2 verbalization. But if the subject does not attend to these processes, then he will have to make intermediate inferences, which may or may not represent correctly his actual motives or the actual causes linking his thoughts. Moreover, an instruction requiring a subject to explain his thoughts may direct his attention to his procedures, thus changing the structure of the thought process.

Alternative Thinking-Aloud Instructions

The subject's TA protocol and (specifically whether it will correspond mainly to Levels 1, 2, or 3) may well be influenced by the exact wording of the TA instructions. We will begin by examining the instructions given by the two psychologists, Duncker and Claparède, who are usually credited with introducing the method of thinking aloud. In describing their instructions, we will distinguish a *main* part from some *complementary* parts.

The main part of the instruction to think aloud is usually very short, making reference to a procedure that is presumed to be already familiar to the subjects.

> "*Try to think aloud.* I guess you often do so when you are alone and working on a problem." (Duncker, 1926; italics in original)
>
> "Think, reason in a loud voice, tell me everything that passes through your head during your work searching for the solution to the problem." (Claparède, 1934.)

Thus the subjects are asked just to vocalize their thoughts, which are apparently presumed to have the form of inner speech. If the presumption is correct, it is not surprising that such short instructions could elicit the desired behavior. Verbalizing, under this assumption, would be quite simple (Level 1), because of the oral code and sequential structure of the internal speech.

In Claperède's instructions, however, the subject is asked to verbalize "everything that passes through [his] head," whether encoded orally or not. In order to comply with such an instruction, the subject would in many cases have to label and encode the content of STM, thus requiring the kinds of recoding processes that are postulated in Level 2 verbalization.

TA instructions frequently contain other information or requirements, which we have already refered to as "complementary." One complementary instruction that is often included is a request for *completeness.* Here are two examples:

> "The chief thing is to talk aloud constantly from the minute I present the picture, for I want to get everything you happen to think of, no matter how irrelevant it may seem." (Patrick, 1935)
>
> "I am not primarily interested in your final solution, still less in your reaction time, but in your thinking behavior, in all your attempts, in whatever comes to your mind, no matter whether it is a good or a less good idea or a question. *Be bold!* I do not count your wrong attempts, therefore speak them all out." (Duncker, 1926; italics in original)

There appear to be at least two ways in which this instruction could be implemented by a verbalization procedure: (a) A superordinate monitoring process could be incorporated in the procedure to guarantee that every item would be verbalized; (b) In quite an opposite direction, the effect of the instruction could be to eliminate any censorship process intervening between awareness of STM contents and their vocalization. Some TA instructions imply the presence of, and require the elimination of, such censorship:

> "Don't plan what to say or speak after the thought, but rather let your thoughts speak, as though you were really thinking out loud." (Silveira, 1972)

Another kind of complementary instruction is a request for *explanation.* In order to get as full an understanding as possible of the subjects' processes, they are asked to explain their thinking:

> "In order to follow your thoughts we ask you to think aloud, explaining each step as thoroughly as you can." (Smith, 1971)

As follows from our earlier discussion, inducing the subject to explain his solution very likely changes the structure of his thought processes. In a school situation with mathematics tasks, where the students were accustomed to explaining their solutions aloud, Krutetskii (1976) took special pains to warn his subjects against confusing the instruction to think aloud with that of explaining the solution:

> "Do not try to explain anything to anyone else. Pretend there is no one here but yourself. Do not tell about the solution but solve it." (Krutetskii, 1976)

Relatively frequently, complementary instructions are included to influence the *content* of the verbalizations. Investigators are sometimes specifically interested in certain types of mental content or mental processes, and ask their subjects to take special care to report instances of such content or processes (Johnson, 1964; Scheerer & Huling, 1960; Webb, 1975). Other investigators have asked subjects to verbalize what they are doing physically as well as what they are thinking (Hafner, 1957; Mayzner, Tresselt, & Helbock, 1964). Still other investigators have asked subjects to verbalize what they are perceiving visually (Farley, 1974; Huesmann & Cheng, 1973). In many of these studies, the exact wording of the instructions is not reported, which makes it difficult to discuss what verbalization procedures are necessary to comply with them. In most cases, however, it would be necessary to scan for specific content and to encode it for vocalization when found. Such monitoring would be, in fact, self-observation at the third level we described earlier.

One can only conjecture to what extent subjects pay attention to, understand, and comply with TA instructions, especially their complementary portions, and especially when the instructions are only presented at the beginning of sessions that may be an hour or more in length. There are many other conditions in a TA experiment that may affect the verbalization procedures of subjects as much as or more than instructions do. We turn next to these additional factors.

Effects of Training and Reminders

In many TA experiments, the subject is given initial warm-up problems to acquaint him with the experimental situation and accustom him to the microphones and tape recorders. In some studies, more extensive warm-up procedures are used explicitly to *train* the subjects to conform to the TA instructions.

During warmup, the experimenter feels free to interfere with and disrupt the subject, while during the experiment, he is very concerned not to interfere. An essential merit of TA as compared with introspection, however, is that the former is a normal mode of processing that does not call for the extensive training needed for the latter. Hence, the training in TA paradigms is negligible compared with that employed in classical introspection experiments.

The experimenter is generally, though not always, present during TA experiments. In the earlier studies, the experimenter had to be present, since there was no other means for recording the subject's verbalizations. Although tape recorders are now used almost universally, the experimenter is still usually present, primarily to monitor the verbalizations by reminding the subject to speak when he lapses into silence. These reminders, given after 15 sec to 1 min pauses (the interval being different in different studies), are generally standardized, taking the form of "keep talking" or "what are you thinking about?"

In some studies the experimenter also monitors the content of the verbalizations, and when necessary, asks the subject to explain what he means by something, and/or asks him to explain his solution processes. One may conjecture that the mere presence of the experimenter may induce some subjects to provide descriptions or explanations that they would omit in a non-social situation.

Reminders to verbalize of the "keep talking" variety should have a very small, if any, effect on the subject's processing. However, a reminder of the type, "what are you thinking about?" is more likely to elicit a self-observation process or produce an other-oriented description as a response. In cases where it is desired that the subject produce specific kinds of verbalizations (see the previous section), the experimenter's prompts may elicit the desired behavior, while the subject may be less likely to comply if only initial instructions are given.

Some investigations use TA procedures where the experimenter is not present (Klinger, 1974). It is technically feasable by automatic means to detect pauses that exceed a given length, and to remind the subject by means of an (auditory) signal to resume verbalizing.

REVIEW OF EMPIRICAL STUDIES

In this review of empirical studies that employ TA procedures, we will be concerned primarily with studies that meet the criteria of Level 2 verbalization, for most of the experiments in the literature de-emphasize speed and instruct the subject to "take your time and concern yourself with performance." This does not mean that verbalizing cannot remain at Level 1, even for complex tasks, but in most cases the additional information obtainable when recoding is permitted is judged to be more important than strict invariance of performance.

Studies of Level 2 Verbalization

The criteria for Level 2 verbalization are that only the information attended to (i.e., held in STM) should be verbalized, and that the recoding of this information for purposes of vocalization should not otherwise alter the processing involved in the task performance. It is often difficult to evaluate whether the criteria of Level 2 are satisfied in a given experiment, but in some cases the trace of the solution for a control condition of "silent" subjects provides sufficient information to permit a comparison of the structure and course of their solution process with that of vocalizing subjects.

In only a few experiments explicitly designed to study the effects of verbalizing, have experimental and control groups been exposed to identical conditions. After describing these studies we will go on to others, where experimental and control conditions, while not identical, are closely similar.

Roth (1965) investigated the effects of TA instructions upon performance in a series of problems. In order to test some hypotheses about how much verbalization would be elicited under different conditions, he varied factorially: (1) warm-up or pretraining for verbalizing, and (2) the extent to which subjects were allowed to manipulate objects while generating the solution. Here we will concern ourselves only with performance on the problem. Roth found no difference in time to solution that could be attributed to verbalizing. Since he did find significant differences due to problems, and to an interaction between problems and allowing manipulation of objects, his experimental measurements were not insensitive to manipulations of conditions.

In a study of discrimination learning, Karpf (1972) compared 40 subjects who were instructed to think aloud with 20 control subjects. The subjects were divided into two matched groups on the basis of ten preliminary problems, and were then given 15 experimental problems where the experimental group was asked to think aloud. Finally, five problems were given, where *all* subjects were instructed to be silent, to detect after-effects of thinking aloud. No reliable differences in number of correctly solved problems were found between the TA group and the control group, for either the experimental problem or the final problems. However, the TA group took about 50 percent more time than the control. Karpf gave some evidence suggesting that the effect of verbalization may have differentially affected the subjects (i.e., helping some and impeding others). This kind of effect will be discussed later.

Walker (1982) instructed subjects, under TA and standard recall conditions, to recall members of familiar categories (e.g., automobiles, soups, detergents). There were large differences among categories in the numbers of members recalled, but no differences between the TA and silent conditions. As would be expected from our theoretical analysis, the verbalizations of TA subjects consisted of retrieval cues and previously recalled names.

When Carroll and Payne (1977) compared TA with control subjects in a study of parole decision-making, no reliable differences were found for speed of decision, type of decision, or information requested while making the decision. In another study, Smead, Wilcox, and Wilkes (1981) had subjects choose between brands of coffee-makers, while confronted either with the products or with verbal descriptions of their attributes. Eye fixations were recorded, and subjects were asked afterwards to rate the realism and difficulty of the judgments and their certainty in their final choices, and also to state what attributes were most important in their choices. No differences between TA and control subjects even approached significance, though several differences were found between subjects shown actual products or verbal descriptions, respectively.

In another experiment (Johnson & Russo, 1978) the subjects were making choices among consumer items, some of the subjects with an additional instruction to think aloud. In a subsequent test for memory of the stimulus information, no differences attributable to thinking aloud could be found either in accuracy or latencies. Feldman (1959) studied a single subject's predictions in a binary choice situation with a TA instruction. Two large control groups of subjects were run in the same condition without instruction to think aloud. No differences in types of choices and number of correct predictions were observed. In a study on the effects of covert modeling of assertive behavior, Kazdin (1976) included a comparison between a silent condition and a condition where subjects narrated aloud their ongoing imagery during the correct modelling. Due to insufficient specification of the instruction to verbalize, we are not certain that it should be seen as a pure TA instruction. Still, it is informative that no differences due to the instruction to verbalize were found in an extensive analysis of the effects of the covert modeling.

Several studies allow comparisons between a silent and a vocalizing group, for identical or similar tasks, and with only slightly different experimental conditions for the two conditions. In some instances, these studies measured success or speed in solving problems, in others, they also reported data on the solutions paths that were discovered.

In several studies, evidence can be found that verbalizing had no effect on the basic performance measures or the gross structure of the thought process. The proportion of subjects solving Duncker's candle problem did not differ between a TA condition and a silent conditon in a study by Weisberg and Suls (1973). In a study of intransitive preferences, Montgomery (1977) replicated a study by Tversky (1969), but with the addition of a thinking-aloud instruction, and found approximately the same choice distributions. Studying anagrams at different levels of difficulty, Sargent (1940) was able to show, by correlational techniques, that difficulty level produced changes in the types of processes used in *both* silent and TA subjects. No more detailed comparisons were made.

More detailed comparisons of the structures of the thought processes were made by Newell and Simon (1972) on problems in propositional logic, by Bulbrook (1932) on insight problems, by Ericsson (1975b) on a block puzzle, and by Flaherty (1974) on algebra word problems. For subjects discovering proofs in propositional logic, Newell and Simon compared the number of solutions and the detailed solution paths of their seven TA subjects with the solutions (collected by different investigators at Yale) to the same two problems by 64 subjects under silent conditions. With so few subjects in the TA condition, it could not be expected that differences could be found between the two conditions in gross measures of performance, nor were they. More significant is the fact that when the detailed structures of the solutions (i.e., the specific proof steps) were examined and compared between the two groups, no differences were found. Both groups explored essentially the same parts of the problem space with about the same relative frequencies.

For a large number of different insight problems, Bulbrook (1932) found close agreement between the solution patterns reflected in subjects' TA protocols, and answers by a large group of silent subjects to a retrospective questionnaire.

In Ericsson's (1975b) study on problem solving with the Eight Puzzle, two separate experiments were conducted for the two conditions, but the same sequence of puzzles was presented to the subjects in both conditions. In one experiment, the subjects were instructed to think aloud and had to tell the experimenter what tiles to move; in the other experiment, the subjects, sitting alone, pressed keys on a teletype to make the computer change the puzzle configuration. Differences in the structures of the solutions could be investigated through analysis of the sequences of moves, which were recorded in both conditions (Ericsson, 1975a, 1975b). No differences were found in the subjects' attainments of objectively defined subgoals and subrelations of the goal, nor in the structure of the search trees. However, the mean number of moves was significantly larger for the verbalizing subjects for the first couple of problems. Ericsson (1975b) attributed this difference to a change in the effort that subjects devoted to mental planning, induced, in turn, by differences in the design of the two experimental conditions. In the study with the TA instruction, the subjects had to tell the experimenter what moves he should make for them, while in the "silent" study, the subjects were alone and inputted their moves on a teletype. This difference alone is enough to explain the observed difference between the numbers of moves in the two conditions.

Flaherty (1974) studied the effects of practice and verbalizing in a factorial design with algebra word problems. Flaherty used an extensive coding schema of 17 variables to describe the TA solutions; ten of these variables could also be assessed from the written solutions of the silent subjects. Analysis of variance showed no effects on problem solving performance of either practice or verbalizing. The verbalizing subjects did not even take a longer time to solve the problems. T-tests for the ten coded variables showed only one significant difference between verbalizing and silent subjects: the verbalizing subjects made more computational errors. Since, unfortunately, the silent subjects solved their problems in groups and their written solutions constituted the data, while the TA subjects were tested individually by the experimenter and their written solution as well as the verbal protocol made up the data, it is difficult to know whether the observed difference in number of mistakes is due to verbalizing or to the other differences in experimental conditions or the data available for analysis.

In several studies using verbalization, models have been constructed on the basis of the verbal data, and used to predict the subjects' behavior in subsequent silent situations. In a study on consumer choice, Bettman (1970) showed that his formalized decision-net models of two consumers, based on TA protocols obtained on earlier shopping trips with these subjects, could predict 85% to 90% of the products selected by the subjects. Clarkson (1962) studied and modeled a trust officer's selection of portfolios of securities on the basis of several TA protocols. When, a year later, the trust officer selected four different portfolios, the model predicted (without interaction with the subject) the companies whose shares he would buy with 80% to 85% accuracy. These results suggest that the information provided by the TA protocols for the model building does accurately reflect the normal course of the thought processes.

Finally, we describe a couple of miscellaneous results that have implications for our problem. In a study by Marks (1951) where all subjects were thinking aloud, performance was significantly affected by asking subjects, "What are the elements of the problem?" This result suggests that thinking aloud does not by itself enforce an analytical approach, as has sometimes been alleged.

Subjects have occasionally been asked to judge the effects of their verbalizing. No interference was reported by subjects who were making decisions (Svenson, 1974). The experimental setup, including the TA requirements, was judged representative by most artists and non-artists engaged in writing poems (Patrick, 1935) and drawing pictures (Patrick, 1937). Some of Dominowski's (1974) subjects reported that verbalizing led them to more careful information processing. However, Dominowski's experiment may not have met the criteria of Level 2 verbalization, for he instructed his subjects to describe the stimuli and to give reasons for their classification responses, even though guesses and other informal explorations were explicitly said to be acceptable.

Differential Effects of Thinking Aloud

Several investigators (Claparède, 1934; de Groot, 1965) have observed marked differences in the spontaneity and richness of TA protocols and suggested that these may be related to the ease and skill with which subjects transform non-orally coded information to external speech. Johnson (1964) noted individual differences in the pattern of verbalizing, some subjects giving fluent verbalizations and some adopting a "think-then-summarize" procedure. Some subjects may transform automatically, whereas others may require a conscious effort for this transformation and verbal production. These differences may be due to the compatibility of codes, some subjects using visual images while others are more verbal. Although there are no studies of individual differences across tasks requiring different codes, some support for the transformation hypothesis can be found in an observation by Klinger (1971) that only highly verbal females could give good verbalizations in a reverie condition, which is known to employ mostly visual images.

An instruction to verbalize and think aloud could help some and impede others. We know of only two studies that have found reliable differential effects on subjects. Flaherty (1973, 1974) separated subjects who were sensitive to the fact that some problems were physically impossible from subjects who were not sensitive (the latter presumably relying mainly on verbal cues for generating the solution). She found a reliable difference in performance between these two groups of subjects only in the think-aloud condition. She suggested that subjects who were verbally oriented problem solvers would be impeded by having to perform a second verbal task (i.e., thinking aloud). However, a reanalysis of Flaherty's (1973) data shows no reliable difference ($p > .15$ two-tailed t-test) between verbally oriented problem solvers in the TA and the control conditions, which would be the proper test for her hypothesis.

In the study by Karpf (1972) discussed earlier, the subjects were asked after the experiment whether thinking aloud helped, had no effect, or interfered. He found reliable differences between the "no effect" group and the "interference" group for the experimental problems requiring them to think aloud, but also for the final problems when *no* subject verbalized. One may have considerable doubt whether the subjects could assess that their worse performance was due to verbalizing, especially since their performance did not improve during the final "silent" trials.

Unless the experimenter, by constant probing, tries to eliminate the differences in completeness of verbalizing between subjects, thus possibly changing their cognitive processes, there appears to be little reason to anticipate differential effects of verbalizing. All the studies reviewed above support the conclusion that the observable structure of cognitive processes is not affected significantly by the instruction to think aloud when the experimental conditions are consistent with the criteria for Level 1 or 2 verbalizations. These findings suggest that the internal structure of the thought processes also is not changed as a result of the additional verbalizing activity. The few differences that were encountered in the published reports were not of major importance, and where differences appeared, they were very likely attributable to differences in the experimental conditions. In evaluating our conclusion, it is important to know that we have included all comparisons in studies complying with the conditions of Level 1 and 2 verbalization that we have discovered in the literature.

Studies Not Conforming With Level 2 Conditions

In the review of studies meeting the criteria of Level 2 verbalizing, we found no evidence of changes in the course or structure of the cognitive processes induced by verbalizing. We would not expect this result to hold in studies where the subject is asked to verbalize information that would not be heeded in the normal course of processing, or that would not normally be encoded in an oral code.

This distinction bears on the completeness of verbalizations. Our model assumes that *only information in focal attention* can be verbalized. In most theories of the structure of the human information processing system a distinction is made between fast, automatic processes that are not necessarily conscious (and which are often thought to proceed in parallel) and the slow serial processes that are executed under cognitive control–a distinction, that is, between pre-attentive and focally attended processes (Neisser, 1967), perceptual and cognitive processes (Simon, 1975), and automatic and cognitively controlled processes (Shiffrin & Schneider, 1977). In our discussion, we will adopt this distinction. We will also assume that with increase in experience with a task, the same process may move from cognitively controlled to automatic status, so that what is available for verbalization to the novice may be unavailable to the expert.

Several types of processes generally occur automatically, in this sense, and rapidly (in a matter of tens or hundreds of milliseconds): perceptual-encoding processes (recognition), memory retrieval processes, and motor processes. There are many instances during thinking-aloud studies where a subject acquires hypotheses "instantaneously" and directly, without evidence of prior related or intermediate stages; for instance, interpretations of complex pictures (Claparède, 1934), and "insights" in geometry problems (Henry, 1934). Retrospective reports of sudden acquisition of hypotheses are common in experimental settings (Woodworth, 1938) and in more informal demonstrations (Miller, 1962).

We have observed that there is no fixed and stable boundary between automatic and focally attended processes. With extensive practice, cognitive processes will develop into fixed automatic processes; and there is suggestive evidence that practice leads to a successive fading from consciousness of information about the process. There is such evidence, for example, in retrospective reports after each trial in experiments by Ach and Watt (Woodworth, 1938).

The fact that processes can continue to occur after retrospective reports about them have disappeared is urged by Leeper (1951) as a principal argument against the validity of verbal reports as data about processes. Even without retrospective reports, the Shiffrin and Schneider (1977) study makes it abundantly clear that an automatic process is very different from its cognitively controlled antecedent in terms of modifiability, speed, and so on. The automatic processes can be seen as units or chunks, and Schneider and Shiffrin suggest that the growth of skill can be viewed as the development of successively higher-level codes and organizations of lower-level processes. Hence, we should expect the verbalizations of highly skilled individuals to be less complete than those of less skilled ones.

If information is available (the process not automated) but not held in an oral code, then verbalization will require recoding (Level 2), and this recoding will have to *share focal attention* with other task-directed processes. We hypothesize that, in the case of such competition, the task-oriented processes will have priority over the recoding and verbalization processes. There is supporting evidence for this hypothesis in three different types of situations:

1. Subjects tend to stop verbalizing in conditions where they are giving indications of being under a high cognitive load. Such indications may take the form of reorganizations of the problem representation or strategy (Durkin, 1937), or direct expressions of feeling difficulty (Johnson, 1964).

2. When subjects attend to information that leads to direct recognition of appropriate actions, this information tends not to be verbalized (Duncker, 1945). In Maier's research on the pendulum problem (1931), subjects were retrospectively questioned on solving a problem just after being given a hint. Subjects who described the solution as emerging in a single step did not report any memory of the hint; while subjects who mentioned several steps in the solution all reported the hint that had been administered. With the available data it is not possible to determine whether all were aware of the hint *at the time it was given.*

3. In situations where the subject is judged not to have any competing task-directed processes, verbalization tends to be relatively complete. Ericsson (1975b) calculated how regularly an observable feature of the process (e.g., reversal of a pair of moves) was associated with a specific kind of verbalization, and found a very high correlation (e.g., on each occurrence of a reversal, the subject verbalized a negative evaluation of the original moves).

There are broad areas of cognitive behavior, however, where investigators consistently experience difficulties in obtaining rich verbalizations from subjects. We will focus on two of the most important kinds of processes where verbalization is poor: perceptual-motor processes, and visual encoding proceeses. These will be taken up in the next two sections. A small number of miscellaneous studies relating to other processes will be discussed in a third section.

Verbalization of Perceptual-Motor Processes

The problems of verbalizing perceptual-motor processes are most clearly visible in problem situations where the problem is represented physically (e.g., the disks and pegs of the Tower of Hanoi puzzle), and performance involves manipulation of this physical representation.

Thomas (1974) found that TA protocols from a variant of the Missionaries and Cannibals puzzle provided insufficient information on the course of the solution to specify a definite model of the process. Other studies with different kinds of puzzles (Ruger, 1910) and instructions to think aloud (Durkin, 1937) have shown that subjects have difficulties in expressing their thoughts verbally in these situations, and are biased towards direct manipulation of objects. Manipulation appears to lack a mediating symbolic representation that can readily be encoded into the verbal code. In the study by Durkin, many subjects expressed the feeling that thinking and manipulation were distinctly different, alternative, activities, even when both were directed to the solution of the problem.

From a study by Klinger (1974) using pure TA instructions, it appears that the amount of verbalization that occurs is not significantly less in these kinds of tasks than in others, but that there is a significant difference in the verbalized mental content. In manipulative tasks there is a relatively high frequency of higher-level evaluations of unverbalized solution attempts (e.g., "Yep," "Damnit," etc.), and of verbalizations denoting attention to control processes (e.g., "Let's see," "Where was I?"). When engaged in perceptual-motor manipulation, the subject does not appear to be aware of, and does not verbalize, the lower-level content or structure of his thought processes.

In order to increase verbalization of content, some experimenters using the TA method have changed the task by constraining the manipulations partially (Durkin, 1937) or wholly (Benjafield, 1971), thereby forcing the subjects to form an internal representation of the content that can be verbally encoded. When this is done, more of the content of the thought processes is, in fact, verbalized.

From a behavioral point of view, constraining manipulation significantly changes the task, for the overt moves have now become covert and are no longer amenable to direct observation. The subjectively perceived polarity between thinking and manipulation, mentioned above, may very well correspond to this difference between covert and overt trials. Duncan (1963) showed that giving subjects an explicit instruction to *think* resulted in significantly fewer overt trials in a switch-setting task, but an actual increase in solution time. Hence, the covert processing, in the condition that induced more planning before manipulation, took longer than the corresponding overt trials. A similar tradeoff was noted by Shipstone (1960) in a concept-learning task, with an instruction to "slow down and use logical procedures." In a switch-setting task, Ray (1957) found that an instruction that required the subject to tell the experimenter what he was going to do before he started to manipulate the switches significantly decreased the number of overt trials to solution. Ray does not report the corresponding solution times.

Most of these results could be explained by the hypothesis that, in response to the instruction to "think" or verbalize, the subject did not change the structure of his processing, but simply substituted overt trials (the measured index of performance) for non-observable covert trials. However, it is reasonable to assume that the internal representation generated for the covert processing improves memory and the organization of the processing.

Our theoretical model predicts that an instruction to verbalize motives or reasons for one's thinking will likely change the course of processing. In general, when the thinking-aloud instructions do not require such verbalizations, the protocols do not contain them. We now review three studies where subjects were specifically instructed to report motives or reasons for their actions.

The study by Gagné and Smith (1962) with the Tower of Hanoi problem was aimed at investigating the effects of different verbalization instructions upon performance during some training tasks (two-disk to five-disk problems), and upon transfer to a similar but more complex task (six-disk problem). One of the two factorially combined manipulations during the training tasks required the subjects to state verbally a reason for each move. This requirement greatly improved performance on the transfer task, both as to number of moves required and time taken to find a solution. As a second manipulation, the subjects in one pair of conditions were also instructed to search for a general principle behind the different versions of the problem. This instruction did not affect performance on the transfer task.

In the training tasks, the verbalization group produced more efficient solutions (solutions having fewer moves), indicating that the instruction to verbalize the reasons induced more deliberate planning, in addition to its effect on transfer. Although no formal record was kept of the time taken for each move, the experimenters judged this to be longer for the overt verbalization group, but they reported that this extra time was "filled with the articulation of the reasons for the moves." They suggested that the instruction to verbalize the reasons for the moves affected performance by "forcing the subjects to think."

In a follow-up study with the same basic design, Wilder and Harvey (1971) investigated whether the overt verbalization was crucial, or if equivalent results could be obtained with a firm instruction to state the reasons covertly; in addition they checked the time taken to achieve solutions during the training tasks. The results showed no difference between the two verbalizing conditions, but a clear reduction in the number of moves in both those conditions as compared with a control condition. The time to solution did not differ among the three conditions, during either the training tasks or the final task. This finding eliminates the hypothesis that the advantage in transfer shown for the verbalization condition in this experiment was attributable to extra learning time during the training sessions.

In another follow-up of Gagné and Smith, Davis et al. (1968) included the presence of the experimenter as an additional dimension in a factorial design. In this study the subjects were instructed to verbalize their reasons at the beginning of the experiment, but the instruction was not repeated; in the earlier experiments, the subjects were closely monitored during the entire session to make sure they followed the instructions. In the Davis study, verbalization had no effect on the training task (5-disk problem), but significantly reduced the number of moves required to solve the test problem (6-disk problem). The experimenter's presence increased verbalization on the training problem, but not on the test problem. There was no interaction between amount of verbalization and presence or absence of the experimenter. Unfortunately, no information is given about the solution times.

These three studies show that a requirement to verbalize reasons and motives has substantial effects on both immediate performance and learning, and thus indicates that generating verbalized reasons brings about changes, at least in manipulative tasks, in the normal course of the processes. Here, as in the studies discussed earlier, we do not know to what extent forcing subjects to give reasons for their actions simply causes them to substitute overt trials for non-observable covert trials and planning. The negative result of the Davis et al. experiment is most readily interpreted as showing that, in a problem-solving situation with a heavy cognitive load, initial instructions may be disregarded by subjects unless they are monitored by the experimenter. An analysis of the content of the verbalizations, not provided by the authors, would be required

to test the plausibility of this explanation. Finally, we may conjecture that the presence of alternative solutions to the Tower of Hanoi problem probably increases, in comparison with other tasks, the sensitivity of thought processes to instructions to verbalize reasons.

Verbalization of Visual Encodings

There is compelling evidence to support the distinction between a visual representation or code and an oral or symbolic representation or code when subjects are presented with drawings or pictures. An instruction to describe a visual scene verbally should require an oral encoding of the picture, which will imply extensive processing.

The task of viewing a novel visual scene has been studied to detect effects of imposing an additional verbalizing task, in this case to describe the scene verbally. A study by Freund (Loftus & Bell, 1975) showed that subsequent recognition of scenes was much improved by the verbalizing requirement as opposed to viewing without it. On the other hand, when an unrelated verbal task, like counting backwards by threes, is imposed, the subsequent recognition of scenes deteriorates as compared with normal viewing, but not to a chance level (Loftus, 1972; studies of Freund and Szewczuk described in Loftus & Bell, 1975).

Loftus and Bell (1975) have studied the process of recognizing pictures viewed with the instruction to remember them for subsequent recognition. At the time of judging whether or not a picture had been presented before, the subjects were asked if their judgment rested on a specific detail or on general familiarity with the picture. Performance was very much better when a detail could be named than when the subject simply cited general familiarity. From recordings of eye movements during the initial presentation of the pictures in an earlier study, Loftus (1972) found that the detail verbalized during the recognition trials was the detail that had received most fixations at the initial presentation. Loftus and Bell propose that all these results can be explained by the notion of two different codes. The nonverbal encoding is direct and independent of the verbalization instructions. The extent to which an oral encoding is also generated is influenced by the instruction to give a verbal description.

Evidence supporting the distinctness of the two codes comes from an investigation by Schuck and Leahy (1966) on fragmenting visual images. They found that subjects reporting the disappearances verbally tended to report omissions of meaningful complete segments, while control subjects who traced the disappearances on an outline of the image did so to a lesser degree. However, subjects in the two conditions were not treated identically. The verbalizing subjects gave reports continuously, while the others were given time to trace images when they appeared. Schuck (1973, 1977) carried out some subsequent experiments that equated the two conditions more carefully, and then found no statistically reliable differences between the two groups of subjects. It would appear that when sufficient time is allowed them, and the degree of detail that is wanted is specified, the verbal descriptions that subjects give of fragmented images are basically veridical. This is especially interesting since the task obviously calls for recoding of the perceived stimulus.

Studies on concept formation using the Vigotsky test (Hanfmann, 1941; Hanfmann & Kasanin, 1937) have identified interindividual differences in strategies which have a bearing on the effects of instructions to describe stimuli verbally. The test employs a number of blocks varying in color, shape, and size. Each block has one of four names written on its reverse side, and the subject is to find the characteristic that serves as basis for the naming. Subjects classified as "conceptual" did not inspect the blocks visually very much, but spent most of their time in thinking, using a symbolic representation of the blocks in terms of their attributes. In this group, for example, the hypothesis that color might be one of the relevant characteristics is checked by counting the colors rather than moving the blocks. This group showed a strong preference for shape as the relevant attribute, and were disappointed with the solution, which was based on color.

The subjects classified as "perceptual" constantly manipulated the blocks, and appeared to get their ideas for organizing the blocks from looking at them. Typically, they first made groupings by manipulating the blocks, and then discovered or formulated a general principle. Retrospective reports from these subjects on how they found the solution were particularly scanty, and they often had no more to say than that it seemed natural to group the blocks as they had. These subjects seemed to favor color and size as the relevant characteristics, and were fully satisfied with the solution based on these characteristics. Consistently with this preference, the perceptual subjects were significantly better at the task than the conceptual subjects.

A reasonable hypothesis about the effect of requiring subjects to verbalize explanations in such a task would be that subjects with a preference for perceptual processing would alter their processing strategy, and hence their performance. In a study by Brunk et al. (1958), subjects were given an initial test of the Vigotsky type, and then a second, similar test. In one condition on the second test, each subject was "requested to tell why he placed each block where he did." In a control condition, no such explanation was requested. The correlation of subject scores between initial and second test was significantly lower under the instruction to explain than under the control condition, as the hypothesis would predict.

Using several highly perceptual tasks, like the Rorschach, the stencil test and a block test, Goldner (1957) was able to identify in data from TA protocols consistent individual differences along the perceptual-cognitive dimension (in his terminology, between processing wholes and parts of the stimulus information). For the stencil test, Hafner (1957) found (though the differences only approached significance) that instructing subjects to verbalize what "they were thinking and doing" improved their test performances, but also required more time. His instruction to verbalize is poorly documented, but the theoretical context of his study suggests that there were important differences from the assumptions of Level 2 verbalization. Hafner's results seem to reveal the same effect as Brunk et al.'s (1958) study.

In a series of studies reported in Merz (1969), the effect of verbalization on performance in intelligence tests was investigated. In a study by Kesting using the Figure Reasoning test, the 13- to 17- year-old subjects who always had to say aloud how the figures were alike or different were significantly better than the subjects who were asked to say the same thing to themselves silently; but they were also significantly slower. In this study, there were also two other conditions with additional verbal interference tasks; one group of subjects had to say "eins, eins ..." rhythmically, while another group had to sing "la, la ..." while solving the test items. These conditions have the same performance (in terms of number of correct solutions and time required) as the silent verbalizing condition; which suggests that the same processes, mainly nonverbal in character, were used in all three conditions.

This oral-encoding interpretation is supported by Waszak (reported in Merz, 1969), who showed that hearing one's own voice in a heavy noise condition does not eliminate the improvement associated with verbalizing. The crucial mechanism underlying these results appears to be that the instruction to explain answers increases the tendency to encode figures and their differences verbally. This interpretation is also supported by the observation of Merz that it is difficult to induce subjects to verbalize in this kind of perceptual task.

To test the hypothesis that the instruction to verbalize made the subjects assume a more analytic problem-solving style, Hofgren (Merz, 1969) compared performance on a parallel form of the Figure Reasoning test between a group that had previously verbalized on an initial form and a control group that had not. The verbalizing group performed significantly less well when not required to verbalize, but still somewhat better than the control group, whose performance hardly differed between the two occasions. Merz suggests that the loss of performance by the "verbalizers" on the second occasion was due to transfer of thought content, as the verbalizing subjects solved more test items on the first occasion, rather than to transfer of a more analytic style of thought. Supporting evidence is given in a study by Waschke (Merz, 1969): the verbalizing and control groups performed equally on the first occasion, when simple items were used; and the performance of both groups was the same on the second occasion, with respect to time and number of items correct. These investigations indicate that the instruction to verbalize forces subjects to encode the figures orally, and that this, in turn, induces verbal symbolic processing that is especially well adapted to discovery of the prescribed solutions in the Figure Reasoning test.

In the above studies, performance was averaged over subjects and different tasks. We now consider evidence of differences between individuals, and of differential effects, on performance with tasks of varying difficulty. Merz (1969) reports a smaller study showing that with simple test items, verbalization did not help the subject and amounted more or less to a retrospective description of the solution "seen" immediately. The verbalized description was only helpful with more difficult items that required some search for the right solution.

Inner speech, measured by electrical activity in the speech apparatus with Raven's Progressive Matrices items and without verbalization shows a similar pattern, in that inner speech activity increased with difficulty of the items (Sokolov, 1972). Analyzing the protocols from verbal reconstructions of the problem-solving process, Sokolov showed that the simple items were solved in a predominately visual way, while with the more difficult problems, verbal designations of some features of the figures were used to aid solution. In outlining a scheme for the interplay of visual and verbal processes in solving such problems, Sokolov points to the influence of verbalization in attending to features that would have gone unnoticed in the purely visual analysis. In this interpretation, the directed verbalization provides the subject with additional noticed features, which in turn facilitate performance.

There is little evidence as to whether directed verbalizing has a general or a differential effect on subjects. A general effect is suggested by the fact that the variance of the increased performance with verbalizing is equal to or less than the variance of performance under control conditions in the studies of Kesting and Waszak and Hofgren (Merz, 1969). The investigation by Sokolov indicates differences between subjects, leading Sokolov to propose a differential reliance on verbal and visual processing; but this result is not incompatible with the idea that there may also be general effects upon all subjects.

In this context we also want to discuss the well-known effect of providing verbal labels to visual stimuli, which was originally demonstrated by Carmichael, Hogan, and Walters (1932). We will rely heavily on the relevant parts of Riley's (1962) excellent discussion of memory for form. Carmichael et al. showed clearly that naming one of two possible labels for an ambiguous stimulus before its presentation led subjects to draw the pictures more like the label in a subsequent memory test. A series of studies have shown that this effect is not a purely perceptual one, at least when the presentation time is sufficiently long for complete perception (Bruner, Busiek, & Minturn (1952) found effects for presentations shorter than a tenth of second). Hanawalt and Demarest (1939) showed that verbal labels presented at recall gave similar effects when the pictures had been unlabeled at presentation. Prentice (1954) showed that presenting verbal labels with the pictures produced no bias in perception.

Additional insight into the effects of labels is provided by a study by Herman, Lawless, and Marshall (1957), who included a control group that was simply shown the pictures. (The original study by Carmichael et al. had a very small control group.) Herman et al. (1957) found that the subjects in the control group showed as much tendency as the others to distortions of memory. However, in the control group the memories were distorted equally in the directions of the two labels, while in the experimental groups, they were distorted mainly toward the label that was presented. Yet even in the experimental groups a few subjects deviated toward the verbal label that was *not* presented. Hanawalt and Demarest (1939) found that subjects who were not given a verbal label at presentation reported meaningful associations to the pictures, and a substantial number of these associations were similar or even identical to the labels used originally by Carmichael et al. Herman et al. (1957) confirmed this by asking a new group of 29 subjects to give free associations to the pictures. On the average, 63 % of the associations were identical to or very similar to the ones used by Carmichael et al. Furthermore, Herman et al. (1957) showed that the effects of the labels could be markedly reduced if subjects were told in advance that they would be asked to draw the pictures later. In sum, the original effect observed by Carmicheal et al. appears to be caused by inducing subjects to use certain associations at the expense of others. Hence, it is entirely possible, even likely, that subjects could easily perform this memory task with TA or retrospective report instructions without changes in performance.

Other Verbalization Studies

None of the studies that remain to be discussed employ highly manipulative tasks or pictorial stimuli. In a study on clinical judgment (Baranowski, 1975), the subjects, who were psychologists, made two successive series of judgments. On the first occasion, all subjects performed the task under identical conditions. On the second occasion, the subjects were divided into: (a) a group working under the same instructions as on the first occasion, and (b) a group instructed to verbalize and monitored by the experimenter who asked the subjects questions whenever "a particular profile could use more explanation" (1975, p. 21). There was no difference between control and verbalizing conditions (measured by variance accounted for by linear and non-linear models). However, the cross-validated linear models over the two occasions accounted for significantly less variance for the verbalizing group than for the control, suggesting that the instruction to verbalize changed the utilization or subjective weights of the cue-variables.

In a related study, Fidler (1979) had subjects choose for admission to graduate study between two students described in terms of several attributes. Subjects gave both concurrent and retrospective reports. (Unfortunately, the concurrent reporting instruction asked subjects to provide certain specific information, like the reasons for judgments, how they evaluated information, and so on.) Retrospective reports were elicited by asking subjects how they had reached their decisions. Fidler compared the performance under different reporting instructions (including no verbal report at all). No reliable differences in choice were observed, nor were subjects more consistent in making choices on repeated trials with verbal report instructions than without such instructions. When regressions were estimated to determine each subject's decision model for verbal-report trials and no-report trials, respectively, no

reliable differences were found for 10 of the 13 subjects. The reliable differences in the rules for the three remaining subjects could be traced to stable changes in their decision rules that occurred in the transition between blocks of report and no-report trials. Whether these changes in decision rules were induced by the reporting instructions or other factors is not certain. Fidler proposes that it may be due to the inexperience of the subjects. Since they were not experts at the judgment task, they were aware of their decision rules, hence could report them. The consistency between reported reasons and decision rules, on the one hand, and actual choices (even on no-report trials) lends further support to that interpretation.

In a concept-learning study by Bower and King (1967), one group of the subjects was required to verbalize their hypotheses before classifying the stimuli, while a control group was not. In preparation for the experiment, the subjects described the stimuli, in order to ensure that subject and experimenter agreed in their descriptions. Under these circumstances, we would expect that no further oral encoding need occur in the verbalizing condition. The number of irrelevant dimensions of the stimuli was varied, although the instructions indicated which two features were relevant to the solution in each case. The requirement to verbalize hypotheses improved performance significantly (i.e., number of responses to criterion), but only for the first problem. Bower and King found that variation in the number of irrelevant features or dimensions was only effective on this initial problem, suggesting that the verbalizing of hypotheses helped the subjects initially to ignore the irrelevant attributes. It should be noted that no training trials were used in this study. Karpf and Levine (1971) found no effect on performance in a discrimination learning experiment from asking subjects to repeat their hypotheses verbally before each trial.

In a cue-probability learning task, Brehmer (1974) required one group of subjects to describe the rule underlying their predictions just after each prediction was made, but before feedback was received. The subjects' descriptions should be so explicit that another subject could understand and use it; if they did not meet this standard, the experimenter prompted for more information. Explanations like "I guessed," or "I remembered from the previous trial," were accepted as verbal descriptions. An analysis of variance showed no significant effect or interactions associated with verbalization. In a subsequent study (Brehmer, Kuylenstierna & Liljegren, 1975) the subjects wrote down their current hypotheses in a booklet at the beginning of the test blocks, without any significant effect on performance.

According to our model, requiring verbal explanations of behavior should not alter the normal processes unless the information required for the verbalizations would not otherwise be generated. Unfortunately, there is little evidence for the tasks used in the above studies about the content of undirected verbalizations. In the cue-probability experiment of Brehmer (1974) with very simple stimuli (a straight line varying in length), the number required for explaining the rule was most likely consciously generated even in the silent condition. That the effects of verbalization were limited to the first trial in the Bower and King (1967) experiment could be attributed to the fact that verbalizing helped the subjects ignore irrelevant features. Alternatively, one might speculate that verbalizing may have speeded up the generation of an internal representation, thus making the subject more independent of his direct perceptions. In the study of Baranowski (1975), unlike the other studies, the subjects were highly skilled. In our model, verbal explanation of automated activities would be cumbersome and change the course of the processing from a largely perceptual (recognition) to a more cognitive one. In support of this hypothesis, the time taken by the clinical psychologists to perform the task with verbalization was two or three times the time taken in the silent condition.

Effects from Retrospective Verbalization

We turn now to some experiments that are often cited to show that verbalizing information (even retrospectively) may change and deform it, and hence affect subsequent task behavior. Hendrix (1947) showed that an instruction to describe a concept or principle verbally after learning reduced ability to use the concept in a transfer situation. The results were substantiated in subsequent work by Schwartz (1966) and by Phelan (1965). Careful analysis shows that these studies do not address the question of verbalization, as such, but rather verbalization of explicit and logical concepts. There are two issues. The first is that if the subjects do not normally organize what they learn in these experiments in verbalizable concepts and general principles, then verbalization forces them to generate such concepts and principles from whatever information is currently available to them. The reformulation may not at all reflect the way in which the learning was actually encoded. For example, Phelan found that the verbal descriptions of certain pictorial stimuli tended to contain discriminative features different from those that defined the concept the subjects had learned. In Chapter 3 we return to the issue of what information should be reportable in such cases.

The second issue relates to the detail and explicitness called for by the instruction to verbalize. Sowder (1974), examining the effects of various sorts of verbalizations of learned gerneralizations, found no differences as compared with a control condition. He also cited two studies that found no effects of producing written descriptions of learned generalizations. Sowder proposed that the important differences between his study and Hendrix's was that Hendrix (1947) specified precisely the content of her subjects' verbalization (quantifiers, domain, and so on), while he did not.

Rommetveit (1960, 1965) and Rommetveit and Kvale (1965a, 1965b) studied concept formation in a situation where 12- and 13-year-old subjects played on a wheel of fortune, different pictures being displayed when subjects were to win or lose, respectively. They found that instructing the subjects that they were subsequently to describe differences between the "win" and "lose" figures, as opposed to just playing on the wheel, influenced retrospective descriptions of the two figures. Other procedural variations–like demonstrating before the experiment how the figures differed (Rommetveit, 1965)– tended to eliminate a (correct) tendency toward associating roundness with good figures. In these studies, therefore, the effective variable is not verbalization per se, but directing the cognitive processes by instructions. Without such direction, verbalization appears to have no effect on the cognitive processes.

The fact that encoding and reporting verbally takes time creates a procedural difference, which is important in some studies, between the reporting and control conditions. Time is known to be an important variable in LTM phenomena. Boersma, Conklin, and Carlson (1966) allowed their subjects in the verbal report condition an additional minute to specify their encodings of each stimulus. Retention scores were superior for the verbal report condition, but the experimental design confounds the effect of the additional time with effects of generating a written description of the stimulus.

In sum, the empirical data from the modest number of systematic studies of the effects of verbalizing (e.g., Brehmer, 1974; Brehmer, Kuylenstierna & Liljegren, 1974, 1975; Karpf & Levine, 1971; Wilson & Spellacy, 1972) support consistently our assertion that producing verbal reports of information directly available in propositional form does not change the course and structure of the cognitive processes. However, instructions that require the subjects to perform additional cognitive processes to generate or access the information to be reported may affect these processes.

New Research on Effects of Verbalization

The renewed interest in psychology in using verbal reports as data has generated new research activity aimed at casting further light on the effects of verbalization upon cognitive processes. This new work seeks to go beyond proving the presence or absence of effects of verbalization, and to measure the extent and nature of these effects under different experimental conditions. We will conclude our discussion by a summary of three important recent studies of this kind.

In a dissertation, Karat (1983) analyzed the verbalizations of subjects solving the Tower of Hanoi with TA instructions. He found that subjects verbalized plans and goals before making some, but by no means all, of their moves. If subjects were required to give reasons for each move, as in the original study by Gagné and Smith (1962), they would have to generate these reasons through additional cognitive processing. Since Karat found no effects due to TA, the effects found by Gagné and Smith can very likely be attributed to their more specific and demanding verbalization instructions.

The most ambitious attempt to explore and document effects of instructions to verbalize is an unpublished study by Russo, Johnson, and Stephens (in preparation). In their study, 24 subjects performed 55 trials of 4 different tasks with 5 different instructions. The five instructions included a control condition, concurrent verbalization, retrospective reports with the problem still in view, retrospective reports without any stimulus information, and retrospective reports cued by the eye-movements recorded during the immediately preceding solution. In the following, we will only discuss the data for the control condition, and for concurrent and retrospective reports without stimulus information. The four tasks were: selection between gambles, items from Raven's progressive matrices, mental addition of three 3-digit numbers, and anagrams. For each task 55 items were selected to yield an average percent correct of 75%. The eye movements of subjects were recorded on every trial.

In this thorough study, the only reliable over-all effect of verbalization was that solutions took longer with concurrent verbalizing than in the silent condition. Although the average solution times with verbalization were longer for all four tasks, a post-hoc analysis showed that the

differences were statistically significant only for gambles and anagrams, and the differences were also small relative to the variability of the solution times. For gambles, the solutions with TA took 7.5 seconds longer than the average of solutions (44.6 seconds) under the silent condition, an increase of about 17 per cent.

The analysis of accuracy of solution showed that with concurrent verbalization subjects performed the gambles task a little better than silent subjects ($p < .01$), while in the addition task concurrent verbalizers performed slightly more poorly than silent subjects ($p = .06$). Hence concurrent verbalizing leads subjects to take longer and perform better on the gambles task, and to take longer but perform more poorly on the additions task.

To help us pinpoint the sources of differences, E. S. Johnson (personal communication) reports results from a preliminary analysis of the data. "We did not find any differences between conditions for the first ten practice trials, but control subjects subsequently increased their accuracy in the addition task more rapidly than subjects giving verbal reports. Both concurrent and retrospective subjects improved over trials in their selection of gambles, while control subjects gradually speeded up but maintained constant accuracy levels."

While awaiting the complete analysis of the protocols, eye-movement sequences, and errors, we can only speculate about the reasons for the differences in final performance. A comparison of the verbalizations during practice trials and test trials should reveal the changes in processing that took place in the gambling task. It is noteworthy that retrospectively reporting subjects, who did not have to think aloud, improved in performance in that task just as concurrently reporting subjects did. The difference for mental additions is especially surprising in view of the fact that earlier studies showed no effects of verbalizing in mental multiplication. From a preliminary analysis of numbers of protocol statements (Russo et al., in preparation) it appears possible that concurrent subjects failed to speed up because they continued to verbalize simple operations that could be elicited by recognition. This "superfluous" verbalization could slow down performance and interfere with recall of intermediate steps. Verbal reporting might not alter initial performance, but might slow down changes toward automaticity during continuing practice.

In sum, Russo et al. (in preparation) found reliable effects of concurrent verbalization for gambles and additions, but the differences reported are small in relation to other variability in the performances.

Since the effect was virtually the same for the concurrent and retrospective verbalization conditions, and since there was no effect of verbalization in the early trials, we are rather uncertain as to the real causes for the differences that were observed.

A recent dissertation by Gerhard Deffner (1983) provides additional extensive information about the effects of concurrent verbalization. Using anagrams, a geometric puzzle, and the N-term series problem (see Ohlsson (1980) for a description), and varying task difficulty and amount of perceptually available information, he found that thinking aloud increased solution times, but did not affect average numbers of solutions, and did not interact with the other experimental variables. Deffner's analysis of his subjects' strategies, and rate and content of their verbalizations will be discussed elsewhere in this monograph.

SUMMARY

The picture that emerges from our review of empirical studies of verbalization is quite clear and consistent. We have seen that the effects, or absence of effects, of verbalization depend upon characteristics of the tasks and of the verbalization instructions in ways that can be predicted from our model of the processing system.

When the instructional procedures conformed to our notion of Level 1 or Level 2 verbalization, the studies gave no evidence that verbalization changes the course or structure of the thought processes. A small number of minor differences between verbalizing and silent subjects can most plausibly be attributed to procedural differences between the experiments conducted with the two groups.

Studies that deviated in task, instructions, or experimental procedure, from the criteria of Level 2 verbalization were mainly experiments on perceptual-motor tasks and tasks requiring visual encoding processes, and experiments where specific verbal content was asked for by the instructions. Evidence was presented, for these kinds of experiments, that the information requested by the experimenter was not normally available to the subject in an oral encoding, and hence that producing the verbalization required intermediate processing that changed the course and structure of the thought process.

From our review of the evidence, we conclude that the processes subjects use to verbalize while thinking are neither illusory nor elusive, but can be understood and modeled. The processes associated with ver-

balization should be treated as an integral part of any model of the cognitive processes for a given task whenever the articulation takes the form of direct verbalization (i.e., vocalization of heeded information). The model should also include the processes for storing information in LTM, to account for the phenomena of retrospective verbalization at the end of experimental trials. The gross model we have proposed is focused on the verbalization of ongoing cognitive processes, but the postulated close link between information attended to and information stored should make it a relatively straightforward matter to model retrospective verbalization.

Finally, in the review of studies comparing different instructions to verbalize, we found substantial evidence that differences in performance were induced by telling the subject *how* to verbalize. In order to verbalize the information called for by the instructions, instead of the information he would normally have attended to, he had to change his thought processes.