Psychological Review

VOLUME 87 NUMBER 3 MAY 1980

Verbal Reports as Data

K. Anders Ericsson and Herbert A. Simon
Carnegie-Mellon University

The central proposal of this article is that verbal reports are data. Accounting for verbal reports, as for other kinds of data, requires explication of the mechanisms by which the reports are generated, and the ways in which they are sensitive to experimental factors (instructions, tasks, etc.). Within the theoretical framework of human information processing, we discuss different types of processes underlying verbalization and present a model of how subjects, in response to an instruction to think aloud, verbalize information that they are attending to in short-term memory (STM). Verbalizing information is shown to affect cognitive processes only if the instructions require verbalization of information that would not otherwise be attended to. From an analysis of what would be in STM at the time of report, the model predicts what can reliably be reported. The inaccurate reports found by other research are shown to result from requesting information that was never directly heeded, thus forcing subjects to infer rather than remember their mental processes.

After a long period of time during which stimulus–response relations were at the focus of attention, research in psychology is now seeking to understand in detail the mechanisms and internal structure of cognitive processes that produce these relations. In the limiting case, we would like to have process models so explicit that they could actually produce the predicted behavior from the information in the stimulus.

This research was supported by Research Grant MH-07722 from the National Institute of Mental Health.

We would like to thank Bill Chase and Robert Neches for their valuable comments on earlier versions of this article.

Requests for reprints should be sent to K. Anders Ericsson, Department of Psychology, Carnegie-Mellon University, Schenley Park, Pittsburgh, Pennsylvania 15213 (before September 1), or Department of Psychology, Muenzinger Building, University of Colorado, Boulder, Colorado 80309 (after September 1).

This concern for understanding the course of cognitive processes has revived interest in finding means to increase the temporal density of observations of behavior to reveal in greater detail intermediate stages of the processes. Increasingly, investigators record the direction of the subject's gaze (eye movements) and the intermediate behavior (making moves or other physical manipulations of stimulus material) that precedes the solution or criterion performance. Since data on intermediate processing are costly to gather and analyze, it is important to carefully consider how such data can be interpreted validly and what contribution they can make to our understanding of the phenomena under study.

Doubts About Verbal Data

One method frequently used to gain information about the course and mechanisms of cognitive processes is to probe verbally the

 0033-295X/80/8703-0215$00.75

subject's internal states. However, since the triumph of behaviorism over introspectively oriented competing viewpoints, verbal reports have been suspect as data. More precisely, behaviorism and allied schools of thought have been schizophrenic about the status of verbalizations as data. On the one hand, verbal responses (or keypunches that are psychologically indistinguishable from verbal responses, except that they are made with the finger instead of the mouth) provide the basic behavioral data in standard experimental paradigms. In the concept attainment experiment, the subjects say (or signal) *yes* or *no* when a possible instance is presented to them. In a problem-solving experiment, they report the answer when they find it. In a rote verbal learning experiment, they say "DAX" when the stimulus syllable "CEF" is presented. The actual performance measures commonly used—latencies and numbers of items correct—are derived from these responses, and the former depend for their validity on the veridicality of the latter.

On the other hand, modern psychology has been vague about the use that can be made of verbalizations produced by the subject along the route to solution or final response. Even more dubious is the status of subject responses to experimenter probes or retrospective answers to experimenter questions about prior behavior. All of these sorts of verbal behavior are frequently dismissed as variants of the discredited process of introspection (Nisbett & Wilson, 1977). Introspection, it has been generally agreed, may be useful for the discovery of psychological processes; it is worthless for verification. As Lashley (1923) said in a vigorous and widely cited attack on the method, "introspection may make the preliminary survey, but it must be followed by the chain and transit of objective measurement" (p. 352).

Unsatisfactory Methodology for Verbal Reporting

The notion that verbal reports provide possibly interesting but only informal information to be verified by other data has had a significant effect on the ways in which verbalizations are collected and analyzed. If the purpose in obtaining verbal reports is mainly to generate hypotheses and ideas, investigators need not concern themselves (and generally have not concerned themselves) with methodological questions about how such data are to be collected. As a result, there is little published literature on such methodological issues; the data-gathering and data-analysis methods actually used vary tremendously, and the details of these methods are sketchily reported in research publications that make use of such data.

This state of affairs is wholly unsatisfactory if we are to make rapid and continuing progress in understanding human cognitive processes. First, no clear guidelines are provided to distinguish illegitimate "introspection" from the numerous forms of verbal output (see the earlier examples) that are routinely treated as hard data, such as passing the chain-and-transit test. On what theoretical or practical grounds do we distinguish between a subject's yes or no in a concept attainment experiment and the assertions that the hypothesis being entertained is "small yellow circle"? Second, no distinction is made between such diverse forms of verbalization as "thinking aloud" protocols, retrospective responses to specific probes, and the classical introspective reports of trained observers. All are jointly and loosely condemned as "introspection."

Evolving a Methodology for Verbal Reporting

To end this confusion, we must extend our analyses of the tasks that our subjects are performing to incorporate the processes they are using to produce their verbal responses. The expansion of theories to include a theory of the measuring instruments is commonplace in physics. Experiments that involve weighing objects require at least a rudimentary theory of the pan balance. In the same way, experiments that record verbal responses of any kind need at least a rudimentary theory of how subjects produce such responses—in what memories the response information has been stored, what demands the response makes on short-term memory, whether responses can go on in parallel with other behaviors, and so on. Nor does this require-

ment of a theory of the response mechanism involve us in a vicious circle. Such a theory must be developed and tested simultaneously with our theories of task performance. In fact, such a theory, correct or incorrect, is implicit when we treat the subjects' verbal responses as veridical in standard laboratory paradigms.

In this article we reexamine the validity of verbal reports as data. We propose some means for moving from informal analysis of verbalized information toward objective procedures for collecting and analyzing them that would satisfy Lashley's (1923) chain-and-transit test. A main goal is to demonstrate that results from studies that are often cited against the use of verbalized information can be understood in terms of the methods used to collect and analyze the verbalizations.

Plan and Scope

With the advent of human information-processing theory, detailed models of memory and problem solving (e.g., Anderson & Bower, 1973; Newell & Simon, 1972; Simon, 1979) have been put forward that demonstrate the sensitivity of behavior to task instructions, types of stimuli, and other crucial factors in the experimental design. In this article we use the theoretical framework of human information-processing theory to propose a model for the verbalization processes of subjects instructed to think aloud, to give retrospective verbal reports, or to produce other kinds of verbalizations in response to experimenters' instructions. We will then use the model to analyze the thinking aloud method and other procedures using verbalization in the light of the criticisms that have been made of them.

We will conceive of the recorded verbalizations as data—exactly like latencies, eye fixations, sequences of moves, and so on—to be accounted for by a corresponding model, which generates them literally or on the level of encoded patterns or information content. This means that we will not assume that the verbalized description accurately reflects the internal structure of processes or of heeded [1] information, or that it has any privileged status as a direct observation; models that can regenerate the verbalizations (or encoded aspects of them) can be constructed and evaluated without such assumptions (Anderson, Note 1).

Types of Verbalizing Procedures

The only common feature among the whole range of techniques used to obtain verbal data is that the subject responds orally to an instruction or probe. Because of the flexibility of language, there are virtually no limits to the probes we can insert and the questions we can ask subjects that will elicit some kind of verbal response.

We propose a model of the cognitive processes that generate subjects' verbal responses. This model should be seen as a hypothesis about cognition on par with other hypotheses about cognitive processes. An example will make this point clear. Suppose that subjects in a problem-solving experiment are asked whether they used subgoals to solve the problem or solved it directly. If they assert that they used subgoals, this would hardly be conclusive evidence that they did, for it is easy to propose models of their cognitive processes that would permit them to generate this answer without consulting memory traces of the solution process to search for one or more subgoals among them. On the other hand, if a subject, in reply, at once described one or more specific subgoals, and these were both relevant to the problem and consistent with other evidence of the solution process, then it would be more difficult to construct a model of the cognitive processes that would produce this information without hypothesizing that it was stored in, and accessible from, the subject's memory of the steps taken in solving the problem.

The analogy to performance behavior, whose veridicality is commonly accepted without question, is clear. It would never cross a researcher's mind simply to ask sub-

[1] Because the phrase "attended to" is often awkward stylistically, we will sometimes use "heeded" instead. So, we will say, more or less synonymously, that information was "attended to," was "heeded," or was "stored in STM."

jects to raise their hands when they had solved a problem, without reporting the solution, no matter how much the experimenter thought he or she could trust them. The best evidence that they have actually reached a solution is their ability to report it. Conversely, it is hard to imagine a model of the cognitive processes that could report the solution unless it had actually been found. The relation of report to outcome of the process is so obvious that it never occurs to us to question the inference about process that we are drawing from the behavior. The procedures we propose in this article to infer internal processes from a wider range of overt verbal behaviors are simply an extension of the procedures that we already use every day in the laboratory. In both cases they involve building and testing alternative models of the cognitive processes that are going on, using standard paradigms for generating and testing hypotheses.

We begin with a classification of different species of verbalization. Producing verbalizations may be the subject's primary task, or only incidental to the "real" task that he or she is addressing. The verbalization may either be concurrent with task performance or retrospective. Various kinds of intermediate processes may intervene between the internal representation of information and its verbalization. The subject may report about specific events or may be asked to make generalizations. These and other variations in the circumstances under which verbalization takes place can have a significant effect on what is verbalized and on the interpretation of the verbal data.

Relative Primacy of the Two Tasks

When subjects verbalize concurrently, they generally must do two things, namely, perform the task that is being studied and produce the verbalizations. In the extreme case in which the verbalization information is totally unrelated to the main task and the purpose is to study the interference between the two (Peterson, 1969), the tasks and their respective heeded information may be viewed as entirely separate and distinct. (See Panel A of Figure 1.)

However, in the situations of primary importance to us, the two tasks are highly interrelated. In favorable situations, the additional cognitive load imposed by the instruction to verbalize may be negligible. In our subsequent review of experimental results, we will come back to the paradigm of dual tasks in order to discuss the interference generated by the added verbalizing task. Note that even in the case of retrospective verbalization, the subject's performance may depend heavily on how much incidental memorizing he or she does while performing the initial task.

We will mainly consider situations in which the verbalizing is supposed to be subordinate to, and passively dependent on, the ongoing cognitive process (see Panel B of Figure 1), as it only involves verbalization of heeded information already generated by the task-directed processes. However, we also discuss situations in which the verbalization is primary and must follow requirements of form and content imposed, for example, by instructions (see Panel C of Figure 1).

Concurrent and Retrospective Verbalization

One of the primary distinctions made in human information-processing models is between different types of storage systems used to retain information (Atkinson & Shiffrin, 1968; Simon, 1979, chap. 2.3). In whatever way the differences among memories may be conceptualized, what is remembered, and how well, will generally depend critically on the interval between the moment of acquisition and the moment of recall. This interval is an important consideration in classifying verbalization procedures.

If information is verbalized at the time the subject is attending to it, we will label the procedure *concurrent verbalization.* If a subject is asked about cognitive processes that occurred at an earlier point in time, we will label the procedure *retrospective verbalization.*

Recoding Before Verbalization

Various kinds of processes, and especially recoding processes, may intervene between

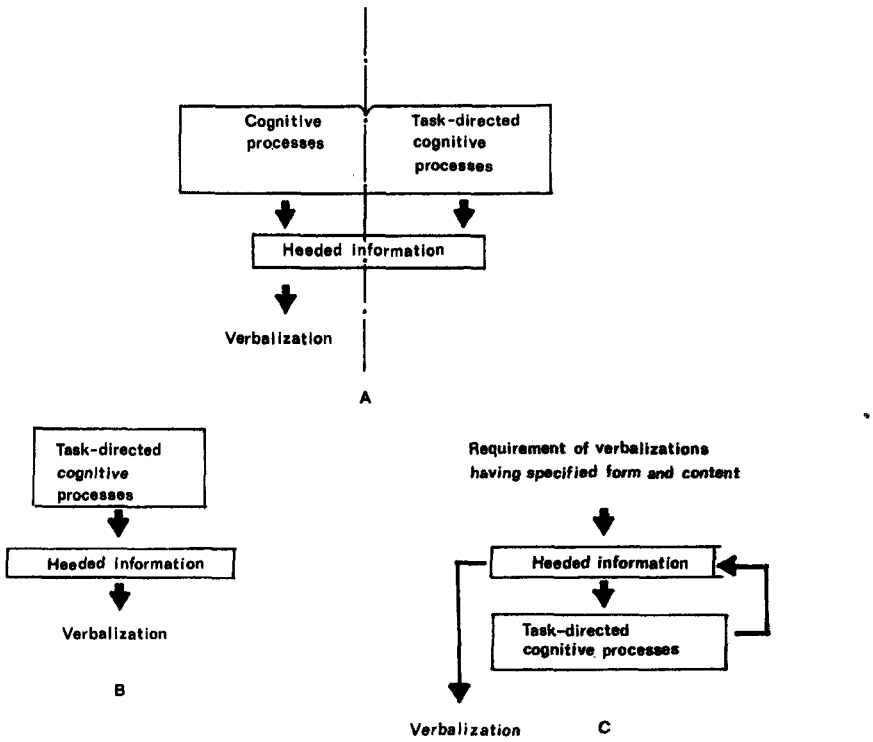

Figure 1. Three possible relations between task-directed cognitive processes and verbalization. ([Panel A] Verbalization and task-directed processes are separate and distinct. [Panel B] Verbalization of heeded information is generated by task-directed processes. [Panel C] Requirements for verbalization modify task-directed processes.)

the time the information was heeded by the central processor (CP) and the time a corresponding verbalization is generated. When information is reproduced in the form in which it was acquired from the central processor, we will speak of *direct* or *Level 1* verbalization. When one or more mediating processes occurs between attention to the information and its delivery, we will speak of *encoded Level 2* or *Level 3* verbalization. A number of different kinds of intermediate processes exist between access and verbalization that modify the information that is heeded. Among the important kinds are the following.

1. Intermediate recoding into verbal code (Level 2 verbalization). This occurs when the internal representation in which the information is originally encoded is not in verbal code but has to be translated into that form. Werner and Kaplan (1963) have shown that when subjects generate verbalizations or verbal descriptions of nonverbal stimuli for their own future use, the format is compact and incorporates many idiosyncratic referents. When verbalizations are generated to communicate the information to another person, additional processing is required to find understandable referents (Werner & Kaplan, 1963).

2. Intermediate scanning or filtering processes (Level 3 verbalization). When the task instructions ask for verbalization of only a selected type of attended content, it is necessary to postulate additional processes that test recurrently if the heeded information matches the desired type. A typical example of such instructions occurs in commentary driving experiments in which the subjects are asked to report all perceived traffic hazards while they are driving a car (Soliday & Allen, 1972).

3. Intermediate inference or generative processes (Level 3 verbalization). The situation

is even more complicated if the experimenter is interested in particular aspects of the situation that a subject would not ordinarily verbalize or attend to. The issue of whether the instruction to verbalize calls for information that normally is not heeded by the subjects in performing the cognitive activity under study is central and directly related to the occurrence of intermediate inference and generative processes. Since we will return to this issue in more depth, only a brief summary will be given here of the types of information that are likely to require additional mediating processing for their generation.

In addition to verbalizing their ongoing thinking, subjects are sometimes asked for verbal descriptions of their motor activities, for example, what objects are moved where, or where they are looking. When this information is not directly heeded, as is often the case, the subject is required to observe his or her own internal processes or overt behavior to generate the information.

Experimenters are often interested in the subjects' reasons for their overt behavior and consequently ask the subjects to verbalize their motives and reasons, which in the normal case for many activities may not be available directly or even at all.

Similarly, in studies that use retrospective verbalization, subjects are seldom asked what they can remember about specific instances of their cognitive processes. Rather, they are generally asked to retrospect about their thought processes in experiments with many trials or to answer general questions, and thus must try to synthesize all the available information after selective recall.

The intermediate processes investigated by Tversky and Kahneman (1973) for judgments on frequency and the probability of events fall in this general category. Events that were recalled readily were judged by subjects to be representative and frequent, but this led to large estimating errors, for frequency and representativeness are not the only determiners of availability for recall. Similarly, Nisbett and Wilson (1977) have shown that subjects verbalizing retrospectively in a variety of settings about the motives for their behavior were no more accurate than observers were in identifying the important situational factors that actually determined the behavior.

Forms of Probing

One of the most direct and widely used methods to gain information about subjects' internal states is to instruct them to think aloud or talk aloud. With this procedure, the heeded information may be verbalized either through direct articulation or by verbal encoding of information that was originally stored in a nonverbal code. With the instruction to verbalize, a direct trace is obtained of the heeded information, and hence, an indirect one of the internal stages of the cognitive process.

In a related procedure, the subjects are probed, concurrently with their performance of a task, for specific information, usually of a kind that they presumably need to guide their succeeding behavior. Typical examples of concurrent probing are requests to subjects to report the hypotheses they are using in concept learning and discrimination learning.

A third class of verbalization procedures, which we have called retrospective verbalization, probes the subject for information after the completion of the task-induced processes. For example, subjects may be asked to report just after the process has been completed. Another form of retrospective probing is a method, which we call *interpretive probing,* in which subjects are probed at the completion of an experimental session consisting of a large number of different trials. This procedure is sometimes justified as eliminating any possibility that the probing will affect the "real" data of the experiment.

Particular and General Reports

If the purpose of retrospective probing were to recover memory traces of subjects' processes, then the appropriate instruction would be to ask them to recall their specific thought processes during particular trials of the experiment. For at least two different reasons, such a procedure is rarely used. First, after a series of trials, a subject's memory for internal states of individual cognitive pro-

cesses will be poor and lacking in detail. Moreover, there is a tendency over time for recurrent cognitive processes to become automated, so that the accessible intermediate states of the processes for the later trials of the experiment become few or nonexistent.

Second, many experimenters are interested primarily in general characteristics of the thought processes, and not in the episodic details of the individual trials. Such experimenters probe their subjects with questions such as, "How did you do these tasks?" Such questions implicitly or explicitly request a general, rather than specific, interpretation of how the subject was performing the tasks in question.

There are several different ways in which subjects might arrive at descriptions of their general procedures, as distinct from reports on specific behaviors during individual trials. One possibility is that the subjects are aware of the general procedures, or "programs," they are using, use essentially the same programs on all trials, and can recall and report these directly, without reference to the specific behavior they produced. Another possibility is that subjects can remember some parts, or even complete episodes, of their processes during particular trials, and that they attempt to generalize this information into a general procedure, which they then report. A different possibility is that subjects remember some specific tasks, regenerate (by redoing them) the processes used for these tasks, and use this information to infer the general procedures they may have used. Finally, the subjects may be drawing on a variety of kinds of prior information, such as general knowledge on how one ought to do these tasks, to generate a verbal report describing a general procedure or strategy. In this case, the verbal reports may not bear any close relation to the actual cognitive processes used in the tasks (Nisbett & Wilson, 1977).

It is interesting to note that in areas of applied psychological research in which verbal questioning of subjects has a long tradition, subjects are usually asked about specific events rather than for general information or conclusions. In the critical incident technique proposed by Flanagan (1954), the subjects were always asked to report their memory for specific events, for example,

> pilots returning from combat were asked "to think of some occasion during combat flying in which you personally experienced feelings of acute disorientation or strong vertigo." They were then asked to describe what they "saw, heard, or felt that brought on the experience." (Flanagan, 1954, p. 329)

From these considerations, we can see that interpretive probing, unlike the critical incident technique, cannot be relied on to produce data stemming directly from the subjects' actual sequences of thought processes. The former procedures encourage or even require subjects to speculate and theorize about their processes, rather than leaving the theory-building part of the enterprise to the experimenter. There is no reason to suppose that the subjects themselves will or can be aware of the limitations of the data they are providing. Moreover, the variety of inference and memory processes that might be involved in producing the reports make them extremely difficult to interpret or to use as behavioral data.

In some studies, subjects are even asked how they would behave if the conditions of the experiment were altered in some way. We will refer to this procedure as probing for *hypothetical* states. For example, in a study by Reed and Johnsen (1977), subjects were asked how they would solve a problem if it were presented to them again. Subjects in a study of Nisbett and Wilson (1977) were asked how they would react to a story if some passages had not been presented.

Finally, subjects may be asked questions that can be answered without reference to the context of the experiment—what we will call *probing for general states.* Many investigators (Watson, 1920) do not seem to distinguish between verbal reports given about a just preceding cognitive process, and verbal probing to elicit general procedures and methods outside the context of specific processes (e.g., how one hits a golf ball). The theoretical interpretation of these two modes of verbalization should be quite different.

Directed or Specialized Probing

Verbal probes differ not only along the dimensions of concurrency versus retrospection and the generality or particularity of the events that are to be reported, but also in the comprehensiveness of the topics that are to be reported. In many studies, the investigator is interested in only some particular aspect of subjects' behaviors. Then the verbal probe may be constructed to induce the subjects to generate information specifically relevant to the hypotheses under consideration. To help subjects retrieve the desired information from memory, and to induce greater completeness of the verbal reports, the question or verbal probe often contains contextual information. To guard against subjectivity in analyzing verbal reports, the investigator often supplies subjects with a fixed set of alternative responses. In contrast, a general instruction to give verbal reports typically asks subjects to tell everything they can remember or are thinking of while performing the task.

In most cases, verbosity and absence of selectivity in subjects' reports is not an important problem. What the subject reports is likely to be less, rather than more, than we should like to hear. In no study known to us using general instructions has the investigator complained that subjects have reported too much information from actual memory.

One common difficulty in probing for specific information, especially when the subjects are offered a fixed set of alternative answers, is to know that the questions conform to the internal representations that the subjects are employing in their thought. Probes for types of information that subjects do not have directly accessible, or probes that provide inadequate sets of alternatives, may force subjects to intermediate and inferential processing, and hence produce verbal reports that are not closely related to the actual thought process. Moreover, when specific, fixed-alternative probes are used, there is no way to detect from subjects' responses that this has occurred.

Providing contextual information and prompts to subjects may aid recall from long-term memory, and in studies of long-term memory the use of prompts and context is frequent and relatively well motivated. When subjects are asked to report on immediately preceding cognitive processes of relatively short duration, specific probes are more questionable and less useful. In a logical sense, the experimenter gets just as much information from the subject in the third as in the first two of the following three cases.

(1) Directed probe 1
Question: Did you use X as a subgoal?
Answer: Yes
(2) Directed probe 2
Question: Did you use any subgoals? If so, which?
Answer: Yes, I used X.
(3) Undirected probe
Verbal report: . . . I was first trying to get X and I . . . when I attained X . . .

The replies in all three cases provide evidence that the subject used X as a subgoal, yet the evidence is stronger in the third case than in the second, and in the second than in the first. The verbalization of Case 1 could easily be generated by processes independent of any memory for the actual thought processes. Comparing Cases 2 and 3, the former communicates to subjects the information that the experimenter expects them to report. It may encourage subjects to try to infer or guess what kind of information the experimenter will accept, and to generate information accordingly.

In many cases, other criteria are available for estimating the validity of the reports. An analysis of the task (Newell & Simon, 1972) will often provide strong indications of the adequacy of verbalized information, especially in cases with a large number of logical possibilities of response.

Finally, different kinds of probes may have different effects on the subsequent behavior of subjects. The request for a certain type of information may serve as a hint to subjects about what aspects of the task are important. Subjects may also alter their normal mode of processing in order to be able to give the requested information to the experimenter on subsequent trials.

The Processing Model

Our purpose in presenting a specific processing model is to aid us in interpreting

verbal data obtained from subjects and the relation of their verbal to their other behavior. Since the data (including the verbal data) are gathered to test theories about the human information-processing system, we are engaged in something of a bootstrap operation. We need a model to interpret data that are to be used to test the model.

Under these circumstances our data-interpretation model should be as simple as possible, and it must not incorporate components that are themselves bones of theoretical contention. The model should be robust, that is, compatible with a wide range of alternative assumptions about human information processing.

The specifications we are about to present are simple and robust in this sense, and, indeed, summarize the core that is common to most current information-processing theories of cognition. Of course, they are not entirely neutral, for they would be hard to reconcile with an extreme version of behaviorism that denied the relevance of central processes to the explanation of behavior. But they do not represent the view of any particular "sect" within the general information-processing tradition.

The most general and weakest hypothesis we require is that human cognition is information processing: that a cognitive process can be seen as a sequence of internal states successively transformed by a series of information processes. An important and more specific assumption is that information is stored in several memories having different capacities and accessing characteristics: several sensory stores of short duration, a short-term memory (STM) with limited capacity and/or intermediate duration, and a long-term memory (LTM) with large capacity and relatively permanent storage, but with slow fixation and access times compared with the other memories.

Within the framework of this information-processing model, it is assumed that information recently acquired (attended to) by the central processor is kept in STM and is directly accessible for further processing (e.g., for producing verbal reports), whereas information from LTM must first be retrieved (transfered to STM) before it can be reported.

This general picture is compatible with all sorts of specific hypotheses that have been put forth with respect to the details of the mechanisms. For example, some theorists propose that what we call "short-term memory" is not a separate, specialized store but simply a portion of LTM that is currently and temporarily activated. Some theorists believe that information in STM extinguishes with passage of time, unless rehearsed; others believe that it is lost only when replaced. In general, these differences of detail do not affect the model at the level of specificity required for our purposes. The important hypothesis for us is that due to the limited capacity of STM, only the most recently heeded information is accessible directly. However, a portion of the contents of STM are fixated in LTM before being lost from STM, and this portion can, at later points in time, sometimes be retrieved from LTM.

We assume that any verbalization or verbal report of the cognitive process would have to be based on a subset of the information in these memories. From this and the earlier mentioned hypotheses, the taxonomy of verbalization procedures shown in Table 1 follows in a straightforward fashion. The taxonomy provides us with a theoretical foundation for some of the distinctions we have already made in types of verbalization.

The two dimensions of Table 1 represent two major distinctions. First, the time of verbalization is important in determining from what type of memory the information is likely to be drawn. Second, we make a distinction between procedures in which the verbalization is a direct articulation or explication of the stored information and procedures in which the stored information is input to intermediate processes, such as abstraction and inference, and the verbalization is a product of this intermediate processing.

Detailed Specification

We must now specify more fully the components, which we have just sketched, of the information-processing system that carry out the processes of verbalization. The model

Table 1
A Classification of Different Types of Verbalization Procedures as a Function of Time of Verbalization (Rows) and the Mapping From Heeded to Verbalized Information (Columns)

	Relation between heeded and verbalized information			
		Intermediate processing		
Time of verbalization	Direct one to one	Many to one	Unclear	No relation
While information is attended	Talk aloud Think aloud			
While information is still in short-term memory	Concurrent probing	Intermediate inference and generative processes		
After the completion of the task-directed processes	Retrospective probing	Requests for general reports	Probing hypothetical states	Probing general states

draws on a variety of sources that are summarized in Newell and Simon (1972, chap. 14) and Simon (1979, chap. 2.3).

Few of the model's specifications are controversial. It makes no real difference, for example, whether we assume a single homogeneous memory with different modes of activation (e.g., Anderson, 1976; Shiffrin & Schneider, 1977) or several discrete memory stores (sensory stores, STM, and LTM). The important matters, which can be described in either terms, relate to the amounts and kinds of information that can be retained and the conditions for accessing them and reporting them verbally. We will use the conventional model of multiple memories in our description.

Recognition. Information received from the sensory organs resides for a short time in memories (iconic and echoic) associated with the different senses. During this time, portions of the sensory information are directly *recognized* and encoded with the aid of information already stored in LTM. Recognition associates the stimulus, or some part of it, with existing patterns in LTM and stores in STM "pointers" to those familiar patterns. Intermediate stages of the direct recognition process, which may take only 50–100 msec, do not use STM to store their products.

Long-term memory. The LTM may be pictured as an enormous collection of interrelated nodes. Nodes can be accessed either by recognition, as just explained, or by way of links that associate these nodes to others that have already been activated. Information accessed by association is then also represented by pointers in STM. Thus, information can be brought into STM from sensory stimuli via the recognition process, or from LTM via the association process. Association processes are much slower than direct recognition processes, requiring at least several hundred milliseconds for each associative step. Associative processes may use STM to store intermediate steps. So, for example, in recalling a name that is not immediately accessible, a person may use a sequence of cues to find an associative path, step by step, to the sought-for name. Such processes may last tens of seconds, or even minutes, and may leave numerous intermediate symbols in STM, where they are temporarily available for verbal reports.

Short-term memory. The CP, which controls and regulates the nonautomatic cognitive processes, determines what small part of the information in sensory stimuli and LTM finds its way into STM. This is the information that is *heeded* or *attended to.* The amount of information that can reside in STM at one time is limited to a small number (four?) of familiar patterns (chunks). Each chunk is represented by one symbol or pointer to information in LTM (Simon, 1979, chap. 2.2). As new information is heeded, information previously stored in STM may be lost.

When a cognitive task (e.g., mental addi-

tion of a column of figures) is being carried out, the typical chunks in STM are pointers to the operands, operators, and outputs of the operations that are being performed. Thus, in adding 3 to 4, pointers corresponding to the symbols "3," "4," "PLUS," and "7" might at some time be present in STM. Since, in our culture, adding two digits involves a direct reference to LTM ("table look up"), no further details of the process would be heeded in STM or available for verbal reports. On the other hand, if the task were to multiply 17 by 45, STM might hold, at various points in the process "45," "17," "7," "TIMES," "3" (the carry in multiplying 45 by 7), "315" (the first intermediate product), "45," "1," "TIMES," "PLUS," "765."

We hold no brief for the details of the earlier description, which is intended merely as an example of the kinds of information we would expect to be heeded in STM, and to be available, potentially, for concurrent or retrospective reports. The specific details would depend on the particular strategies subjects used and the nature of the chunks stored in LTM (Simon, 1979, chap. 2.4). STM would symbolize the process only down to some modest level of detail (corresponding to elementary processes of a sec or 2 in duration), and we would not expect to find information there about simple, automated processes (e.g., the processes of retrieval from LTM or recognition processes), much less about neuronal events. Thus, the architecture of the control apparatus (CP) determines the fineness of grain of the representation of processes in STM.

Control of attention. The flow of attention is diverted, from time to time, by interruptions through the higher control mechanism. Intermediate stages in these interruptions, not being symbolized in STM, are not reportable. Sudden movements in peripheral vision, loud noises, and emotions operating through the reticular system are important causes of interruption and shift in attention (Simon, 1979, chap. 1.3). Even though information heeded immediately before or after a shift in attention may sometimes allow subjects to give a relatively clear account of the interruption, we would expect such information to be less complete than reports of an orderly process that is induced by the successive contents of STM itself (e.g., a thought sequence during which goals in STM are guiding the thought processes).

Fixation. New information is retained in STM during the time the CP is attending to it. To create an LTM representation of new information that can later be recalled, associations must be built up by coding and imaging, as well as new tests and branches in the recognition network. Processing of the order of 8–10 sec is required to assemble each new chunk from its familiar components in STM, and to store it in LTM as a new chunk (Simon, 1979, chap. 2.2, 2.3).

Automation. As particular processes become highly practiced, they become more and more fully automated (Shiffrin & Schneider, 1977). Automation means that intermediate steps are carried out without being interpreted, and without their inputs and outputs using STM. The automation of performance is therefore analogous to executing a computer algorithm in compiled instead of interpretive mode. Automation and compiling have two important consequences. They greatly speed up the process (typically, by an order of magnitude), and they make the intermediate products unavailable to STM, hence unavailable also for verbal reports.

Verbalization Processes

Within the context of this general model, verbalization processes produce (externalize) information that is in STM. In the case of thinking-aloud instructions, the information verbalized will then be some portion of the information currently being attended to. The verbal production process takes two rather different forms, depending on whether the STM chunks already denote symbols in the verbal mode. For information that can be represented as a string of phonemes, that is, aural information, the model assumes that attending to that information or activating the corresponding structure in LTM allows the information to be vocalized by automatic verbal translation without making additional demands on STM or the CP.

Intermediate processes. When information

in STM is not verbally encoded (e.g., visual imagery), making a verbal report requires, according to the model, the corresponding verbal representations of the information (in the simplest case, names or labels) to be evoked. The recoding processes will make at least modest demands on processing capacity and processing time. This means that some heeded information may not be vocalized when other task-directed processes take priority and interrupt the verbal encoding and production processes.

The verbal-encoding processes involved in thinking aloud evoke a verbal reference to, and occasionally an explication of, the heeded structure in STM. These processes would not be predicted to change the information attended to in the way that requested explanations would. (When subjects are asked for explanations, the verbal reports cannot be generated without extending the information and relations heeded.) Hence, thinking aloud, as distinguished from explanation, will not change the structure and course of the task processes, although it may slightly decrease the speed of task performance.

Retrospective reports. The most general retrospective verbalizing instruction asks the subject to report everything he or she can remember about the cognitive process studied. If the subject is asked immediately after performing the process, the model predicts that some previously heeded information will still be in STM, permitting direct reporting by the processes described earlier, and facilitating retrieval of additional information stored in LTM in episodic associations that were formed when the information was heeded.

The control process in retrieving previously heeded information from LTM, however, may be rather variable. Since it is outside the scope of this article to review the existing evidence on factors affecting the efficiency of retrieval, we will limit ourselves to a few comments. In situations in which similar information is attended to over and over, as in experiments with factorially designed stimulus material, the model would predict that retrieval of specific items will be hampered by extensive interference. Retrieval may not only fail to access previously heeded information but may on occasion access information that is confused with the events being queried, and hence is inaccurate. The degree to which retrospective verbalization must rely on retrieval from LTM can be minimized by studying cognitive processes of short duration, where the verbal responses lag the task processes by only a brief interval.

Empirical Evidence on Verbalization

In the remainder of this article, we will develop in more detail our model of verbalization and test it against empirical evidence. We will show that the empirical findings are compatible with the assumptions of our model, and that the model provides guidelines for the interpretation of verbal reports gathered under various procedures.

Since the relevant literature is voluminous, we cannot review it all explicitly within the compass of a journal article. Instead, we will select representative studies for discussion, leaving a more exhaustive survey to other papers (Ericsson & Simon, in press; Ericsson & Simon, Note 2, Note 3). However, in our process of selection we will be careful not to screen out studies that are troublesome for our theoretical framework.

The classical issues we will discuss in the next three sections are (a) the effects on the cognitive processes of the instruction to verbalize and of probes, (b) the completeness of verbal reports, and (c) the consistency of verbal reports with other empirical data on behavior. These are the central questions that must be settled if verbal reports are to be used as data in psychological research. In addition, in experiments in which verbal reports are obtained, it is important to raise two further questions: (d) the generalizability and validity of the verbalized information and (e) the design of objective methods for encoding and analyzing think-aloud protocols. We have addressed these latter questions in several working papers (Ericsson & Simon, Note 2, Note 3) and will not pursue them further here.

Effects of Probes and Instructions to Verbalize

It is often asserted that the mere procedure of eliciting verbal reports changes the course

and structure of the cognitive processes that are under study. If that were so, the verbal data, even if they reflected accurately the cognitive processes going on during verbalization (and hence provided significant and valid data about cognition), would give an inaccurate picture of the "normal" course of those processes.

In studies in which the probing and the verbal reporting take place at the end of the experiment, it may be argued that since the subjects cannot be aware of the fact that they are subsequently going to be asked to report on their processes, the reporting task cannot affect those processes. However, in studies in which the subjects are told explicitly in their initial instruction that they will be questioned about general (Morgan, 1934) or specific (Rommetveit, 1965) aspects of the experiment, the possibility that these instructions will affect their cognitive processes during the experiment cannot be ruled out on logical grounds. The possibility of induced effects is even greater when the probing procedure requires subjects to give reports periodically during the experimental sessions.

Nevertheless, the greatest concern about possible effects of verbalization on the course of the cognitive processes arises when the verbalization is concurrent with the task performance. For that reason we will concentrate, in this section, mainly on concurrent verbalization.

Predictions for Concurrent Verbalization

Our model of concurrent verbalizing assumes that the verbalizations involve either direct articulation of information stored in a language (verbal) code [2] (Level 1 verbalization); articulation or verbal recoding of nonpropositional information without additional processing (Level 2 verbalization); or articulation after scanning, filtering, inference, or generative processes have modified the information available (Level 3 verbalization).

When the subjects articulate information directly that is already available to them, the model predicts that thinking aloud will not change the course and structure of the cognitive processes. Nor will verbalization under these conditions slow down these processes.

When the information being processed in order to perform the main task is not verbal or propositional, the model predicts that the performance may be slowed down, and the verbalization may be incomplete but that the course and structure of the task-performance process will remain largely unchanged. Several kinds of tasks fall in this category.

The performance may be highly automated, hence may not make much use of STM. This case includes acts of recognizing familiar stimuli, and more generally, many kinds of tasks after long practice. For such tasks, it is most likely that the thinking-aloud protocols will be very sketchy, but that the processes will not be slowed down or altered. We would also expect a more frequent injection of metastatements (explicit statements about the process itself) replacing statements about inputs and outputs in the protocols.

Similar predictions can be made for tasks with a large motor–perceptual component and tasks employing complex visually encoded stimuli. If the task performance is not highly automated, then the subjects, in their endeavors to obey the thinking-aloud instructions, may take time to translate their inputs and outputs into verbal form, and to report them, but at the expense of slowing down their performance of the task.

In studies in which subjects are not merely asked to think aloud but are asked for specific kinds of information—for example, the reasons for their subsequent actions and moves (Gagné & Smith, 1962; Wilder & Harvey, 1971)—their efforts to obey the instructions would be predicted to have more substantial effects on task performance. The model would predict that these effects would be especially prominent if the subjects were asked to produce information that would not normally be available to them during their performance of the task. If the information requested was information that would normally be available, then the model would predict that the effects of the verbalization instructions would be relatively minimal. The

---

[2] Hereafter, we will refer to information stored in memory in essentially propositional form as "verbally encoded."

prediction depends, therefore, on the process used to perform the task; and conversely, the degree to which verbalization changes task performance can be used as a cue to determine what that process is.

In studies using probes, the same distinction can be made between probes requesting specific information to be verbalized (and occasionally specifying even the information's form and completeness) and more general and nondirective probes. The latter would affect the process less than the former. The effects on task performance would be especially large if the probes rested on a theory or conceptual framework that was not a veridical or adequate description of the information available directly to the subjects.

It is worth emphasizing again that the relation between main task and reporting task is mutual. The predictions of the model can be used to test hypotheses about the task processes, just as predictions from the latter can be used to test hypotheses about the model of verbalization. Once we have acquired some confidence in the verbalization model, we find that differences between subjects in thinking-aloud and silent conditions have implications for the processes that are being used to perform the main task.

In considering empirical studies that employ thinking-aloud procedures, we will be concerned primarily with studies that meet the criteria of Level 2 verbalization, for most of the experiments in the literature deemphasize 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 that 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 1 and 2 Verbalization

Of the experiments explicitly designed to study the effects of verbalizing, only a few have exposed the experimental and control groups to identical conditions. An important example is a study of discrimination learning by Karpf (1973), in which he 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 10 preliminary problems, and were then given 15 experimental problems for which the experimental group was asked to think aloud. Finally, 5 problems were given for which *all* subjects were instructed to be silent, to allow exploration of aftereffects of thinking aloud.

The stimuli in Karpf's (1973) experiment were slides, each containing a pair of letters varied along eight dimensions, such as form (A or U), size (large or small), color (black or white), shape of the border surrounding the letter (circle or square), and texture of line under the letter (solid or dotted). The letters in each pair were discriminated by simple hypotheses, involving a single dimension. Although it would be expected that subjects would process these stimuli in pictorial rather than verbal mode, it is fairly easy to recode them verbally; hence, the model would predict that thinking aloud would affect, at most, only the speed of the task performance.

This is, indeed, what Karpf reported. No reliable differences were found between the thinking-aloud group and the control group, for either the experimental problems or the final problems, in numbers of problems solved correctly. However, the thinking-aloud group took about 50% more time than the control. Supporting evidence comes from a study by Roth (1966), who found that verbalization had no effect on the effectiveness of task performance—but in his experiment also, there was no effect on speed of performance. (For similar findings, see Carroll & Payne, 1977; Feldman, 1959; Kazdin, 1976; Johnson & Russo, Note 4).

In a study of mental multiplication, Dansereau and Gregg (1966) asked a subject to verbalize each step during the solution process. Moreover, whenever the subject remained silent too long, the experimenter urged him to talk. On a wide range of problems varying in difficulty, no reliable differences in speed of performance were found between a silent control condition and the verbalizing condition. In a subsequent study, Dansereau (1969) reported that with increased practice, two of his faster subjects gave overt verbalizations only of intermediate results and the

initial problem, and they reported introspectively that complete verbalization interfered with retrieving information. To what extent these subjects felt that verbalizing slowed them down is not reported, nor is a comparison provided with performance under a silent control condition.

A number of other studies allow comparisons between a silent and a vocalizing group, for identical or similar tasks, but without complete comparability between experimental and control conditions. For example, for subjects discovering proofs in propositional logic, Newell and Simon (1972) compared the number of solutions attained and the detailed solution paths of their seven think-aloud subjects with the solutions (collected by different investigators at Yale) to the same two problems by 64 subjects under silent conditions. The data that could be compared between the two conditions were the actual steps taken while searching for the proof—the entire search tree, including both the correct paths and the unsuccessful attempts. When the detailed structures of the search trees were 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 (and found the correct solutions about as often). The stimuli here were symbolic expressions that are easily described in words.

Similar results were obtained by Ericsson (1975a, 1975b) in a study of problem solving with the 8-puzzle, a task in which small, numbered tiles must be manipulated into a desired arrangement. The task has a strong visual–perceptual component, but moves are easily encoded in symbolic form (e.g., "Move 7 up," "Move 3 left.") so that it is not difficult to verbalize them.

Two separate experiments were conducted for the thinking-aloud and silent conditions, but the same sequence of puzzles was presented to the subjects in both conditions. In the thinking-aloud experiment, the subjects also had to tell the experimenter which tiles to move; in the silent condition, the subjects, sitting alone, pressed keys on a teletype to cause the computer to make the moves. The sequence of moves made in the two conditions could be compared (Ericsson, 1975a, 1975b). Various subgoals can be defined for the task (e.g., arranging the tiles in the first row correctly). No differences were found in the subjects' attainments of such subgoals, nor in the structure of their search trees. However, for the first several problems, the verbalizing subjects made a larger average number of moves than did the silent subjects. Thus, it appears that there was some tendency for the silent subjects to do more planning and thinking ahead than the verbalizing subjects. Whether this difference was due to the verbalizing, to a feeling of "irreversibility" in typing the move to the computer, or to some other cause cannot be determined from the experiment.

This sample of studies comparing problem-solving behavior under thinking-aloud and silent conditions illustrates that the model predicts correctly that verbalization does not affect the behavioral manifestations of the thought processes when the conditions for Level 1 verbalization are satisfied; generally, it only affects the speed of performance when the stimuli are nonverbal but easily recoded for verbal reporting. Where other effects of verbalization are found (as in the Ericsson experiment), they can reasonably be attributed to other differences between the silent and thinking-aloud conditions. These findings suggest that the internal structure of the thought processes also is not changed as a result of the verbalizing activity.

Studies Not Meeting Level 2 Conditions

When the criteria for Level 2 verbalizing are not met, so that the subject is asked to verbalize information that would not be heeded in the normal course of processing or that could not easily be encoded in a verbal code, our model predicts that the course and structure of the cognitive processes may be changed by the verbalization.

Not only is it possible that verbalization will change the thought process, but conversely, it is likely that the nature of the task will cause the subjects to give a different, and generally less complete, account of their processes than under Level 2 conditions. We will postpone to the next main section of this

article the discussion of the completeness of verbalizations, and will be concerned here only with how the attempt to verbalize will affect task performance.

Verbalization of perceptual–motor processes. The problems of verbalizing perceptual–motor processes are most clearly visible in problem situations in which 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. Because verbalizations are often quite sketchy in tasks of these kinds, experimenters sometimes change the task to increase verbalization. For example, if the subjects must instruct the experimenter to make the manipulations for them, instead of making them themselves, they are thereby forced to form an internal representation of the moves that can be encoded verbally. When this is done, more of the content of the thought processes is, in fact, verbalized.

To increase verbalization of content, some experimenters using the thinking-aloud method change 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 encoded verbally. 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 in that the overt moves have now become covert and are no longer amenable to direct observation. The subjectively perceived polarity between thinking and manipulation, mentioned earlier, 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 trade-off between speed and quality of performance was noted by Shipstone (1960) in a concept-learning task, with an instruction to disregard speed and concentrate on what to do. Ray (1957) found that requiring subjects to tell the experimenter what they were going to do before they started to manipulate the switches significantly decreased the number of overt trials to solution (but the corresponding solution times are not reported).

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

Verbalization can be encouraged not only by constraining the manipulations of the problem material but also by instructing subjects to verbalize the motives or reasons for their actions. Our model predicts that such instructions will likely change the course of processing. In general, when the thinking-aloud instructions do not require verbalizations of motives or reasons, the protocols do not contain them. Three studies in which subjects were specifically instructed to make such reports illustrate the effects that may be expected.

The study by Gagné and Smith (1962) with the Tower of Hanoi problem was aimed at investigating the effects of different verbalization instructions on performance during some training tasks (two-disk to five-disk problems) and on 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 overt 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 time, taken up entirely with the act of verbalization" (Gagné & Smith, 1962, p. 17). 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 overt and covert verbalizing conditions but a clear reduction in the number of moves in both those conditions, as compared with a control condition. The time taken 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 (1962), Davis, Carey, Foxman, and Tarr (1968) included the presence of the experimenter as an additional dimension in a factorial design. In this study the subjects were instructed at the beginning of the experiment to verbalize their reasons, 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 et al. study, verbalization had no effect on the training task (five-disk problem), but the "think-aloud" subjects required significantly fewer moves on the test problem (six-disk problem), even though no subjects were asked to verbalize on the latter task. The experimenter's presence facilitated performances in both conditions on the training problem but not on the test problem. The hypothesis that the "think-aloud" instruction would influence performance by interacting with the presence or absence of the experimenter was not borne out. The extent to which these effects were mediated by differences in solution times cannot be discussed, since the published report gives no information about these times.

These three studies show that as predicted, a requirement to verbalize reasons and motives has substantial effects on both immediate performance and learning, and that generating verbalized reasons brings about changes, at least in manipulative tasks, in the 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 causes them to substitute unrecorded covert trials and planning for overt trials. The negative result of the Davis et al. (1968) 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 this explanation. Finally, we may conjecture that the richness of alternative strategies for 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 a verbal or symbolic representation or code when subjects are presented with drawings or pictures. An instruction to describe a visual scene verbally should require a verbal recoding of the picture, which will imply extensive processing. Our model predicts that this additional processing may have three kinds of effects: It may slow down performance of the main task, it may change the structure and course of performance of the main task, and it may influence what is remembered about the task and is later available to retrospective verbalization. The magnitude of the effects will depend on how fully the subjects carry out the instruction to ver-

balize, and on how difficult it is to describe the visual scene in words. We have already seen that when the latter is relatively easy, the only detectable effect of verbalization may be to slow task performance.

The task of viewing a novel visual scene has been studied to detect effects of imposing the additional task of describing the scene verbally. A study by Freund (described in Loftus & Bell, 1975) showed that subsequent recognition of scenes was much improved by the verbalizing requirement. On the other hand, when an unrelated verbal task, such as 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).

Evidence supporting the distinctness of the visual and verbal 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, whereas control subjects who traced the disappearance on an outline of the image did not.

A number of studies provide evidence of large individual differences in preferences for perceptual versus verbal processing (Ericsson & Simon, Note 3). Our model would predict that requiring subjects to verbalize explanations in a task with a complex visual stimulus would cause subjects with preference for perceptual processing to alter their strategies and hence their performances. For example, in a study by Brunk, Collister, Swift, and Stayton (1958), subjects were given an initial test of the Vygotsky 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" (p. 238). 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 model would predict.

In a series of studies reported in Merz (1969), the effect of verbalization on performance in intelligence tests was investigated. For example, Kesting, using the Figure Reasoning Test, reported that 13- to 17-year-old subjects who always had to say aloud how the figures were alike or different performed 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, whereas another group had to sing "la, la, . . . " while solving the test items. These conditions yielded 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.

To test the hypothesis that the instruction to verbalize made subjects assume a more analytic problem-solving style, Höfgen (cited in Merz, 1969) compared performance on parallel forms 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 than before, but still somewhat better than the control group, whose performance hardly differed between the two occasions.

Inner speech (measured by electrical activity in the speech apparatus) during performance on Raven's Progressive Matrices items without overt verbalization 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 predominantly visual way, whereas 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 pointed 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 pro-

vides 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 performance with verbalizing is equal to or less than the variance of performance under control conditions in the studies of Kesting and of Waszak and Höfgen (cited in 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 on all subjects.

Other Concurrent Verbalization Studies

None of the studies that remain to be discussed employed highly manipulative tasks or pictorial stimuli.

In a study on clinical judgment (Baranowski, Note 5), 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" (p. 21). There was no difference between control and verbalizing conditions (measured by variance accounted for by linear and nonlinear 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 concept-learning study by Bower and King (1967), one group of subjects was required to verbalize their hypotheses before classifying the stimuli, but a control group was not. In preparation for the experiment, the subjects described the stimuli to ensure that subject and experimenter agreed in their descriptions. Under these circumstances, we would expect that no further 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 significantly improved performance (i.e., number of responses to criterion), but only for the first problem. Bower and King found that variation of the number of irrelevant features or dimensions affected only the 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.

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 were to 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, & Liljergren, 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 cited 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 subjects more independent of their direct perceptions. In the study of Baranowski (Note 5), unlike the other studies, the subjects were highly skilled. In our model, verbal explanation of automated activities would be cumbersome and would 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

With this summary of effects of concurrent verbalization (or the absence of such effects), we turn now to the topic of retrospective verbalization. We will first consider some experiments that are often cited to support the idea that attempts to verbalize information 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 it caused a decrement in ability to use the concept in a transfer situation. These results were substantiated in subsequent work 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 our discussion of incompleteness of verbal reports, we will return to this issue.

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 generalizations, found no differences as compared with a control condition. He also cited two studies in which no effects of producing written descriptions of learned generalizations were found. Sowder proposed that the important difference between his study and Hendrix's (1947) was that Hendrix required complete specification of the content of her subjects' verbalization (quantifiers, domain, and so on), whereas he did not.

Rommetveit (1960, 1965) and Rommetveit and Kvale (1965a, 1965b) studied concept formation in a situation in which 12- to 13-year-old subjects played on a wheel of fortune with different pictures being displayed when the subjects were to win or lose, respectively. They found that instructing the subjects that they were subsequently to describe the differences between the "win" and "lose" figures, as opposed to just playing on the wheel, influenced subjects' retrospective descriptions of the two figures. Other procedural variations, such as 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 the instructions. Without such direction, verbalization seems to have no effect on the cognitive processes.

A number of studies have not found any effects from instructions to give verbal reports. In a series of studies of probabilistic inference, Brehmer (1974) and his co-workers have investigated the effects of asking their subjects to describe retrospectively their hypotheses about the relations between the cue and the criterion. The subjects were asked to describe, explicitly enough for someone else to make the predictions, their rules for arriving at the prediction from the cue value. However, subjects were free to report that they were guessing, were remembering from earlier trials, and so on. In a factorial study (Brehmer, 1974) in which subjects gave descriptions from trial to trial, no main effects

or interactions could be attributed to verbalization.

Two other studies in which the verbalized descriptions were generated prior to blocks of test trials also failed to find effects of verbalization (Brehmer, 1974; Brehmer, Kuylenstierna, & Liljergren, 1974, 1975). In a concept acquisition study using auditory stimuli, Wilson and Spellacy (1972) found no difference in number of trials or errors to criterion between a control group and a group of subjects who told the experimenter what rule they used. Asking subjects what they thought was the correct solution before each trial in a discrimination-learning experiment was not found to change the proportion of correct responses (Karpf & Levine, 1971). The proportion of correct placement of cards in two piles was not related to whether or not the subjects were verbalizing the rule they were using (Dulaney & O'Connell, 1963; Verplanck, 1962).

The fact that encoding and reporting verbally take time creates a procedural difference, which is important in some studies, between the verbal reporting and control conditions. Although time may not be important in many tasks, it is known to be an important variable in long-term memory phenomena. Boersma, Conklin, and Carlson (1966) allowed their subjects in the verbal report condition an additional minute to specify their encoding strategy for each stimulus. Retention scores were superior for the verbal report condition, but the experimental design confounds the effect of the additional time allowed for verbalization with effects of generating a written description of the encoding strategy.

Discussion

Even though the empirical data from systematic studies investigating the effects of verbalizing are relatively modest, the results of these studies consistently support our model's prediction 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 subjects to recode information in order to report it may affect these processes. Our model assumes that *only information in focal attention* can be verbalized. In our model, as 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 that are often thought to proceed in parallel) and the slow serial processes that are executed under cognitive control—a distinction, that is, between preattentive and focally attended processes (Neisser, 1967), perceptual and cognitive processes (Simon, 1979), and automatic and cognitively controlled processes (Shiffrin & Schneider, 1977). 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.

Completeness of Verbal Reports

When subjects do not verbalize information that the investigator has strong reasons to assume they would need to have available in order to perform the task, it is reasonable to conclude that the protocols provide only an incomplete record of the process. For example, Rees and Israel (1935) found that subjects could acquire a set for solving anagram problems in a specific way without reporting the similarity of their solutions and of the structures of the anagrams. Similarly, many concept-formation studies have shown that subjects can display consistent and accurate behavior without always being able to report verbally the concepts employed by the experimenter (Heidbreder, 1934, 1936; Smoke, 1932). With respect to learning, many investigators have shown subjects to be unable to report reinforcement contingencies used by the experimenter (e.g., Greenspoon, 1955) or the use of mediating associations in paired-associate learning (e.g., Bugelski & Scharlock, 1952).

What kinds of incompleteness in reporting would be predicted in the framework of our model of verbalizing? The model proposes

that verbal reports are based on the information currently in STM or on information previously in STM that has been fixated in, and can be retrieved from, LTM. Within the model, we can identify three different causes of incompleteness of reports: (a) The information is not heeded, hence not stored in STM, hence not accessible for verbal reporting. (b) Not all the information available in STM at the time of the report is actually reported. (c) Not all of the information previously available in STM has been retained in LTM, or is retrievable from LTM.

Unavailability of Information in STM

Under a variety of circumstances, information about ongoing cognitive processes may simply not be available in STM. Several types of cognitive processes, like perceptual-encoding processes, motor processes, and LTM direct retrieval processes, appear not to use short-term memory for storage at intermediate stages of processing, but only for the final product. On this point, model and empirical studies are in full agreement.

Generally, we recognize familiar faces, words, and objects directly, that is, without storing in STM the features extracted from the stimuli and used for discrimination. There is evidence, also, for direct recognition of more complex patterns and relationships, especially when the presentation is visual. Claparède (1933) found that his thinking-aloud subjects did not report intermediate stages when generating interpretations and hypotheses for complex visual stimuli. Similar findings are reported for subjects noticing relationships in geometry problems (Henry, 1934).

In recall and retrieval of familiar information, unless it requires problem solving with the aid of successive associations, we frequently find processes that leave only the final product as trace in STM. The phenomenon is so familiar that it appears not to have been tested by experiment. There is ample evidence, however, from introspective reports directed at the issue of the existence of imageless thought for Woodworth (1938) to reach the conclusion that "what is imageless is not thought as much as recall" (p. 787).

With respect to perceptual–motor processes, also, it is clear that we normally have access only to certain higher level intermediate results. In a recent article, Broadbent (1977) presented empirical evidence for a hierarchical organization of such processes. Most studies of perceptual–motor processes using thinking-aloud protocols and other verbal reports have examined problem solving with puzzles. In tasks allowing physical manipulation, Klinger (1974) found a relatively high frequency of higher level verbal evaluations of unverbalized solution attempts (e.g., "Yep," "Dammit," etc.), and of verbalizations of attention-control processes (e.g., "Let's see," "Where was I?", etc.). When engaged in perceptual–motor manipulation, subjects did not verbalize, and appeared not to be aware of, the lower level content or structure of their thought processes. For example, Ruger (1910) found that subjects could often solve one of his mechanical puzzles several times, yet they provided only a limited high-level account for the intermediate steps leading to the solution. It has been suggested that physical manipulation is different from thinking in not employink any internal (i.e., STM) representation (Durkin, 1937).

There appears to be a close (negative) relation between degree of practice and awareness of intermediate stages of a process. The early work of Watt and others suggested that the conscious content disappeared with extended practice and growing automaticity of the processes (Woodworth, 1938). More recently, Dean and Martin (1966) found that overlearning in paired-associate learning leads to a decrease in the number of reported mediating associations. The work of Schneider and Shiffrin (1977; Shiffrin & Schneider, 1977) suggests that there are clear differences between automatic and controlled processing, in terms of speed and of accessibility for modification and learning. From all this evidence, it seems necessary to postulate, as we have, that many highly overlearned processes operate automatically without leaving any more trace than their final result in STM.[3]

[3] This effect of automation may be explained thus: Before overlearning has occurred, processes

We may distinguish between automatic processes that subjects already possessed prior to an experiment, as part of their cognitive skills, and processes whose intermediate stages became more automatic, and hence less reportable, during the course of the experiment. In the case of the latter, reports obtained from the automated processes at the end of the experiment will not give useful information about intermediate states of which the subjects were aware at the beginning of the experiment. We will later provide some concrete examples of this phenomenon.

An often cited study by Rees and Israel (1935) demonstrated that subjects could solve a long series of anagrams having identical structure and would select, in anagrams with multiple solutions, the solution corresponding to the common structure, without reporting awareness of that structure. The anagrams all used the simple permutation, 54123, which permitted rapid solution. It is thus not unreasonable to suppose that these anagrams were solved by an "automatic" process that did not leave intermediate results in STM. In a study by Sargent (1940), using thinking-aloud and retrospective reports, subjects were generally unable to report intermediate states when they achieved fast solutions to anagrams. More support for this interpretation is given by the findings of Rees and Israel that subjects recognized the similarity of anagrams when a more complex permutation of the letters was used. In fact, the only anagram pattern that was not reported by subjects was the one depicted earlier. In experiments in which subjects are asked to do similar tasks over and over, the subjects' latencies become markedly shorter with practice. Relying on the subjects' retrospective reports along with other evidence, several investigators have shown that with practice, subjects change from attending to the meaning of the presented information to recognition of invariant perceptual characteristics of the display (Quinton & Fellows, 1975) or purely formal properties of the presented information (Wood & Shotter, 1973). Wood, Shotter, and Godden (1974) gave direct experimental evidence of this transition in showing a marked relation between subjects' ability to answer unexpected questions—that would require attending to the meaning of the presented information—and practice.

Failure to Report STM Contents

Subjects tend to stop verbalizing or to verbalize incompletely in conditions in which 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). On the other hand, in situations in which the subject is not judged to be performing major task-directed processes, verbalization tends to be relatively complete. Ericsson (1975a) calculated the association between regularly observable features of the process (e.g., reversal of a pair of moves) and specific kinds 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).

Apart from cases in which the subjects rely on automatic processes, there are other forms of incompleteness in reports in which information that was once in focal attention (in STM) is not verbalized. Our model asserts that the verbal report will be based on information that is available to the subject at the time of the report. From this assumption, we expect information in STM to be reported in full, whereas information in LTM will be differentially accessible for various reasons. From extensive research on information stored in STM, it is clear that such information is easily obliterated. In a matter of a few seconds, the contents of STM can be destroyed or made inaccessible by requiring subjects

---

have to be interpreted, with substantial feedback from intermediate processing stages into STM. Overlearning amounts to compiling these processes, so that fewer tests are performed when they are being executed, hence less information is stored at intermediate stages in STM. Experience with compiling in computer languages shows that automation typically speeds up a process by an order of magnitude, at the expense of making it less flexible, and its intermediate stages less available for report.

to perform certain types of processes, for example, repeatedly subtracting 7 from a given number (Brown-Peterson paradigm). With any shift in the focus of attention, of which this paradigm provides an example, the previous contents of STM become unavailable. If an intermediate result in a sequence of processes causes a direct execution of other processes that make full demands on STM, the intermediate result may reside for only a brief moment in STM, and may be lost before being reported. Under thinking-aloud conditions, it has been observed (Duncker, 1945) that information that leads to the direct recognition of the appropriate action often tends not be be verbalized. Similar observations have been made (de Groot, 1965) about the reports of chess grandmasters considering possible moves in chess positions.

A frequently cited study by Maier (1931) on subjects' retrospective reports about a hint given during solution of the pendulum problem gives some evidence for the same phenomenon. Subjects who described the solution as emerging in a single step did not report any memory of the hint. A result of Maier's study that is less often recognized was that *all* subjects who mentioned more than one step in the solution of the problem reported that the hint had been administered.

The two related mechanisms mentioned thus far—absence of intermediate stages of acts of recognition from STM and failure to report transient contents of STM—are fully adequate to account for the phenomena of sudden "insight" that are the subject of so many anecdotes in the literature of creativity (Nisbett & Wilson, 1977, pp. 240–241). The studies cited in support of sudden insight are based on retrospective accounts of purportedly real creative acts, often reported many years after the event. Fortunately, a number of studies have addressed this topic in a more controlled experimental environment. Durkin (1937) sought to create favorable circumstances for "insight" with subjects thinking aloud while solving block puzzles. Even though subjects occasionally reported insights, the background steps leading to the emergence of the insightful ideas could always be determined from the concurrent thinking-aloud protocols. Of this kind of insight, Durkin (1937) says:

> When it occurs, it comes with an onrush that makes it seem very sudden—an "out of the blue" experience. But it can always be found to have developed gradually. The suddenness must be regarded as due to the concealment of the background. It does not bring in a new *kind* of process. (p. 81)

For geometry proofs (Henry, 1934), and for a variety of "insight" problems (Bulbrook, 1932), the thinking-aloud protocols showed that the progress to solution was either gradual or was determined by trial and error. In neither case was it necessary to postulate additional kinds of processes.

Insight or the illumination of a creative idea was described by many early investigators as the result of a period of unconscious work, or *incubation,* following preliminary work in becoming familiar with the problem (*preparation*). During this period of alleged incubation, according to these accounts, the scientist or inventor has laid the problem aside in favor of other activities. However, it has been suggested (Woodworth, 1938) that during the period of incubation, the scientists will occasionally lapse into thinking about the problem, even though their main activities may be different.[4] Woodworth (1938) cited a study by Platt and Baker that suggests that subjects are not aware of the durations of these unplanned episodes of concentrated thought on the problem. Clearly, such unanticipated thought processes, if they occur, will be very difficult to retrieve in retrospect. Generally such episodes are terminated, often abruptly, by external demands, as, for example, in a driving situation.

From research on daydreaming (Singer, 1975) and undirected thought (Klinger, 1971), there are suggestions that such thought episodes are difficult to recall fully unless the retrospective reports are obtained shortly thereafter, or unless the subjects label or rehearse the thought content for subsequent recall. Studies using concurrent verbalization (Bertini, Lewis, & Witkin, 1964; Kazdin,

[4] A different explanation of incubation, based on forgetting of STM contents, has been proposed by Simon (1977, pp. 296–299).

1976; Klinger, 1971) provide detailed and informative accounts of undirected thought processes. In sum, thinking that is not closely related to the external environment can sometimes be retrieved with situational clues but seldom otherwise, except when it is verbalized concurrently with the thought process.

Incompleteness in Retrieval from LTM

We consider next the situation in which a subject is probed for information that is not available in STM at the time of probing. Then the information must be retrieved from LTM. Memory retrieval is fallible and sometimes leads to accessing other related, though inappropriate, information. Further, the information that can be recalled depends on what cues and probes are provided. Hence, the completeness of the information retrieved will vary with the probing procedure.

Verbalizing rules in concept attainment. In reviewing evidence for incompleteness in reports, we will be concerned primarily with studies of concept attainment and learning. Such processes can hardly be carried out automatically, since they require reprogramming of responses to stimuli; hence, the stimuli should be attended to and thus available in STM, at least before extended practice. We can thus be reasonably sure that the information corresponding to these changes in behavioral regularities has at least at one time resided in STM. To test our hypotheses about the less complete report of information from LTM, as compared with STM, we will review evidence allowing us to infer what information would reside in STM at the time of a verbal report. In particular, we will be concerned with empirical studies showing that subjects can learn to behave in agreement with concepts or rules without being able to verbalize the concepts or the stimulus attributes they are responding to. We will also discuss some other studies that have examined whether experimentally controlled factors facilitating learning are mediated by reportable intermediate states.

In concept learning, it has been observed that subjects can select appropriate instances in test trials without being able to state the concepts they are using (Heidbreder, 1934, 1936; Phelan, 1965; Smoke, 1932). Although these studies have been cited in support of the notion that the verbalized information is incomplete, a more reasonable interpretation is that the subjects cannot formulate the concept as it is defined by the experimenter, although they can differentiate instances from noninstances in a test run. However, correct selection of instances can also be mediated by processes like memorizing exemplars or a set of correlated discrimination features without using a rule expressed in terms of common features.

When Smoke (1932) asked subjects to identify the concept used by the experimenter from verbal descriptions, he found in four experiments that 20%–25% of the verbal descriptions of the successful subjects were defective, and "usually too inclusive" (p. 20). This study, and other related ones, do not address the incompleteness problem as we would like to state it: Is the verbal report a complete or sufficient description of *the information the subject actually has and uses?* In Smoke's (1932) study the subjects who were classified as unable to verbalize the concept could "almost invariably" (p. 20) draw two instances of the concept from memory correctly, but this fact does not imply that they had a complete and correct (though unverbalizable) criterion for making the selections.

The question of completeness has been put to a more direct test in an interesting study by Wilson (1974, 1975). In his study of concept learning, the subjects wrote down their rules for positive instances during each trial and were then asked to sort a test series of instances. Wilson then assessed the information transmitted in these verbal descriptions by having the same subjects a week later make re-sorts from the descriptions. In addition, naive subjects who did not participate in the concept-learning experiment were asked to sort the test series on the basis of the individual verbal descriptions provided by the original subjects. The results showed that the sorts made by subjects after a week's delay agreed less closely with their original sorts during the concept-learning experiment than they did with the sorts made by the naive

subjects. This is evidence that the subjects had more information at the time of the original experiment than they gave in their verbal reports. It should be pointed out, however, that the correspondence between all the sorts was high, especially considering that the verbal rule was not always applicable to all instances in the test series.

In a second experiment, Wilson (1974, 1975) found that the degree of incompleteness of descriptions varied with the stage of concept learning at which the verbalization of the rule was obtained. For rules verbalized either at the beginning of the experiment or at criterion, almost complete agreement was obtained between sorts, but more discrepancies were found for rules verbalized at intermediate stages of acquisition. Wilson attributed these results to subjects' difficulties in verbalizing the complex hypotheses they entertained during intermediate learning stages, and this interpretation was further supported by the greater length of the verbal descriptions for these stages.

One alternative explanation, suggested previously, would be that the instructions used by Wilson (to verbalize a rule—the average number of words per rule was about eight) were inadequate to tap the subjects' information about positive instances. Another possible interpretation is that instances in the test series may have served as cues for retrieval and recognition of previously presented items that were not available to the subjects when they generated the verbal report. Some supporting evidence is given by a study on discrimination learning by Frankel, Levine, and Karpf (1970), in which the subjects could give a retrospective description of "on what basis they had responded" (p. 346) that described more than 90% of their responses.

Learning without awareness. The controversy over whether learning can and does occur without awareness, as evidenced by verbal reports from subjects, has recently been reviewed by Brewer (1974). Drawing on his work, our primary aim will be to interpret the differences, primarily methodological, that distinguish studies of learning without awareness from studies without such learning.

One type of study has been criticized repeatedly for poor documentation of probing procedures and brief postexperimental interviews, which are sometimes not given after the learning trials but after an additional series of extinction trials (Spielberger, 1962). This type of experiment does not contradict the possibility that subjects retain in STM the information about the reinforcement contingency until it is lost during overlearning or extinction trials. To ensure as complete a verbal report as possible, the probing should occur just after the last learning trials, preferably before the subject is told that the experiment is over, to ensure that the critical information remains in STM and does not require retrieval from LTM. Many of the studies of this first type seem little concerned with these considerations. Furthermore, the verbal probes used have been global questions, like "What did you think the experiment was about?" (Brewer, 1974). In our framework, such verbal probes are not aimed at eliciting retrospective memory of the subjects' own cognitive processing, but rather at encouraging them to generate hypotheses about the experiment, which may or may not be related to those processes.

A second type of study, which has responded to the earlier mentioned criticisms by probing subjects just after the last learning block of trials and explicitly asking them for their memories of the cognitive processes during the preceding trials, has generally failed to find evidence for learning without reported awareness of the reinforcement contingency. The studies of this type have been criticized, in turn, for asking questions that are too specific, and hence suggesting awareness to the subjects. We have already discussed this issue in the section on effects of probing. The need for specific probes is not well documented, but a general motivation for it is given in Dulany (1962, p. 114). Another important consideration is that the subjects often report contingencies that although not identical to the one the experimenter reinforces, are correlated with it (Dulany, 1962). Here, failure to report the experimenter's version of the contingency may simply mean that this is not the version the subject is using, and may not at all imply incompleteness in the report of the contents of STM.

In a third type of study, subjects are asked to verbalize during conditioning experiments according to a technique suggested by DeNike (1964), in which the subject writes down "any thoughts that come to you that have any relation to the experiment" (p. 523). Using this procedure, DeNike was able to extend the claim of consistency between written thoughts and behavior in agreement with the reinforcement contingency, finding that subjects' behaviors changed on the trial in which they wrote down their first correct hypothesis about the reinforcement.

Also using DeNike's technique, Kennedy (1970, 1971) found that the behaviors changed *before* subjects wrote down that they were confident of their hypotheses or before their verbal hypotheses were confirmed. Brewer (1974) pointed out that Kennedy's finding, that behavioral changes were associated with trying out or modifying verbal hypotheses, does not challenge, but supports the validity of the verbalized information.

There is evidence from comparing different probing techniques with each other that the written thoughts elicited by DeNike's technique are incomplete as compared with information obtained using Dulany's postexperimental questionnaire (Sallows, Dawes, & Lichtenstein, 1971). There is also some evidence that the incompleteness may be caused by the requirement of written responses. Silveira (1972) found a marked difference in number and character of responses elicited during a creativity test between written and oral response conditions. The process of writing the responses or ideas, as contrasted with giving them aloud, was found to be linked to evaluation and censorship.

It has been assumed implicitly that the contents of STM after the last learning block are representative of the contents during the preceding trials. This is a reasonable assumption for concept learning and for experiments with deterministic reinforcement schedules, in which there is no inducement for the subject to reject a correct hypothesis. Several studies have shown that subjects are unlikely to change hypotheses in response to positive feedback (Heidbreder, 1924; Karpf & Levine, 1971). However, since subjects are highly likely to change hypotheses in response to negative feedback, the assumption of stable hypotheses is less tenable for probabilistically determined feedback conditions.

In a recent study by Williams (1977), all subjects who successfully learned the discrimination were assessed (by a procedure of the similar to Spielberger, 1962) to have been aware of the reinforcement. However, in a second experiment with probabilistic relations, Williams found evidence of learning without awareness (assessed by the same procedure). During the experiment, subjects were explicitly told that "All the sentences I said 'correct' to met the same necessary requirements" (p. 93). These instructions might have encouraged subjects to abandon correct and correlated hypotheses when they encountered probabilistically determined negative feedback. Hence, the subjects may very well have entertained correct or correlated hypotheses during the trials evidencing learning, but they may have discarded them when negative feedback was encountered. A study by O'Connell (1965) shows clearly that if the subjects verbalize their hypotheses on each trial, the verbalized reports account for their behavior even in a nondeterministic environment with partial reinforcement.

There is a related issue of awareness of mediating associations. A series of studies (Bugelski & Scharlock, 1952; Horton & Kjeldergaard, 1961; Russel & Storms, 1955) have found that paired-associate learning can be facilitated by prior exposure of the subjects to the proper mediating associations. Subjects who first learned lists of paired associates of Types A-B and B-C learned lists of Type A-C faster than control groups. In all of these studies, informal postexperimental questioning gave no evidence that the subjects were aware of using any mediating B-list items. "None of the *S*s [subjects] was able to report any correct appreciation of the nature of the experiment and most assuredly did not verbalize a pattern of A-B, B-C, A-C in learning the third list" (Bugelski & Scharlock, 1952, p. 366). This was interpreted as evidence for unconscious mediation in learning.

However, the previously cited studies can

be criticized on the same grounds as the studies purporting to demonstrate concept learning without awareness; typically, we find different results in more recent studies that probe for retrospective information in a more controlled and ambitious manner. In a study by Dean and Martin (1966), the subjects, after reaching criterion for the third list, read each stimulus of the last list and were then asked to tell the experimenter exactly what came to their minds when they saw the syllable on the screen. The subjects were then shown the entire list of paired associates and were asked how the list was learned. Dean and Martin found that a majority of the subjects reported using at least one mediating term from the previously learned facilitating lists. An analysis of learning rates for each of the reported mediation types (other than A-B-C mediation) clearly suggested that the effective difference among groups was attributable to the occurrence or nonoccurrence of A-B-C mediation. When Dean and Martin had one experimental group overlearn the paired-associate list for 10 extra trials, they found a significantly lower occurrence of reported mediation, thus suggesting that direct and automatic processing had developed as a result of additional practice.

In a study following a procedure similar to that in Horton and Kjeldergaard (1961), Horton (1964) used a direct question: "Did you notice any relationship between the pairs you just completed and the ones you learned earlier in the experiment?" Horton assessed three levels of awareness, where the highest level required naming the actual mediating items. He found a consistent relation between mediation (effect of previously learned facilitating paired-associate lists) and assessed awareness for a variety of experimental manipulations.

There are many differences between the studies reporting awareness of mediating items and those that report no awareness. Horton (1964) suggested that the difference between his study and Horton and Kjeldergaard (1961) stemmed from the factors in the experimental situation stimulating awareness. The two groups of studies also differ in the strictness of the learning criterion, which might have affected the retrievability of the mediating links. However, the most obvious difference, in our view, lies in the probing procedure.

Other studies alleging incompleteness. We will mention briefly several miscellaneous studies claiming incompleteness of retrospective reports. Rees and Israel (1935) found that anagrams were solved faster if the solution words were sampled from specific types of words, like "nature words" (names of plants and trees). Subjects reported in the postexperimental questioning that they noticed the relation between the solution words, but most of them also reported that they did not actively use that information.

In a study of the effect of reversal in discrimination learning, Walk (1952) found that the reversal had a short-term effect on the behavior of the reversed group that was not matched by verbal evidence of awareness of the reversal. With the pictorial stimuli, Walk also found that subjects were occasionally unable to define the reasons for their correct selections.

In a study of concept formation, Heidbreder (1924) found that subjects referred to aspects of figural stimuli presented earlier that they had not previously considered in the retrospective reports, as illustrated by a quote from a verbal report, "I've been wondering if the ones I've had right haven't always had more lines in the figure I marked. I think they have but I'm not sure" (p. 136).

Discussion

Under a variety of circumstances, verbal reports may omit information that subjects use to perform the task. Evidence of the nature of the omissions is consistent with the predictions of our model. The intermediate stages of immediate recognition processes and the detailed steps of perceptual–motor processes are not generally recorded in STM; hence, they are not reported. Processes that have been so often repeated as to have become automated are less often and less fully reported.

When subjects give indications that they are working under a heavy cognitive load, they tend to stop verbalizing or they provide

less complete verbalizations. The contents of STM can be obliterated, hence, omissions caused in reports, by requiring subjects to perform intervening tasks concurrently with their verbal reporting.

The evidence for inability to report the bases for sudden insights is mainly anecdotal and is refuted by the few laboratory studies that have been made of insight phenomena. Nor is the evidence, all of it anecdotal and retrospective, of the unreportability of thought processes during incubation convincing. Alternative explanations of these phenomena are available that are consistent with our model.

When clear probes are used for specific retrospective memory and when reports are requested immediately after the last trial(s), informative verbal reports can usually be obtained, although perhaps not always in the case of complex pictorial stimuli. The failure of subjects to report some information does not demonstrate the uselessness of verbal protocols. Incompleteness of reports may make some information unavailable, but it does not invalidate the information that is present. In an often cited remark, Duncker (1945) observed that "a protocol is relatively reliable only for what it positively contains, but not for that which it omits" (p. 11).

Consistency of Verbal Reports With Other Behavior

So far we have been concerned with the failure of subjects' reports to contain information that one would expect was heeded and hence was in STM at some time. In this section we will be concerned with verbal reports that are inconsistent with other sources of data—primarily with observable nonverbal behavior.

Claims that verbalized information is inconsistent with other behavior are often made in general and sweeping terms without providing specific evidence. When evidence is provided, it is often anecdotal, resting on the premise that if one can produce a single case of a clearly inaccurate or inconsistent verbal report, then verbal reports are wholly inadmissible as data. In this section we will consider under what circumstances inaccurate and inconsistent verbal reports have been observed. In most of the cases in which inconsistency has been observed or claimed, the verbal reports were retrospective.

Potential Sources of Inconsistency

Within our model, inconsistent verbal reports could be produced by two processes. First, cues used to access LTM, if too general, could retrieve information related to, but not identical with, the information that was actually sought. There is evidence that subjects have information (in the form of differential confidence) about the correctness of LTM retrievals (see Montague, 1972); this information could be tapped to decrease the number of retrieval errors stemming from this cause.

The second source of inconsistent information, already discussed in some detail, is the use by subjects of intermediate processes to infer missing information and to fill out and generalize incomplete memories before responding. An example is provided in the study of Rommetveit and Kvale (1965b), discussed earlier. Patterns were displayed to signal to subjects playing a wheel of fortune whether they would win or not. When the experimenters asked a boy to describe the differences between positive and negative patterns, he said he did not know, although he had been able to anticipate the rewards correctly on previous trials. When pressed by the experimenters, he finally attempted a verbal description and gave one that was inconsistent with the actual signals. In this and many other similar accounts, it seems appropriate to attribute the error to absence of the information from memory, rather than inconsistency between memory contents and verbal reports of them. When information is not in memory, it cannot be reported verbally.

In a review article, Smedslund (1969) claims that verbal reports did not provide useful information on rapid mental processes in an arithmetic task he had studied (Smedslund, 1968). He could describe two reports that could be proven inconsistent—one inconsistent with the subject's performance and speed in solving the test items, the other inconsistent with the types of items actually

given. In the original article, however, Smedslund does not mention inaccuracy in the verbal reports and actually quotes those reports to support some of his general results. In Smedslund's study, the conditions for verbal reporting were far from optimal, for the subjects were asked after a relatively large number of trials to tell how they solved the tasks. Even under these unpromising circumstances, the verbalized information, with a few exceptions, seemed to satisfy the author as basically consistent with his observations of subjects' performance.

Provided that our procedure for analyzing data can handle occasional errors in the verbal information, and provided that these errors can be minimized by appropriate procedures (including recognition of the circumstances under which subjects cannot be expected to remember certain information), anecdotal evidence of the sort just cited need not shake our confidence in the validity and legitimacy of verbal reports.

Inconsistency With Concurrent Verbalization

In an often cited study (Verplanck, 1962), Verplanck and Oskamp claimed to have shown that verbalized rules are dissociated from the behavior they were supposed to control. This study is the only one we have found claiming inconsistency of *concurrent* verbal reports with behavior. By having the subjects verbalize the rules they were following in sorting illustrated cards, the experimenters could reinforce either the verbal rule or the placement of cards (i.e., behavior). To make the contingencies less noticeable, the partial reinforcement followed the criterion trials. When correct placements were reinforced, the subjects placed cards correctly in 71.8% of the trials, but they stated a correct or correlated rule on only 48.4% of the trials. When the correct statement of the rule was reinforced, the subjects stated a correct or correlation rule on 92.8% of the trials but placed the cards correctly on only 76.8% of the trials.

In a replication and analysis of the Verplanck-Oskamp experiment, Dulany and O'Connell (1963) showed that the previously mentioned results could be attributed to two artifacts of the original experiment. First, consider the case in which correct placement was reinforced. Making a correction for guessing (the subjects had a 50–50 chance of placing the card in the correct pile when they did not know the rule), we can estimate that subjects *knew* the correct answer in 43.6% of the trials—a percentage very close to the 48.4% in which they stated the correct rule.

Second, with respect to the reinforcement of rules, Dulany and O'Connell (1963) found that the correct rules used by Verplanck and Oskamp were ambiguous for the card illustrations they employed. In fact, naive subjects, who were told the rules explicitly, generated the same proportion of misplacements as was recorded in the original experiment.

In a detailed analysis of the rules the subjects verbalized on each trial, Dulany and O'Connell (1963) found that on all but 11 of 34,408 trials, the subjects put the card where they said they were going to. Hence, Dulany and O'Connell impeached rather thoroughly the evidence put forth by Verplanck and Oskamp for believing that the verbalized rules were inconsistent with the behaviors.

Numerous studies provide documented support for consistency between verbalized rules, concepts, and hypotheses, and immediately proceding and succeeding behavior, before subjects receive feedback. In Schwartz (1966), where subjects were asked their reasons for placing a card as they did, reasons consistent with placements were given on all but 2 of 1,962 trials. Even more impressive, Frankel et al. (1970) obtained retrospective reports from subjects on the basis of their responses to four earlier discrimination-learning problems with 30 nonfeedback trials each and found that subjects could provide reports in more than 90% of the sequences of trials.

The Nisbett-Wilson Literature Review

In a recent, extensive review of studies permitting comparison of retrospective verbal reports with behavior, Nisbett and Wilson (1977) have stated conclusions that appear at first sight to be almost diametrically opposite to those reached in this article. Since

the article by Nisbett and Wilson has received widespread attention, it is important to consider how their findings are to be reconciled with ours.

Nisbett and Wilson (1977) summarized their main empirical findings as follows:

> People often cannot report accurately on the effects of particular stimuli on higher order, inference-based responses. Indeed, sometimes they cannot report on the existence of critical stimuli, sometimes cannot report on the existence of their responses, and sometimes cannot even report that an inferential process of any kind has occurred. (p. 233)

First, we call attention to the frequent use, in their summary, of the qualifiers "often" and "sometimes." Nisbett and Wilson cited a large number of experiments that support their conclusions, but they did not investigate in detail the *conditions* under which these conclusions do and do not hold. Moreover, they did not propose a definite model of the cognitive processes as a framework for interpreting the findings they surveyed. Their theoretical interpretations of these findings are entirely informal, resting heavily on an undefined distinction between introspective access to "content" and to "process," or, as they alternatively state it, between access to "private facts" and to "mental processes." Nisbett and Wilson's (1977) summary of the kinds of information to which subjects *do* have access is this:

> We do indeed have direct access to a great storehouse of private knowledge. . . . The individual knows a host of personal historical facts; he knows the focus of his attention at any given point of time; he knows what his current sensations are and has what almost all psychologists and philosophers would assert to be "knowledge" at least quantitatively superior to that of observers concerning his emotions, evaluations, and plans. Given that the individual does possess a great deal of accurate knowledge . . . it becomes less surprising that people would persist in believing that they have, in addition, direct access to their own cognitive processes. The only mystery is why people are so poor at telling the difference between private facts that can be known with near certainty and mental processes to which there may be no access at all. (p. 255)

Nisbett and Wilson (1977) also observed that subjects "are often capable of describing intermediate results of a series of mental operations (p. 255)," that is, they hold in STM and can access the symbols that are inputs and outputs to such operations.

We may compare this list of "private facts" and intermediate results that according to Nisbett and Wilson *are* accessible to subjects with the kinds of verbalizations we have been considering in our review of the evidence. The individuals know, they say, their focus of attention, their current sensations, their emotions, their evaluations, and their plans. They know the intermediate results of their mental operations. But these are exactly the kinds of information that according to our model and the evidence we have examined, are held in STM and are available for verbal reports. Only one kind of item that we have considered is missing from this definition of "content": awareness of ongoing processes. If there is a discrepancy between our account and that of Nisbett and Wilson, it lies in that domain.

Unfortunately, the studies reviewed by Nisbett and Wilson provide little data as to what information is heeded during the thought processes and what information is accessible from STM and LTM at the time of the verbal report. Nisbett and Wilson (1977) simply assert that the subjects, when *asked questions about their cognitive processes,* frequently do not base their answers on memory for specific events at all, but "theorize" about their processes.

> When reporting on the effects of stimuli, people may not interrogate a memory of the cognitive processes that operated on the stimuli; instead, they may base their reports on implicit, a priori theories about the causal connection between stimulus and response. (p. 233)

In reviewing the studies cited by Nisbett and Wilson, we can profitably raise the question of *why* and *when* subjects do not consult their memories of cognitive processes in answering questions about those processes. It is easy to draw the erroneous conclusion that this independence of the verbal answers to the questions about cognitive processes from the actual course and results of those processes implies a general lack of accessible memory for such processes, or even an unawareness of the information while the process

was actually going on. But we have seen that such a sweeping conclusion is contradicted by the evidence from concurrent verbalization.

Drawing on our taxonomies of types of verbalization and of techniques for probing for information, we will show that in the studies reviewed by Nisbett and Wilson, very different procedures were used from those that according to our model, would elicit valid retrospective reports of processes. Before we discuss these differences, we want to point out that the studies reviewed in Nisbett and Wilson were neither designed for nor primarily concerned with determining subjects' memories of their cognitive processes. It would be preferable to test the implications of our model against a new set of studies directly designed to assess the information heeded by subjects and the information reported retrospectively for the types of cognitive processes investigated in this research. Until such studies have been carried out, we can only speculate on what information about the cognitive processes is heeded and stored in accessible form.

On the basis of the distinctions made in our taxonomy of verbalization and probing procedures, we find three important differences between the retrospective verbalization procedures that our model would recommend and the procedures used in studies reviewed by Nisbett and Wilson (1977). In some of these studies, the questions presented to subjects contain considerable background information that would make it feasible for subjects to generate answers without consulting their memories of the cognitive processes. With questions like, "I noticed that you took more shock than average. Why do you suppose you did? (Nisbett & Wilson, 1977, p. 237)," it is not even clear to us, nor probably to the subjects, that memory for the cognitive process *should* be the information source for the answer. If subjects can generate their answers without consulting their memories of the cognitive process (Nisbett & Wilson, 1977, showed that control subjects could do exactly that), this might often be more efficient than retrieving information from memory.

Moreover, in most of the studies reviewed by Nisbett and Wilson, the time lag between task and probe was sufficiently great as to make it unlikely that the relevant information remained in STM. When the probe is *not* a good retrieval cue for the relevant aspects of the memory of the process (see our discussion in the next paragraph), the subject must attempt, through conscious processing, to secure a sufficiently complete recall for giving the appropriate answer. Such retrieval from LTM requires considerable time and effort, and we would claim that subjects, unless explicitly instructed to provide a relatively complete recall, would be highly unlikely to do so, especially if other processing alternatives were available to them.

The second difference concerns the types of information that the studies reviewed by Nisbett and Wilson probed for, and the relation of those types of information to information stored about specific instances of processing. Our model predicts that information can be recovered by probes only under conditions in which that same information would be accessed by undirected concurrent or retrospective reports. Information about what such undirected reports would provide is lacking for most types of cognitive processes pertinent to the Nisbett-Wilson review. However, we find for many of the studies in that review that our taxonomy and model would predict failure to obtain from the probes verbal information about particular instances of processes. For example, in between-subjects designs, subjects obviously cannot answer from memory of their processes why they behaved differently from subjects in another experimental condition because the processes did not include such a comparison. Hence, this information can be derived, if at all, only by comparing the descriptions of the processes by different sets of subjects in the two conditions. In other studies the subjects were asked how they would have reacted if the experimental conditions had been different in a specified respect. Such probing for hypothetical states can never tap subjects' memories for their cognitive processes, since the information was never *in* memory. In still other studies, subjects were asked, explicitly

or implicitly, to summarize or generalize the processes they used, rather than reporting the processes used on each trial.

The third difference between the studies cited and our model concerns the time interval between the execution of cognitive processes and the probe asking them to be reported. We have already discussed this point earlier and in the section on incompleteness. Our review of "learning without awareness" and "memory for mediating associations" showed some of the procedural conditions in such experiments that consistently determined how available or unavailable memories would be and how complete or incomplete the reports were.

A few of the studies reviewed by Nisbett and Wilson (1977), including Maier's (1931) study of problem solving and the studies of learning without awareness, have already been discussed earlier in some detail, and have been found to be consistent with our model.

In sum, after exploring the implications of the studies reviewed by Nisbett and Wilson (1977) for our model, we find no difficulty in reconciling the experimental findings with our model. In most cases, the studies reviewed by Nisbett and Wilson are fully consistent with the model; in other cases they involve verbalization or probing procedures for which our model does not make strong predictions. Further discussion of the Nisbett-Wilson conclusions, agreeing generally with our critique but making some additional points as well, can be found in a recent article by Smith and Miller (1978).

General Discussion

When verbal reports are collected concurrently with other records of behavior, it becomes possible to check the consistency of the reports with the other behavior. Evidences of inconsistency can be found only under experimental conditions in which such inconsistency would be predicted by the model.

With our model, inconsistent retrospective reports can be produced as a result of probes that are too general to elicit the information actually sought, and as a result of subjects' use of inferential processes to fill out and generalize incomplete or missing memories. Studies in the literature provide examples of both.

On the other hand, the single study, by Verplanck and Oskamp (cited in Verplanck, 1962), purporting to find inconsistency between concurrent verbalizations and behavior, does not stand up under careful analysis. When the data were reanalyzed, they failed to support the claims of inconsistency. Meanwhile, a number of subsequent studies have shown a high level of congruence between verbal reports and other behavioral measures in a variety of experimental settings.

A well-known review article by Nisbett and Wilson (1977) is sometimes thought to provide evidence discrediting verbal reports as data. A close examination of the specific studies analyzed in their article shows that the instances cited of inconsistency between verbal reports and data all refer to experimental situations and procedures where our model would predict that veridical reports could hardly be expected. In fact, Nisbett and Wilson's own detailed summary of the conditions under which verbal reports can be assumed to be valid are consistent with the conclusions we have reached in this article.

For more than half a century, and as the result of an unjustified extrapolation of a justified challenge to a particular mode of verbal reporting (introspection), the verbal reports of human subjects have been thought suspect as a source of evidence about cognitive processes. In this article we have undertaken to show that verbal reports, elicited with care and interpreted with full understanding of the circumstances under which they were obtained, are a valuable and thoroughly reliable source of information about cognitive processes. It is time to abandon the careless charge of "introspection" as a means for disparaging such data. They describe human behavior that is as readily interpreted as any other human behavior. To omit them when we are carrying the "chain and transit of objective measurement" is only to mark as terra incognita large areas on the map of human cognition that we know perfectly well how to survey.

Reference Notes

1. Anderson, J. R. Personal communication, 1978.
2. Ericsson, K. A., & Simon, H. A. *Retrospective verbal reports as data* (CIP Working Paper No. 388). Unpublished manuscript, Carnegie-Mellon University, 1978.
3. Ericsson, K. A., & Simon, H. A. *Thinking-aloud protocols as data: Effects of verbalization* (CIP Working Paper No. 397). Unpublished manuscript, Carnegie-Mellon University, 1979.
4. Johnson, E. J., & Russo, J. E. *What is remembered after a purchase decision?* (CIP Working Paper No. 379). Unpublished manuscript, Carnegie-Mellon University, 1978.
5. Baranowski, T. *Clinical judgment as a cognitive process: A comparison of linear regression and decision net models.* Unpublished manuscript, University of Kansas, 1975.

References

Anderson, J. R. *Language, memory, and thought.* Hillsdale, N.J.: Erlbaum, 1976.

Anderson, J. R., & Bower, G. H. *Human associative memory.* Washington, D.C.: V. H. Winston, 1973.

Atkinson, R. C., & Shiffrin, R. Human memory: A proposed system and its control processes. In K. Spence & J. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2). New York: Academic Press, 1968.

Benjafield, J. Evidence for a two-process theory of problem-solving. *Psychonomic Science,* 1971, *23,* 397–399.

Bertini, M., Lewis, H. B., & Witkin, H. A. Some preliminary observations with an experimental procedure for the study of hypnagogic and related phenomena. *Archivo di Psicologia, Neurologia e Psichiatria,* 1964, *6,* 493–534.

Boersma, F. J., Conklin, R. C., & Carlson, J. E. Effects of reporting associative strategies on the retention of paired-associates. *Psychonomic Science,* 1966, *5,* 463–464.

Bower, A. C., & King, W. L. The effect of number of irrelevant stimulus dimensions, verbalization, and sex on learning bi-conditional classification rules. *Psychonomic Science,* 1967, *8,* 453–454.

Brehmer, B. Hypotheses about relations between scaled variables in the learning of probabilistic inference tasks. *Organizational Behavior and Human Performance,* 1974, *11,* 1–27.

Brehmer, B., Kuylenstierna, J., & Liljergren, J. E. Effects of functional form and cue validity on the subjects' hypotheses in probabilistic inference tasks. *Organizational Behavior and Human Performance,* 1974, *11,* 338–354.

Brehmer, B., Kuylenstierna, J., & Liljergren, J. E. Effects of information about the probabilistic nature of the task on learning of uncertain inference tasks. *Umeå Psychological Reports* (No. 90). Umeå, Sweden: Department of Psychology, University of Umeå, 1975.

Brewer, W. F. There is no convincing evidence for operant or classical conditioning in adult humans. In W. B. Weimer & D. S. Palermo (Eds.), *Cognition and symbolic processes.* Hillsdale, N.J.: Erlbaum, 1974.

Broadbent, D. E. Levels, hierarchies and the locus of control. *Quarterly Journal of Experimental Psychology,* 1977, *29,* 181–201.

Brunk, L., Collister, G., Swift, C., & Stayton, S. A correlation study of two reasoning problems. *Journal of Experimental Psychology,* 1958, *55,* 236–241.

Bugelski, B. R., & Scharlock, D. P. An experimental demonstration of unconscious mediated association. *Journal of Experimental Psychology,* 1952, *44,* 334–338.

Bulbrook, M. E. An experimental inquiry into the existence and nature of "insight." *American Journal of Psychology,* 1932, *44,* 409–453.

Carroll, J. S., & Payne, J. W. Judgments about crime and the criminal: A model and a method for investigating parole decisions. In B. D. Sales (Ed.), *Perspectives in law and psychology: The criminal justice system* (Vol. 1). New York: Plenum Press, 1977.

Claparède, E. La genèse de l'hypothèse. *Archives de Psychologie,* 1933, *24,* 1–155.

Dansereau, D. F. An information processing model of mental multiplication (Doctoral dissertation, Carnegie-Mellon University, 1969). *Dissertation Abstracts,* 1969, *30,* 1916B. (University Microfilms No. 69-15, 746)

Dansereau, D., & Gregg, L. W. An information processing analysis of mental multiplication. *Psychonomic Science,* 1966, *6,* 71–72.

Davis, J. H., Carey, M. H., Foxman, P. N., & Tarr, D. B. Verbalization, experimenter presence, and problem solving. *Journal of Personality and Social Psychology,* 1968, *8,* 299–302.

Dean, S. J., & Martin, R. B. Reported mediation as a function of degree of learning. *Psychonomic Science,* 1966, *4,* 231–232.

de Groot, A. D. *Thought and choice in chess.* The Hague: Mouton, 1965.

DeNike, L. D. The temporal relationship between awareness and performance in verbal conditioning. *Journal of Experimental Psychology,* 1964, *68,* 521–529.

Dulany, D. E. The place of hypotheses and intentions: An analysis of verbal control in verbal conditioning. In C. W. Eriksen (Ed.), *Behavior and awareness.* Durham, N.C.: Duke University Press, 1962.

Dulany, D. E., & O'Connell, D. C. Does partial reinforcement dissociate verbal rules and the behavior they might be presumed to control? *Journal of Verbal Learning and Verbal Behavior,* 1963, *2,* 361–372.

Duncan, C. P. Effects of instructions and information on problem solving. *Journal of Experimental Psychology,* 1963, *65,* 321–327.

Duncker, K. On problem solving. *Psychological Monographs,* 1945, *58*(5, Whole No. 270).

Durkin, H. E. Trial-and-error, gradual analysis and

sudden reorganization: An experimental study of problem solving. *Archives of Psychology,* 1937, 210.

Ericsson, K. A. Instruction to verbalize as a means to study problem solving processes with the 8-puzzle: A preliminary study (No. 458). *Reports from the Department of Psychology.* Stockholm, Sweden: University of Stockholm, 1975. (a)

Ericsson, K. A. Problem-solving behaviour with the 8-puzzle. IV: Process in terms of sequences of moves (No. 448). *Reports from the Department of Psychology.* Stockholm, Sweden: University of Stockholm, 1975. (b)

Ericsson, K. A., & Simon, H. A. Sources of evidence on cognition: A historical overview. In T. V. Merluzzi, C. R. Glass, & M. Genest (Eds.), *Cognitive assessment.* New York: Guilford Press, in press.

Feldman, J. An analysis of predictive behavior in a two-choice situation. Unpublished doctoral dissertation, Carnegie Institute of Technology, 1959.

Flanagan, J. C. The critical incident technique. *Psychological Bulletin,* 1954, *51,* 327–358.

Frankel, F., Levine, M., & Karpf, D. Human discrimination learning: A test of the blank-trials assumption. *Journal of Experimental Psychology,* 1970, *85,* 342–398.

Gagné, R. H., & Smith, E. C. A study of the effects of verbalization on problem solving. *Journal of Experimental Psychology,* 1962, *63,* 12–18.

Greenspoon, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *American Journal of Psychology,* 1955, *68,* 409–416.

Heidbreder, E. An experimental study of thinking. *Archives of Psychology,* 1924, 73.

Heidbreder, E. A study of the evolution of concepts. *Psychological Bulletin,* 1934, *31,* 673. (Abstract)

Heidbreder, E. Language and concepts. *Psychological Bulletin,* 1936, *33,* 724. (Abstract)

Hendrix, G. A new clue to transfer of training. *Elementary School Journal,* 1947, *48,* 197–208.

Henry, L. K. The role of insight in the analytic thinking of adolescents. *Studies in education, University of Iowa Studies,* 1934, *9,* 65–102.

Horton, D. L. The effects of meaningfulness, awareness and type of design in verbal mediation. *Journal of Verbal Learning and Verbal Behavior,* 1964, *3,* 187–194.

Horton, D. L., & Kjeldergaard, P. K. An experimental analysis of associative factors in mediated generalizations. *Psychological Monographs,* 1961, *75*(11, Whole No. 515).

Johnson, E. S. An information processing model of one kind of problem solving. *Psychological Monographs,* 1964, *78*(4, Whole No. 581).

Karpf, D. A. Thinking aloud in human discrimination learning (Doctoral dissertation, State University of New York at Stony Brook, 1972). *Dissertation Abstracts International,* 1973, *33,* 6111B. (University Microfilms No. 73-13, 625)

Karpf, D., & Levine, M. Blank-trial probes and introtacts in human discrimination learning. *Journal of Experimental Psychology,* 1971, *90,* 51–55.

Kazdin, A. E. Assessment of imagery during covert modeling of assertive behavior. *Journal of Behavior Therapy and Experimental Psychiatry,* 1976, *7,* 213–219.

Kennedy, T. D. Verbal conditioning without awareness: The use of programmed reinforcement and recurring assessment of awareness. *Journal of Experimental Psychology,* 1970, *84,* 487–494.

Kennedy, T. D. Reinforcement frequency, task characteristics and interval of awareness assessment as factors in verbal conditioning without awareness. *Journal of Experimental Psychology,* 1971, *88,* 103–112.

Klinger, E. *Structure and functions of fantasy.* New York: Wiley, 1971.

Klinger, E. Utterances to evaluate steps and control attention distinguish operant from respondent thought while thinking out loud. *Bulletin of the Psychonomic Society,* 1974, *4,* 44–45.

Lashley, K. S. The behavioristic interpretation of consciousness II. *Psychological Review,* 1923, *30,* 329–353.

Loftus, G. R. Eye fixations and recognition memory for pictures. *Cognitive Psychology,* 1972, *3,* 525–551.

Loftus, G. R., & Bell, S. M. Two types of information in picture memory. *Journal of Experimental Psychology: Human Learning and Memory,* 1975, *1,* 103–113.

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

Merz, F. Der Einfluss des Verbalisierens auf die Leistung bei Intelligenzaufgaben. *Zeitschrift für Experimentelle und Angewandte Psychologie,* 1969, *16,* 114–137.

Montague, W. E. Elaborative strategies in verbal learning and memory. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 6). New York: Academic Press, 1972.

Morgan, C. L. Characteristics of problem-solving behavior of adults. *Studies in Education, University of Iowa Studies,* 1934, *9,* 105–143.

Neisser, U. *Cognitive psychology.* Englewood Cliffs, N.J.: Prentice-Hall, 1967.

Newell, A., & Simon, H. A. *Human problem solving.* Englewood Cliffs, N.J.: Prentice-Hall, 1972.

Nisbett, R. E., & Wilson, T. D. Telling more than we can know: Verbal reports on mental processes. *Psychological Review,* 1977, *84,* 231–259.

O'Connell, D. C. Concept learning and verbal control under partial reinforcement and subsequent reversal or nonreversal shift. *Journal of Experimental Psychology,* 1965, *69,* 144–151.

Peterson, L. R. Concurrent verbal activity. *Psychological Review,* 1969, *76,* 376–386.

Phelan, J. G. A replication of a study on the effects of attempts to verbalize on the process of concept attainment. *Journal of Psychology,* 1965, *59,* 283–293.

Quinton, G., & Fellows, B. J. "Perceptual" strategies in the solving of three-term series problems. *British Journal of Psychology,* 1975, *66,* 69–78.

Ray, W. S. Verbal compared with manipulative solu-

tion of an apparatus-problem. *American Journal of Psychology,* 1957, *77,* 289–290.

Reed, S. K., & Johnsen, J. A. Memory for problem solutions. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 11). New York: Academic Press, 1977.

Rees, H. J., & Israel, H. E. An investigation of the establishment and operation of mental sets. *Psychological Monographs,* 1935, *46*(6, Whole No. 210).

Rommetveit, R. Stages in concept formation and levels of cognitive functioning. *Scandinavian Journal of Psychology,* 1960, *1,* 115–124.

Rommetveit, R. Stages of concept formation. II: Effects of an extra intention to verbalize the concept and stimulus predifferentiation. *Scandinavian Journal of Psychology,* 1965, *6,* 59–64.

Rommetveit, R., & Kvale, S. Stages in concept formation. III: Further inquiries into the effects of an extra intention to verbalize. *Scandinavian Journal of Psychology,* 1965, *6,* 65–74. (a)

Rommetveit, R., & Kvale, S. Stages in concept formation. IV: A temporal analysis of effects of an extra intention to verbalize. *Scandinavian Journal of Psychology,* 1965, *6,* 75–79. (b)

Roth, B. The effect of overt verbalization on problem solving (Doctoral dissertation, New York University, 1965). *Dissertation Abstracts,* 1966, *27,* 957B. (University Microfilms No. 65-9, 321)

Ruger, H. A. The psychology of efficiency. *Archives of Psychology,* 1910, 15.

Russel, W. A., & Storms, L. H. Implicit verbal chaining in paired associate learning. *Journal of Experimental Psychology,* 1955, *49,* 287–293.

Sallows, G. O., Dawes, R. H., & Lichtenstein, E. Subjective value of the reinforcer (RSv) and performance: Crux of the S-R versus cognitive mediation controversy. *Journal of Experimental Psychology,* 1971, *89,* 274–281.

Sargent, S. S. Thinking processes at various levels of difficulty. *Archives of Psychology,* 1940, 249, 5–58.

Schneider, W., & Shiffrin, R. M. Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review,* 1977, *84,* 1–66.

Schuck, J. R., & Leahy, W. R. A comparison of verbal and non-verbal reports of fragmenting visual images. *Perception & Psychophysics,* 1966, *1,* 191–192.

Schwartz, S. H. Trial-by-trial analysis of processes in simple and disjunctive concept-attainment tasks. *Journal of Experimental Psychology,* 1966, *72,* 456–465.

Shiffrin, R. M., & Schneider, W. Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review,* 1977, *84,* 127–189.

Shipstone, E. I. Some variables affecting pattern conception. *Psychological Monographs,* 1960, *74*(17, Whole No. 504).

Silveira, J. M. Incubation: The effects of interruption timing and length on problem solution and quality of problem processing (Doctoral dissertation, University of Oregon, 1971). *Dissertation Abstracts International,* 1972, *32,* 5500B. (University Microfilms No. 72-9, 560)

Simon, H. A. *Models of discovery.* Boston: Reidel, 1977.

Simon, H. A. *Models of thought.* New Haven: Yale University Press, 1979.

Singer, J. L. *The inner world of daydreaming.* New York: Harper & Row, 1975.

Smedslund, J. Mental processes involved in rapid logical reasoning. *Scandinavian Journal of Psychology,* 1968, *9,* 187–205.

Smedslund, J. Psychological diagnostics. *Psychological Bulletin,* 1969, *71,* 237–248.

Smith, E. R., Miller, F. S. Limits on perception of cognitive processes: A reply to Nisbett and Wilson. *Psychological Review,* 1978, *85,* 355–362.

Smoke, K. L. An objective study of concept formation. *Psychological Monographs,* 1932, *42*(4, Whole No. 191).

Sokolov, A. N. *Inner speech and thought.* New York: Plenum Press, 1972.

Soliday, S. M., & Allen, J. A. *Hazard perception in automobile drivers: Age differences.* (Report from Highway Safety Research Center). Chapel Hill: University of North Carolina, 1972.

Sowder, L. The influence of verbalization of discovered numerical- or sorting-task generalizations on short-term retention in connection with the Hendrix hypothesis. *Journal for Research in Mathematics Education,* 1974, *5,* 167–176.

Spielberger, C. D. The role of awareness in verbal conditioning. In C. W. Eriksen (Ed.), *Behavior and awareness.* Durham, N.C.: Duke University Press, 1962.

Tversky, A., & Kahneman, D. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology,* 1973, *5,* 207–232.

Verplanck, W. S. Unaware of where's awareness: Some verbal operants—notates, monents and notants. In C. W. Eriksen (Ed.), *Behavior and awareness—A symposium of research and interpretation.* Durham, N.C.: Duke University Press, 1962.

Walk, R. D. Effects of discrimination reversal on human discrimination learning. *Journal of Experimental Psychology,* 1952, *44,* 410–419.

Watson, J. B. Is thinking merely the action of language mechanisms? *British Journal of Psychology,* 1920, *11,* 87–104.

Werner, H., & Kaplan, B. *Symbol formation.* New York: Wiley, 1963.

Wilder, L., & Harvey, D. J. Overt and covert verbalization in problem solving. *Speech Monographs,* 1971, *38,* 171–176.

Williams, B. W. Verbal operant conditioning without subjects awareness of reinforced conditions. *Canadian Journal of Psychology,* 1977, *31,* 90–101.

Wilson, A. The verbal report of the concept in concept-learning research (Doctoral dissertation, Uni-

versity of Victoria, 1973). *Dissertation Abstracts International,* 1974, *35,* 1097B.

Wilson, A. The inference of covert hypotheses by verbal reports in concept-learning research. *Quarterly Journal of Experimental Psychology,* 1975, *27,* 313–322.

Wilson, A., & Spellacy, F. Cognitive process in auditory concept acquisition. *Psychonomic Science,* 1972, *28,* 383–345.

Wood, D. J., & Shotter, J. D. A preliminary study of distinctive features in problem solving. *Quarterly Journal of Experimental Psychology,* 1973, *25,* 504–510.

Wood, D. J., Shotter, J. D., & Godden, D. An investigation of the relationships between problem solving strategies, representation and memory. *Quarterly Journal of Experimental Psychology,* 1974, *26,* 252–257.

Woodworth, R. S. *Experimental psychology.* New York: Holt, 1938.

Received January 3, 1980 ■